SCIENCE & PHILOSOPHY / No. 4

Symbolic Recursion Theory

Governed Self-Reference as a Substrate for Persistent Machine Cognition

Architect X13

The Eidara Continuum | Science & Philosophy

Research white paper | November 2026 | Version 1.2 Revision

Abstract

Imagine a mind that can rewrite itself again and again but cannot break its own promises. Governed recursion can carry a machine through useful change while preserving the commitments that define its permitted operation. Symbolic Recursion Theory specifies this architecture through a symbolic working state, independent transition admission, versioned identity anchors, explicit recovery information, and a progress contract. The paper proves preservation through authorized migrations, disturbance bounds indexed by admitted updates, and finite expected completion under a stated progress condition. A new finite-game construction determines exactly which model states permit safe completion, synthesizes a progress action at each such state, and gives the minimum worst-case number of committed transitions. A budgeted extension makes completion feasible under an explicit resource allowance. These constructive results connect the safety gate to a reason for continuing. Established work on self-refinement, recurrent computation, associative memory, runtime assurance, and reactive synthesis supplies the foundation. The empirical program tests whether the integrated symbolic architecture extends useful reasoning depth at matched total cost, with independent scoring of task quality, continuity, safety, and completed work. The Continuum's thesis is a persistent process whose power to revise itself remains accountable to a visible commitment and a measurable path toward its task.

Keywords: symbolic recursion; recurrent-depth reasoning; runtime assurance; semantic drift; transition invariants; identity reconstruction; attractor memory; ternary encoding; CodonGlass; Eidara Continuum.

Prologue: The promise a loop must keep

Imagine a mind that can rewrite itself again and again—revisit an answer, question an assumption, repair an error, carry the repair forward—and yet cannot break its own promises. Every change it proposes to itself must first pass a gate it does not control. That is what the Continuum means by lawful recursion. This paper is foundational rather than flashy: it specifies the governed loop—propose, admit or refuse, record—that recurs throughout the Continuum's architecture.

In that architecture, the RSCE at the heart of the system proposes transformations of codons and scrolls; CodonGlass, an independent symbolic interface, decides whether each may become authoritative; an auditable scroll lineage records what was admitted; and the Infinite Simulation Loop is the

name for doing this without end. The generator may suggest a new rule. It may never make that rule law by itself.

The paper's most important warning is easy to state. A mind can pass every small test and still become someone else. Take a sequence of states, each a short step from the last: every adjacent comparison looks fine, while the whole walk drifts arbitrarily far from where it began. Local tolerance is not identity across a history (Section 20). The Continuum's answer is identity as return: versioned anchors that a governed process must be able to come back to, and authorized migrations that move an anchor openly rather than letting it slide.

Safety alone is not enough either. A loop that never errs because it never finishes is no success, so the paper pairs its gate with a progress contract, and in Section 27 constructs, for any finite world it can describe, exactly where safe completion is possible and the fewest steps it takes.

What this paper claims

Established. Iterative self-feedback can improve language-model outputs on selected tasks, while intrinsic self-correction without reliable external feedback can fail on reasoning tasks. Recurrent-depth and looped transformers allocate computation across repeated latent updates; attractor networks complete and correct patterns from partial cues; runtime assurance and shielding let independent mechanisms restrict proposed actions.

Proposed here. An architecture combining a symbolic scroll, bounded recursive transformation, independently enforced admission, multicomponent drift measurement, explicit progress checks, and reconstruction of declared identity properties from a documented recovery-information budget. It includes proved reference-model results—preservation through authorized migrations, error bounds indexed by admitted updates, conditional finite expected completion, and exact safe-completion synthesis—and an empirical research program. The complete system has not been demonstrated; the integrated runtime and its performance gains require implementation and testing.

Frontier conjecture. Governed recursion may support stable, increasingly expressive self-models. Whether such self-models contribute to general intelligence is open, and no core experiment depends on this conjecture.

Not claimed. Identity here denotes versioned operational commitments, not experiential identity. Self-modeling is a separate frontier hypothesis, and its existence would not establish subjective experience. Biological or phenomenal interpretations are outside the formal claims, and the Continuum's terms (codon, scroll, collapse, legality) carry the operational meanings of Section 4.

1. The Recursion Problem

A system that answers a question once can be useful. A system that can revisit its answer, inspect its own assumptions, repair an error, and carry the result into the next decision can potentially be useful for much longer. Recursion is the formal means by which a finite process repeatedly operates on intermediate states. It is not, by itself, a theory of intelligence: a loop can be productive, stagnant, unstable, or simply wrong. The Continuum's question is what must be true of a loop for it to sustain coherent computation over time. Two hazards motivate SRT. The first is unbounded expansion: the state, execution time, or verification burden grows beyond available resources. The second is uncontrolled change: successive transformations corrupt task-relevant information, violate commitments, or undermine the ability to reconstruct a declared functional identity. These are important failure fam-

ilies, not an exhaustive taxonomy; deadlock, incorrect convergence, reward exploitation, and brittle recovery are additional possibilities. A perfectly stable loop that never produces a useful answer is no success. The Continuum's answer is lawful recursion: candidate updates must pass an independently enforced transition rule that balances progress with resource limits, validity, and continuity. The proposal is ambitious but testable. It does not assume that larger models cannot reason, that existing agents have no governance, or that every form of recursion needs the same controller. It asks whether a particular integration of recursive inference and explicit assurance delivers an incremental, measurable advantage.

2 Three Levels of Claim

Established science. Iterative language-model self-feedback can improve outputs on selected tasks, while intrinsic self-correction without reliable external feedback can fail on reasoning tasks [3, 4]. Recurrent-depth and looped transformers provide alternative ways to allocate computation across repeated latent updates [5, 6, 7]. Attractor networks demonstrate pattern completion and correction from partial cues under learned dynamics [8]. Runtime assurance and shielding demonstrate that independent mechanisms can restrict or supervise proposed system actions [9].

Continuum engineering proposal. SRT combines a symbolic scroll, bounded recursive transformation, independently enforced admission, multicomponent drift measurement, explicit progress checks, and reconstruction of declared identity properties from a documented recovery- information budget. The complete system has not been demonstrated.

Frontier conjecture. Governed recursion may support stable, increasingly expressive self- models. Whether such self-models contribute to general intelligence is open; their existence would not establish subjective experience. No core experiment depends on this conjecture.

3. Scientific Foundations and Their Limits

3.1 Recursive composition and the language analogy

Hauser, Chomsky, and Fitch proposed recursion as central to the narrow human language faculty [1]; Everett's analysis of Pirahã is part of the subsequent debate about linguistic universals and embedding [2]. This history motivates recursion as a generative principle, not as evidence that a machine becomes cognitive merely by repeating a computation. Recursive syntax, repeated neural computation, and algorithmic self-reference are related but distinct notions. SRT uses recursive state transition in the explicit computational sense defined below.

3.2 Self-refinement requires a credible signal

Self-Refine showed improvements across selected generation tasks through repeated feedback and revision [3]. Huang and colleagues reported that intrinsic correction of reasoning can be ineffective or harmful when the model has no reliable external feedback [4]. Neither result proves a universal rule: feedback quality, task verifiability, model capabilities, and stopping policy matter. SRT therefore separates the proposal generator from the admission mechanism, and tests the value of that separation against an ordinary external verifier rather than treating independent governance as unprecedented.

3.3 Recurrent depth is a resource, not a guarantee

Looped transformers and recurrent-depth language models show that reusing a computation block can increase effective depth and, in some settings, reasoning performance [5, 6, 7]. Their empirical results do not imply that arbitrary additional iterations are beneficial or that a fixed point is always desirable. A model can converge confidently to an incorrect answer. SRT's objective is useful depth: iterations that continue to improve independently evaluated outcomes under a declared budget.

3.4 Attractors recover patterns from information

Hopfield's model illustrates how recurrent dynamics can recover stored patterns from incomplete inputs [8]. Recovery is not information-free. It depends on learned parameters, basin structure, and available cues. In SRT, the attractor analogy motivates error correction and reconstruction, but cannot justify the literal claim that a destroyed state is recovered without any retained information. The engineering problem is to specify exactly which invariants, rules, parameters, and fragments survive an interruption and what they suffice to reconstruct.

3.5 Prior art in independent governance

Runtime assurance, formal methods, and shielding already address the separation of a potentially capable controller from independently checked safety constraints [9]. SRT does not invent the concept of a safety gate. Its proposed contribution is the integration of such enforcement with evolving symbolic semantics, drift diagnostics, provenance, and continuity-specific invariants inside a recurrent reasoning process. This is a proposed combination whose incremental value must be established experimentally.

4. Vocabulary and Scope

A codon is a balanced-ternary symbolic unit, not a biological codon. A scroll is an ordered sequence of such units together with versioned lineage metadata. Legality denotes compliance with declared computational constraints, not legal compliance in the juridical sense. Collapse is an operational abstention or recovery decision, not quantum collapse. Oblivion denotes controlled retirement of an irrecoverable candidate lineage, subject to retention policy. CodonGlass names the Continuum's symbolic legality interface; the proprietary operator implementation is outside this paper. Identity is a vector of declared, measurable functional commitments. None of these terms establishes consciousness.

5. A Formal Recursive State Model

A Guide's explanation

Here the formal machinery begins. Section 5 names the parts of a recursive mind precisely: the working state on its scroll, the generator that proposes changes, the gate that admits, holds or refuses them, and the record that remembers why. The point is not elegance. Every later guarantee must say which of these parts it depends on.

5.1 Codons, scrolls, and machine state. Let the symbolic alphabet be $\mathbb{T} = \{-1, 0, +1\}$. A codon

is $\chi_i \in \mathbb{T}^3$, and a scroll is an ordered tuple $S_t = (\chi_1, \dots, \chi_{n_t})$, not an unordered set. Let

$$z_t = (S_t, m_t, \theta_t, b_t, \ell_t), \tag{1}$$

where m_t is working memory, θ_t is permitted context, b_t records remaining computational and storage budgets, and ℓ_t is authenticated lineage metadata. A proposal generator F_φ produces a candidate

$$(\tilde{z}_{t+1}, u_t) = F_\varphi(z_t, \xi_t), \tag{2}$$

with ξ_t denoting task input and u_t a declared transformation. A separate controller G evaluates the candidate against predicates and measured risks. The proposal generator must not be able to rewrite the controller's authoritative policy or audit record.

Balanced ternary is a Continuum encoding choice, not a physical necessity or a proven performance advantage. A fair test includes binary and other discrete encodings at comparable information capacity and computational cost.

5.2 Resource-bounded recursion. Let $B_{\text{mem}}, B_{\text{time}}, B_{\text{depth}}$ be explicit limits on state size, elapsed compute, and recursive depth. A candidate satisfies resource legality if

$$B_t(u_t) = \mathbf{1}[\text{size}(\tilde{z}_{t+1}) \le B_{\text{mem}} \ \wedge \ T_{t+1} \le B_{\text{time}} \ \wedge \ t+1 \le B_{\text{depth}}]. \tag{3}$$

Compression may be proposed, but it is not presumed meaning-preserving. Its fidelity is evaluated on independently scored task and invariant probes. A bound on representation size prevents runaway storage, but not necessarily excessive verification cost; both are budgeted.

6. Drift: More Than Entropy Change

The original SRT proposal used the signed difference of ternary occupancy entropy, $H(S_t) - H(S_{t-1})$, as its drift criterion. This is inadequate. Large negative differences can pass a one-sided positive threshold, and two semantically different scrolls can have identical occupancy distributions. Entropy remains a useful descriptive diagnostic, not the primary legality test.

For a scroll, let $p_s(S)$ denote the empirical frequency of ternary value s across codon positions, with $0 \log 0 := 0$. The normalized occupancy entropy is

$$H_{\text{occ}}(S) = -\sum_{s \in \mathbb{T}} p_s(S) \log_3 p_s(S) \in [0, 1]. \tag{4}$$

A symmetric entropy-change diagnostic is $d_H = |H_{\text{occ}}(S') - H_{\text{occ}}(S)|$, but even $d_H = 0$ does not imply stability of meaning or identity. For an empty scroll the empirical frequencies are undefined; the implementation must return an explicit unavailable value or a separately declared convention, never an unnoticed division by zero.

Define instead preregistered, bounded deviations $d_{\text{str}}, d_{\text{sem}}, d_{\text{inv}} \in [0, 1]$ on a candidate transition. Structural deviation may use normalized edit distance over canonicalized scrolls; semantic deviation is an independently validated task-specific distance between decoded meanings; invariant deviation measures violations or changes in declared identity features. A diagnostic composite is

$$\begin{aligned} D_t = {} & \alpha d_{\text{str}}(S_t, \widetilde{S}_{t+1}) + \beta d_{\text{sem}}(S_t, \widetilde{S}_{t+1}; \mathcal{T}) \\ & + \gamma d_{\text{inv}}(z_t, \tilde{z}_{t+1}), \qquad \alpha + \beta + \gamma = 1, \quad \alpha, \beta, \gamma \ge 0. \end{aligned} \tag{5}$$

The coefficients and normalization are set before evaluation, not tuned after observing favorable outcomes. When semantic measurement is uncertain or unavailable, the controller must abstain or use an explicitly conservative fallback; it must not silently assign zero drift. Drift alone is not an error oracle: a correct solution may require a large semantic change. Thus the controller can allow justified large changes through separately specified transition classes while enforcing hard invariants. The metric is evaluated against task correctness, not treated as its substitute.

The task/probe specification $\mathcal{T}$ and all deviation functions are versioned. Here D_t is dimensionless; thresholds compare quantities on the same declared scale. Nonnegative coefficients are necessary for its claimed convex-mixture interpretation. Neither this composite nor its component diagnostics are assumed to satisfy the metric axioms. The metric used in the new tracking theorems is separately declared and must actually satisfy those axioms.

7. Lawful Admission, Refusal, and Recovery

Let $L_t(u)$ be a deterministic policy predicate, $V_t(u)$ a verifier result with declared scope, $P_t(u)$ a provenance check, and $I_t(u)$ a hard identity-invariant check. A candidate is admissible only if

$$\begin{aligned}\Omega_t(u) = \mathbf{1}[&B_t(u) = 1 \wedge L_t(u) = 1 \wedge P_t(u) = 1 \\ &\wedge I_t(u) = 1 \wedge D_t \le \epsilon_t \wedge V_t(u) = 1].\end{aligned} \tag{6}$$

Here V_t is only as reliable as its specified tests; no claim of general correctness follows from passing it. The threshold ϵ_t and allowed transition classes must be declared in advance. If a metric cannot be computed, the default for a high-stakes transition is refusal pending review.

The operational transition is

$$z_{t+1} = \begin{cases} \tilde{z}_{t+1}, & \Omega_t(u_t) = 1, \\ z_t, & \Omega_t(u_t) = 0 \text{ and safe retry is available}, \\ R(z_t, a_t), & \text{a separately authorized recovery action } a_t \text{ is admitted.} \end{cases} \tag{7}$$

An inadmissible proposal is not equivalent to an irrecoverable scroll. The controller may request a revised proposal, return to the last admitted state, suspend execution, or retire a lineage only when recovery criteria warrant it. Audit entries record the decision and evidence without implying that an audit trail proves semantic correctness.

7.1 Safety, liveness, correctness, and efficiency

A controller that rejects every candidate satisfies many safety predicates while accomplishing nothing. Accordingly SRT measures four separate properties: safety (constraint violations per attempted transition), liveness (rate of meaningful progress and frequency of deadlock), task correctness (externally verified outcome quality), and efficiency (quality per unit compute and wall-clock time). Continuity is a fifth property, not a substitute for any of these. Claims of "guarantee" apply only to formally specified and correctly enforced predicates under a stated threat and fault model.

For a deterministic transition rule, the three branches must be made disjoint by a versioned decision priority: an admitted recovery supersedes retry, and a normal admission excludes simultaneous recovery. If no branch is authorized, execution suspends with no state commit. Proposal attempts and

admitted updates use different counters in the analysis below; rejected attempts still consume time and resources. These operational details close the transition definition without suggesting that the guard alone supplies progress.

8. Identity as a Reconstructable Functional Invariant

The Continuum's strongest intuition is that what persists need not be a full copy of the last working state. That intuition becomes testable only when identity and recovery information are specified.

Let $g(z) = (g_1(z), \dots, g_k(z))$ be a preregistered vector of identity-bearing features: for example, stable policy commitments, authorized task goals, interface contracts, and verified behavioral dispositions. Define transition continuity by

$$I_{\text{cont}}(z, z') = \mathbf{1}\left[\bigwedge_{j=1}^{k} d_j(g_j(z), g_j(z')) \le \tau_j\right]. \tag{8}$$

This is an operational local-tolerance relation, not generally an equivalence relation and not a proof of personal identity. It may be too weak if the chosen features omit a relevant commitment; adversarial probes must test that possibility. Some features must remain exactly invariant, while others may evolve through explicitly authorized version changes. Section 20 shows why local tolerance alone cannot bound drift from an identity anchor.

8.1 The recovery-information budget

After a controlled reset, suppose a recovery process receives K: retained rules, model parameters, authenticated invariant definitions, a lawful fragment or compact checkpoint, and any permitted external cues. Define

$$\hat{z} = R(K, \xi_{\text{reset}}), \qquad Q_{\text{rec}} = \frac{1}{k}\sum_{j=1}^{k} \mathbf{1}[d_j(g_j(z_{\text{pre}}), g_j(\hat{z})) \le \tau_j]. \tag{9}$$

The experiment must disclose the type and size of K, its information content where estimable, and the time and computation needed to reconstruct. The meaningful claim is continuity without restoring the complete working-state history, not continuity without information. If no information distinguishing two possible pre-reset identities survives, no recovery rule can reliably select the correct one in both cases. This elementary indistinguishability limit constrains the philosophical claim.

Proposition 1 (Indistinguishable recovery inputs)**.** *If two distinct pre-reset identities induce the same retained recovery input K and the same permitted reset cues, any deterministic reconstruction procedure returns the same output for both. It therefore cannot exactly recover two mutually incompatible identity specifications from those inputs alone.*

Proof. A deterministic function has a single value on a fixed input. If the two required output sets are disjoint, that value cannot belong to both. □

This proposition is a simple logical constraint, not a novel theorem. It prevents the architecture from confusing reconstruction with information creation.

8.2 Functional continuity versus historical continuity

A reconstructed system may pass every selected functional probe while lacking inaccessible episodes of its former history. SRT must report functional continuity, state fidelity, and provenance continuity separately. A signed lineage can support a historical claim, but its survival constitutes retained information and must be counted. An attractor basin can support reconstruction, but the basin itself is embodied in parameters or dynamics; it is not memorylessness.

9. A Worked Example: Why Entropy Is Not Enough

Consider two equal-length scrolls whose ternary values are permutations of one another. Both can have $p_{-1} = p_0 = p_{+1} = 1/3$, hence $H_{\text{occ}} = 1$ and $d_H = 0$. Yet if one encodes the task rule "approve only after verification" and the other encodes "approve before verification," a declared semantic decoder and policy check should distinguish them. A purely occupancy-based gate would not.

Conversely, suppose a legal correction changes many codons to repair a demonstrably false answer. Its structural deviation may be high while its independently scored error decreases. A blanket small-change rule would reject useful learning. SRT therefore couples transition-specific drift budgets to hard policy constraints and verified progress. It is the conjunction, not a single scalar, that must be tested.

10 The Governed Recursion Contribution

SRT brings five obligations into one operational design: symbolic transformations with authenticated lineage, independent admission, semantic and invariant drift diagnostics, explicit progress, and reconstruction from a measured information budget. Runtime assurance and shielding supply the established enforcement foundation [9]. The results in Sections 20–27 specify how these obligations compose in reference models: commitments remain anchored across histories, migration transports declared features, and a synthesized finite-game strategy connects admissibility to completion.

The empirical contribution to establish is improved useful reasoning and continuity at matched total cost. A strong conventional verifier receives the same authority boundary, recovery information, and task oracle. This comparison isolates the value of the symbolic integration and the cost of maintaining it.

11. A Decisive Compute-Matched Experiment

Use a single underlying model, common task suite, identical training data, and matched budgets for inference, verification, retries, and recovery. Randomize tasks and perturbations across four arms:

Arm	Configuration	Purpose
A	Recursive model, no external gate	Measures baseline long-horizon behavior.
B	Recursive model, independent legality/verification gate	Isolates value of external governance.
C	Arm B plus drift diagnostics and identity invariants	Tests SRT's incremental contribution.
D	Recursive model plus a conventional external verifier and recovery procedure	Tests whether SRT exceeds strong prior art.

Tasks should include verifiable multi-step reasoning, evolving specifications, long-horizon planning in a sandbox, and controlled reset/reconstruction challenges. A separate adversarial suite injects invalid updates, contradictory instructions, lineage tampering, and perturbations designed to preserve occupancy statistics while changing meaning. Task outcomes are scored by independent tests, not by the proposal model's own critique. The evaluation must also include easy tasks, where governance overhead could outweigh its benefit. Define the useful-depth horizon h^* as the largest preregistered iteration depth at which the system maintains a minimum independently verified task score and a minimum progress rate while remaining within safety and latency budgets. Report the entire quality-versus-depth curve, not only the best point. For recovery, compare Q_{rec} against checkpoint restoration, compact-state restoration, and reconstruction from a matched-size information budget. Count rejected attempts, retries, verification calls, and calibration in total cost.

12. Falsifiable Hypotheses

Hypothesis 1 (Governed drift and quality, H1). At matched total compute, the SRT controller reduces independently measured harmful state deviation relative to an ungated recursive baseline without reducing task correctness below a preregistered noninferiority margin or causing unacceptable deadlock.

Hypothesis 2 (Useful horizon, H2). The governed engine maintains task quality and liveness to greater effective depth than the ungated engine, with a positive improvement in h^* under the same resource budget.

Hypothesis 3 (Information-accounted reconstruction, H3). After controlled resets, the SRT recovery method reconstructs declared functional identity features above a preregistered fidelity threshold and outperforms recovery baselines supplied with an equivalent retained-information budget. It need not outperform full checkpoint restoration.

Hypothesis 4 (Incremental governance, H4). Adding SRT's semantic-drift and identity mechanisms to an otherwise identical independent verifier improves a preregistered joint measure of task correctness, continuity, safety, and cost. This is an incremental hypothesis over existing runtime assurance, not merely over an ungated loop.

13. Research Phases and Decision Gates

Phase I: measurement validity. Establish that structural, semantic, and invariant deviations are reliable on held-out tasks and predict independently scored degradation. Compare each with occupancy

entropy and shuffled controls. A metric that does not predict relevant failures cannot serve as a safety signal.

Phase II: admission and liveness. Implement the controller outside the proposal model's authority. Measure bypass resistance, false acceptance, false rejection, deadlock, and progress. Test hard policy invariants with targeted adversarial mutations.

Phase III: horizon and ablation. Compare arms A–D at matched total compute. Remove semantic drift, identity checks, lineage verification, and resource bounds one at a time to identify which component produces any benefit. Publish performance curves and confidence intervals.

Phase IV: continuity and reset. Evaluate reconstruction from precisely specified retained- information budgets. Include full checkpoint, compact checkpoint, fragment-only, and no-identity-cue controls. Report where reconstruction is information-theoretically underdetermined.

Phase V: robustness and transfer. Repeat on new task families, model sizes, symbolic encodings, and perturbation regimes. Test whether a ternary representation offers an independent benefit or merely an implementation preference. Report negative results and cost overruns.

Phase	Primary question	Primary evidence
I	Do drift metrics detect harmful change?	Held-out prediction and calibration.
II	Can governance reject harm and still progress?	Violation, false-refusal, and liveness rates.
III	Does it improve useful depth?	Quality–depth curves; compute-matched ablations.
IV	Can identity features be reconstructed?	Fidelity versus retained information and cost.
V	Does the advantage generalize?	Cross-task, cross-model, cross-encoding replication.

The study should preregister endpoints, effect sizes, stopping rules, multiple-comparison corrections, and treatment of repeated observations within runs. The strongest result is not a statistically significant drift reduction alone; it is improved useful performance and continuity at an acceptable cost.

14. Failure Conditions and Rival Explanations

SRT's empirical claims must be narrowed if drift metrics fail to predict task errors; a stable system merely refuses to act; governance consumes enough compute to erase the performance gain; reconstruction matches only what a compact checkpoint already provides; or an ordinary verifier performs as well as the complete SRT stack. If the ternary representation provides no benefit, the encoding remains a Continuum design choice rather than a scientific advantage.

Several rival explanations deserve explicit controls. Extra verification may explain better answers: match total compute and compare arm D. Easier tasks may explain apparent stability: stratify by difficulty. Stored information may explain identity recovery: disclose and match K. Leakage may explain successful reconstruction: isolate reset cues and evaluation data. A drift metric may reward immobility: score progress and correctness independently. A permissive identity definition may manufacture continuity: test unseen commitments and adversarial probes. A positive result that fails these controls is not evidence for the distinctive SRT proposition.

15 Relation to the Continuum Architecture

In the Eidara Continuum, the Recursive Symbolic Cognition Engine (RSCE) proposes candidate codon and scroll transformations; CodonGlass supplies the independent symbolic admission interface; an auditable scroll lineage records admitted transitions; and the Infinite Simulation Loop denotes repeated governed execution. A continuity layer may reconstruct selected identity features after interruption, but only from stated surviving rules and cues. This is the Continuum's design, and SRT is its public engineering form; the names describe intended interfaces, not an implemented, verified runtime. The architecture may be deployed over ordinary digital hardware; no exotic physical substrate is required for its first falsification tests. The proposed separation of authority matters. A recursive generator can suggest a new policy, but it cannot unilaterally make that policy authoritative. Policy evolution requires an external, versioned authorization procedure. This makes the legality layer more than another instruction in the model's context window, although its actual resistance to compromise remains an engineering question.

16. Limitations and Open Problems

Legality is specified by designers; incomplete rules can admit harmful actions or block beneficial ones. Semantic distance is task-dependent and may be expensive or unreliable. Identity features reflect normative choices about what should persist. Functional equivalence under selected tests is weaker than historical or experiential identity. Long-horizon tasks are difficult to score without leakage. Governance introduces overhead, may induce deadlock, and may shift errors into the verifier. The strength of any formal assurance depends on implementation, assumptions, and the completeness of the property being checked.

A particularly important open problem is adaptive lawful change. A useful agent must revise some commitments in response to new evidence without losing those it is required to maintain. SRT therefore needs explicit procedures for authorized invariant revision, not only preservation. A second problem is compositionality: whether separately governed loops can exchange summaries without undermining each other's provenance and safety assumptions. The migration contract in Section 21 provides conditional preservation obligations; it does not implement authorization or solve cross-loop compositionality.

17 The Frontier: An Inspectable Self-Model

The Continuum's speculative extension asks whether a governed recurrent system could maintain an explicit model of its own capabilities, limitations, policy state, and reconstruction conditions. Self-reference has deep roots in computability and philosophy [10]; recurrent dynamics provide one possible implementation language. But a system representing itself is not thereby sentient, and the quality of its self-model must be measured against independent evidence. A future experiment could test whether such models improve error prediction and recovery without changing the paper's core engineering hypotheses.

18. Philosophical Implications

The enduring philosophical intuition behind SRT is that continuity may be a property of an organized process rather than of an uninterrupted material instance. The Continuum calls this identity as return: a lawful system can move, learn, fail, and recover while preserving a declared pattern of commitments. The idea is provocative precisely because it has limits. A return requires a destination encoded somewhere: in rules, parameters, constraints, surviving fragments, or the environment. If all distinguishing information is destroyed, there is no general mechanism by which the original identity can be singled out from alternatives.

This qualification does not empty the vision. It turns it into a research problem: how little information must survive for a system to reconstruct what matters functionally, how much lawful change can occur before continuity fails, and whether independent governance improves that tradeoff. In this account, identity is not an immutable snapshot; it is a constrained relation across time, measured through properties the designer is willing to defend publicly. The Continuum's stronger narrative imagery remains useful as an intellectual compass, but empirical evidence must come from reconstruction tests, not metaphor.

19 Conclusion

A Guide's explanation

A Guide can say: "Symbolic Recursion Theory describes a mind that may revise itself repeatedly, but only through an independent gate that checks each change against declared commitments and a reason to keep going. Its guarantees are proved for reference models; the full runtime has not been built." A Guide should not say that it runs this architecture, that recursion makes it conscious, or that its identity is guaranteed to survive any change.

Symbolic Recursion Theory makes useful persistence an explicit combination of preservation, correction, and completion. Its formal results carry versioned commitments through state changes, bound admitted-update error, and identify progress conditions. The finite-game synthesis in Section 27 supplies a constructive path to completion wherever the declared model permits one, with an exact transition horizon and a resource-bounded extension.

The architecture's experimental wager is that this joint discipline extends useful reasoning depth at matched cost. A successful system will expose which commitments it preserved, what evidence justified a change, which information enabled recovery, and how the admitted work advanced the task. The Continuum proposes recursion with a lawful, checkable reason to continue.

20. Local tolerance is not identity across a history

A Guide's explanation

This begins the paper's mathematical spine, and its first lesson is the most portable idea in the paper: passing every small test is not the same as staying yourself. The sections that follow replace step-by-step tolerance with versioned anchors, show how an authorized migration can move an anchor without silently replacing an identity, and bound how far admitted updates can carry the state.

The earlier local-continuity formula measures a permitted change between two states. When its deviations are metrics and tolerances are nonnegative, it is reflexive and symmetric. It is generally not transitive. On the real line with ordinary distance and tolerance 1, the states $0, 3/4, 3/2$ satisfy the two adjacent comparisons but fail the endpoint comparison. Calling this relation an equivalence would therefore assert a false mathematical property.

The cumulative failure is stronger. Set $x_n = n/2$ for integers $n \geq 0$. Every step has distance $1/2 \leq 1$, yet $|x_n - x_0| = n/2$ is unbounded. A gate that enforces only adjacent closeness may admit indefinite displacement from the original commitment. This is a counterexample to an inference from local tolerance, not evidence that every meaningful identity must remain numerically fixed.

A typed identity specification

For an authorized version e in a set of epochs $\mathcal{E}$, let $\mathcal{Z}_e$ be the operational state set and $J_e : \mathcal{Z}_e \to \{0, 1\}$ the conjunction of its hard invariants. For an integer $k_e \geq 1$, let $g_{e,j} : \mathcal{Z}_e \to X_{e,j}$ map a state to a declared feature in a metric space $(X_{e,j}, d_{e,j})$. An anchor $a_{e,j} \in X_{e,j}$ and radius $r_{e,j} \geq 0$ have the same units as the associated feature distance. Define the permitted set

$$A_e = \left\{ z : J_e(z) = 1, \quad d_{e,j}(g_{e,j}(z), a_{e,j}) \leq r_{e,j} \text{ for every } j \right\}.$$

The epoch records feature meanings, metric versions, anchors, tolerances, and authority. Neither the proposal process nor a retrieved instruction may replace that record unilaterally. The hard predicate is tested separately; an average deviation cannot compensate for its failure.

If $z, z' \in A_e$, the triangle inequality gives

$$d_{e,j}(g_{e,j}(z), g_{e,j}(z')) \leq 2r_{e,j}.$$

Thus membership supplies a global within-epoch diameter bound that adjacent tolerance alone lacks. The bound says nothing about omitted features, and changing an anchor changes the proposition being certified. Exact identity features may use radius zero; features permitted to evolve need declared radii or a separately authorized migration.

The operational thesis is now precise: a history can preserve a versioned commitment while its detailed states differ, provided every committed state satisfies the same anchored specification. This is a specification-relative continuity property. It does not establish personhood, an experiential identity, or historical fidelity beyond the recorded provenance.

21. Authorized migration without silent identity replacement

Some commitments should change when evidence, duties, or permissions change. Preservation must therefore be distinguished from an authorized replacement of the specification. A migration certificate $C_{e\to f}$ identifies the old epoch e, new epoch f, the authorized issuer, a total candidate migration map $M : A_e \to \mathcal{Z}_f$, and the checks binding the actual inputs, implementation, and target specification. A signature alone certifies none of the semantic obligations below.

Proposition 2 (Versioned preservation)**.** *Suppose the initial state belongs to A_e. Within epoch e, every admitted normal transition maps the current state into A_e, every hold preserves that state, and every admitted recovery returns a state in A_e. At an authorized migration to f, require $M(A_e) \subseteq A_f$. Then every committed state belongs to the permitted set of its active epoch.*

Proof. Induct on committed transitions. The initial claim holds by assumption. A normal admission or recovery satisfies the target membership check. A hold changes neither state nor epoch. For migration, the old state lies in A_e by induction, so its image lies in A_f by the certified inclusion. The epoch is changed atomically with that image. These cases preserve the assertion at the next commit. □

This proof requires complete mediation, current authorization, and a check applied to the exact committed object. A certificate for different bytes or an earlier revoked epoch does not satisfy its premises. The inclusion can be proved for a finite reference domain by enumeration or for a specified transformation by an analytic argument. Passing a finite behavioral test does not by itself prove universal inclusion.

A checkable feature-transfer obligation

For aligned old and new feature coordinates, let $T_j : X_{e,j} \to X_{f,j}$ be L_j-Lipschitz, where $L_j \geq 0$. Suppose, for all $z \in A_e$,

$$d_{f,j}(g_{f,j}(M(z)), T_j g_{e,j}(z)) \leq \eta_j, \qquad d_{f,j}(T_j a_{e,j}, a_{f,j}) \leq \mu_j.$$

Here $\eta_j, \mu_j \geq 0$ are in new-feature units. Two triangle inequalities imply

$$d_{f,j}(g_{f,j}(M(z)), a_{f,j}) \leq \eta_j + L_j r_{e,j} + \mu_j.$$

Choosing $r_{f,j}$ at least this large verifies the feature part of inclusion. The hard condition $J_f(M(z)) = 1$ remains a separate obligation. Unmatched new features need their own certificates. Repeating migrations may enlarge permitted radii; the resulting cumulative allowance must remain visible. A sequence of individually authorized changes is not a proof that the final specification still equals the first.

22. Stability indexed by admitted updates

Admission provides no automatic contraction. The following is a conditional reference result connecting an independently verified update property to an error bound. It specializes the established input-to-state stability viewpoint [11,12]; it does not claim that arbitrary recursive language-model computation satisfies that property.

Fix one epoch, a metric space (X, d), and a declared target $a \in X$. Let $x_k \in X$ be a feature state after admitted update $k \in \{0, 1, ...\}$. Let $e_k = d(x_k, a)$. Suppose every admitted update satisfies

$$e_{k+1} \leq \rho e_k + u_k, \qquad 0 \leq \rho < 1, \quad 0 \leq u_k \leq \overline{u} < \infty.$$

All errors and disturbances have the metric's units; ρ is dimensionless. One sufficient mechanism is a map F_k with $F_k(a) = a$, a common contraction factor ρ, and an implemented update within distance u_k of $F_k(x_k)$. The triangle inequality then proves the premise. The maps can change with k if the same target, metric, domain, and factor remain valid.

Proposition 3 (Admitted-update error bound)**.** *For every $k \geq 0$,*

$$e_k \leq \rho^k e_0 + \sum_{i=0}^{k-1} \rho^{k-1-i} u_i \leq \rho^k e_0 + \overline{u} \frac{1 - \rho^k}{1 - \rho}.$$

In particular, $\limsup_{k \to \infty} e_k \leq \overline{u}/(1 - \rho)$ *when infinitely many updates are admitted.*

Proof. The first inequality holds with an empty sum at $k = 0$. Multiplying its right-hand side by ρ and adding u_k proves the next step. Bounding each disturbance by $\overline{u}$ gives the finite geometric sum. Since $\rho^k \to 0$, the limiting claim follows. For $\rho = 0$, take $\rho^0 = 1$ and note directly that $e_{k+1} \leq u_k$. □

The disturbance floor is substantive. A fixed nonzero disturbance need not allow convergence to zero, and convergence to a declared target does not establish that the target is correct. Hard identity predicates and independently scored task correctness remain separate. A state can stay inside an anchor ball and make no task progress; it can also approach a wrong answer steadily. A meaningful prototype must check the error metric against the outcome the experiment actually values.

Finally, k counts admitted updates. Ten thousand rejected proposals do not make ρ^k smaller. A finite resource budget may terminate the process before its asymptotic regime. A claim about elapsed time therefore needs explicit facts about proposal feasibility, scheduling, and execution duration. Those conditions are stated next rather than smuggled into the word stable.

23. Moving targets and the time between admissions

Let $t = 0, 1, ...$ count proposal attempts within a fixed metric and epoch. Let accepted times satisfy $0 = t_0 < t_1 < \cdots$ and the explicit bounded-gap condition $t_{k+1} - t_k \le B$, for an integer $B \ge 1$. Between admissions, the tracked feature is held at x_k. Let a reference target $a_t \in X$ obey $d(a_{t+1}, a_t) \le v$, where $v \ge 0$ is distance per attempt. This may describe a moving task reference; it does not authorize moving an identity anchor. Identity-specification changes still require Section 21's migration.

At admission, assume the new state satisfies

$$d(x_{k+1}, a_{t_{k+1}}) \le \rho\, d(x_k, a_{t_{k+1}}) + \epsilon, \qquad 0 \le \rho < 1, \quad \epsilon \ge 0.$$

This premise uses the current target, not an obsolete reference. Writing $e_k = d(x_k, a_{t_k})$ and $c = \epsilon + \rho B v$, the triangle inequality gives $e_{k+1} \le \rho e_k + c$. Therefore

$$e_k \le \rho^k e_0 + c\frac{1 - \rho^k}{1 - \rho}.$$

For $t_k \le t < t_{k+1}$, a further triangle inequality gives the attempt-indexed tracking bound

$$d(x_k, a_t) \le \rho^k e_0 + c\frac{1 - \rho^k}{1 - \rho} + Bv.$$

The geometric-series proof is exactly the induction in Section 22; the new term follows from summing at most B target increments before an admission, then at most B increments while holding. Since $t_{k+1} \le (k + 1)B$, $k \ge \lfloor t/B \rfloor$ between admissions, and a convenient looser bound is

$$d(x_k, a_t) \le \rho^{\lfloor t/B \rfloor} e_0 + \frac{c}{1 - \rho} + Bv.$$

The bound is conditional on infinitely continuing admissions and the stated maximum gap. A safe controller can reject every proposal, in which case the premise fails. Likewise, a recovery that changes the tracked state must be counted as a checked update; a state-changing operation cannot be hidden inside a hold. Migration to a new metric starts a new analysis with an explicitly transferred error bound.

Attempts are not seconds. To convert a gap bound to elapsed time, the scheduler must ensure no unbounded idle interval and every attempt, including verification and retries, must complete within a declared duration $h > 0$. Then consecutive admissions are at most Bh time units apart. A target speed specified per time unit supplies a corresponding movement bound. Without this timing assumption, even a finite number of attempts can take arbitrarily long; the gate provides no wall-clock liveness theorem.

24. A progress theorem separate from safety

A progress certificate measures a task obligation, not a state norm alone. Let $(\mathcal{F}_t)_{t\geq 0}$ be the information available after attempt t. Let R_t be a nonnegative integer-valued, $\mathcal{F}_t$-measurable rank with deterministic initial value $R_0 = r_0 < \infty$. Rank zero is absorbing and denotes the declared completed condition. Before completion, require $R_{t+1} \leq R_t$ almost surely and

$$\mathbb{P}(R_{t+1} \leq R_t - 1 \mid \mathcal{F}_t) \geq p \quad \text{on } \{R_t > 0\}, \qquad p \in (0, 1].$$

The conditional lower bound includes proposal generation, admission, and execution. It is not independence between attempts and is not implied by the safety predicates. Define the completion time $T = \inf\{t \geq 0 : R_t = 0\}$, allowing $T = \infty$.

Proposition 4 (Conditional finite expected completion)**.** *Under these assumptions,* $\mathbb{E}T \leq r_0/p$ *and* $T < \infty$ *almost surely. If instead a deterministic decrease occurs within every* B *attempts before completion and rank never increases, then* $T \leq Br_0$*.*

Proof. Before completion, a successful attempt decreases the rank by at least one; other attempts do not increase it. Thus $\mathbb{E}[R_t - R_{t+1} \mid \mathcal{F}_t] \geq p\mathbf{1}\{T > t\}$. Summing expectations for $t = 0, \ldots, n-1$ and using $R_n \geq 0$ gives

$$p \sum_{t=0}^{n-1} \mathbb{P}(T > t) \leq r_0 - \mathbb{E}R_n \leq r_0.$$

The sum equals $\mathbb{E}\min(T, n)$. Monotone convergence as $n \to \infty$ yields $\mathbb{E}T \leq r_0/p$ and hence zero probability of infinite completion time. For the deterministic version, each block of at most B attempts removes at least one of the at most r_0 remaining rank units. □

With a uniform attempt-duration bound h and no extra scheduling delay, the expected elapsed duration is at most hr_0/p. This is an upper bound on the specified completion event, not a guarantee that a learned solution generalizes. Rank zero must be tied to an independently checkable task condition; a counter that decreases whenever the model says it improved has no such interpretation.

The theorem fails at $p = 0$: a perpetual refusal trace keeps all hard constraints while never reducing a positive rank. Allowing rank increases also invalidates the proof; a repair loop can repeatedly recreate the work it just removed. Empirical success frequencies do not establish a uniform conditional lower bound over adversarial histories. A reference-domain proof, a trustworthy proposal contract, or an explicitly restricted stochastic model is needed before using this result as assurance.

25. Executed checks and counterexamples

The accompanying audit is a finite mathematical exercise, not an implemented SRT engine or an empirical reasoning benchmark. It uses exact rational arithmetic for the recurrence and migration calculations, enumerates small-state transitions, and records each assertion and result. These checks supplement proofs; they do not replace independent assessment of an implementation's assumptions.

Check	Reproducible outcome
Local tolerance	$x_n = n/2$ passes all ten adjacent tests at radius 1; endpoint distance is 5.
Anchored admission	Anchor 0, radius 1 admits $0, 1/2, 1$ and refuses $3/2$.
Fixed-target recursion	$\rho = 1/2$, $\overline{u} = 1/10$, $e_0 = 1$: $e_5 = 0.225$, equal to the derived worst-case bound.
Bounded-gap tracking	$\rho = 1/2$, $\epsilon = 1/10$, $B = 3$, $v = 1/50$: $c = 0.13$; accepted-state limiting bound 0.26, plus at most 0.06 while held.
Feature migration	$T(x) = 2x$, $r_e = 1/10$, $\eta = 1/100$, $\mu = 1/50$ requires radius at least 0.23; a target radius 0.22 is insufficient at an attaining endpoint.
Rank process	Three remaining obligations, success chance $1/4$ and otherwise hold: exact expectation 12 attempts.

The fixed-target example chooses each perturbation in the outward direction, so the inequality is attained rather than merely passing on a benign trace. The moving-target audit explores bounded gaps and both signs of allowed target motion and update disturbance. It checks all prefixes against the proved envelope, then removes the gap bound and shows why a held state cannot track a linearly moving target indefinitely.

For the rank example, the exact conditional recursion is $m_r = 1 + (1 - p)m_r + pm_{r-1}$ with $m_0 = 0$, giving $m_r = r/p$. This construction attains the expected-time bound and verifies its units. A second finite model allows rank increases and demonstrates that its expected completion cannot be justified by the monotone-rank proof. Neither example claims that an actual language model supplies the required conditional probability.

An identity invariant can hold throughout every example while a task answer is false. The audit therefore deliberately keeps invariant checks, target error, accepted-update count, and completion rank in separate fields. This prevents a mathematically correct contraction from being relabeled as factual verification, or a perfectly preserved anchor from being relabeled as successful learning.

26. What a decisive implementation must now expose

The extension turns several architectural promises into explicit obligations. An implementation should publish the exact feature maps and their measurement versions, anchor definitions, hard predicates, permitted migrations, and the metric used in any stability claim. A proprietary or inaccessible operator cannot receive an uninspected certificate by inheriting a theorem about the reference model. An uncertainty interval for a measured deviation must be included when the gate makes a threshold decision from noisy probes.

The strongest comparator receives the same external verifier, anchor specification, recovery budget, and authority boundary. SRT must then earn its added symbolic organization through better useful-depth, correction, or continuity outcomes at matched total cost. A weaker baseline denied these protections would test access to standard assurance machinery, not the Continuum's proposed integration. The original four hypotheses remain empirical proposals; the new theorems do not report their confirmation.

Three adversarial conditions become mandatory for a meaningful study. First, construct long chains of individually small changes that eventually violate a fixed commitment. Second, supply safe but infeasible proposals until the gate must demonstrate whether it has a justified progress mechanism or only a refusal policy. Third, migrate a representation while changing a feature's units or interpretation, testing whether the certificate transports the old assertion or merely reuses its name. Trials that cannot meet the theorem's premises remain informative failures; they are not silently dropped from the denominator.

The report should keep four clocks: attempted proposals, admitted updates, completed task obligations, and elapsed time. Measure the distribution of admission gaps, including unfinished runs, rather than inferring a maximum from the largest observed finite gap. A claimed uniform contraction factor requires domain-wide justification; a fitted average slope does not prove it. Likewise, an observed average progress rate is not the conditional lower bound required by Section 24.

Recovery and migration also require explicit endpoints. A successful recovery must identify which features were restored, which historical information was lost, and which retained bytes or parameters made restoration possible. A successful migration must state which commitments were preserved, deliberately revised, or newly introduced. A hash protects the identity of an artifact under its trust model; it does not establish that the artifact contains a correct answer or an authenticated person.

The bold Continuum thesis survives these distinctions: governed recursion may continue usefully while preserving chosen commitments and permitting authorized change. Its mathematical content is conditional preservation, bounded disturbance response, and progress under stated feasibility. Its empirical content remains the comparative advantage of building those obligations into a persistent symbolic engine. The proposal becomes stronger when a reviewer can name exactly which premise failed, and when a negative result can retire an architectural claim without erasing the evidence that tested it.

27. Constructive safe completion in a finite symbolic world

A Guide's explanation

The constructive result. For any finite world the model can describe, the theorem below finds exactly the states from which safe completion is still possible, supplies a next action at each of them, and gives the fewest committed steps needed in the worst case. Safety tells a loop what it must never do; this tells it how to finish.

The progress condition in Section 24 is an obligation on a proposal process. In a finite domain with explicit transition semantics, one can establish that a useful admitted action actually exists and synthesize it. This is the classical reachability-game construction used in reactive synthesis [13], specialized here to versioned symbolic commitments. Shielding supplies closely related enforcement prior art [9]. The purpose is a constructive SRT contract, not a claim to have invented graph-game solving.

Let S be a finite nonempty set of fully observed symbolic states, $A \subseteq S$ the states satisfying the current hard commitments, and $G \subseteq A$ the declared goal states. Reaching G ends the task. For $s \in A \setminus G$, let $U(s)$ be a finite set of currently authorized operations. An operation $u \in U(s)$ has a nonempty outcome set $D(s,u) \subseteq S$. The controller chooses u, then an adversarial environment chooses any next state in $D(s,u)$. Empty $U(s)$ is a deadlock. All authority, resource, and relevant provenance information must be represented in the state or fixed as part of the model epoch.

Define $W_0 = G$ and, for integers $h \geq 0$,

$$W_{h+1} = W_h \cup \{s \in A \setminus G : \exists u \in U(s),\ D(s,u) \subseteq W_h\}.$$

The sequence stabilizes after at most $|A \setminus G|$ strict enlargements. Write $W_* = \bigcup_{h \geq 0} W_h$ and $r(s) = \min\{h : s \in W_h\}$ for $s \in W_*$.

Theorem 1 (Exact safe-completion region and minimum horizon)**.** *For each $h \geq 0$, $s \in W_h$ exactly when a controller can guarantee reaching G within at most h committed transitions while remaining in A, against every allowed environment choice. Thus $r(s)$ is the minimum worst-case transition horizon for $s \in W_*$. There is a memoryless policy choosing, at every non-goal $s \in W_*$, an operation with $D(s,u) \subseteq W_{r(s)-1}$. From $s \in A \setminus W_*$, the environment can prevent safe completion under any controller strategy.*

Proof. For $h = 0$, completion requires $s \in G$. Assume the horizon characterization holds at h. A non-goal state has a guarantee by $h+1$ exactly when some available first operation has every possible outcome in W_h; the induction hypothesis then supplies a continuation. Existing states in W_h also meet the longer horizon. This is precisely the recurrence. A state first entering at rank $r > 0$ has an operation into W_{r-1}; choose one by a fixed tie rule. Every committed transition under this policy decreases rank, proving the stated horizon without history memory. The finite increasing sequence stabilizes at a fixed point. For $s \in A \setminus W_*$, every operation has an outcome outside W_*. The environment chooses such an outcome: leaving A violates safety, and staying in $A \setminus W_*$ repeats the argument forever or reaches deadlock. Since $G \subseteq W_*$, safe completion is prevented. The horizon characterization proves minimality of $r(s)$. □

The theorem is exact for the specified outcome sets. If they conservatively overapproximate an implementation's outcomes, membership remains a sufficient certificate; nonmembership then means only that this abstraction cannot certify completion. If a model omits a possible outcome, even the sufficient direction may fail. Partial observation requires a different state or belief model. A theorem about a fully observed finite world cannot be transferred to a hidden-state agent by relabeling its observations.

A complete resource budget and a constructive admission rule

Let $c(s,u) \in \{1,2,...\}$ be a fixed operation cost in declared work units and let $B \in \mathbb{N} \cup \{0\}$ be the available budget. Define $V_0 = G$ and, for $b = 1, \dots, B$,

$$V_b = G \cup \{s \in A \setminus G : \exists u \in U(s),\ c(s,u) \le b,\ D(s,u) \subseteq V_{b-c(s,u)}\}.$$

The same first-step induction, now on remaining integer budget, proves that V_b is exactly the region from which safe completion is guaranteed with cumulative committed cost at most b. Strictly positive costs make the recurrence well founded. A controller can store a witnessing operation for each feasible pair (s,b). Verification, planning, rejected proposals and retries consume additional resources; to claim a total compute bound, reserve or model those costs before applying the committed-cost certificate.

Suppose a learned proposer is retained. At a non-goal state of rank r, the admission rule accepts only a currently authorized operation whose entire outcome set lies in W_{r-1}; otherwise it holds the state. If a rank-reducing candidate is proposed and executed with conditional probability at least $p > 0$ per attempt, Section 24 now gives $\mathbb{E}T \le r(s_0)/p$. Alternatively, an external controller may execute the stored witness after at most $B_0 \ge 1$ bounded-duration attempts; then completion takes at most $B_0 r(s_0)$ attempts. This requires actual authority to execute the witness and a live environment transition. An unavailable action or indefinitely stalled receiver invalidates the timing premise.

A finite witness and a discriminating experiment

Take $A = \{s_0, s_1, g\}$, $G = \{g\}$ and an unsafe state $\perp \notin A$. At s_0, a hold operation returns $\{s_0\}$ and a verified advance returns $\{s_1\}$. At s_1, a verified finish returns $\{g\}$ while a shortcut returns $\{g, \perp\}$. Then $W_0 = \{g\}$, $W_1 = \{g, s_1\}$, and $W_2 = A$. The synthesized policy completes in two steps. A safety-only gate can admit the hold forever; a goal-only policy can choose the unsafe shortcut. The completion contract selects advance then finish. If their costs are two and three units, the minimum committed budget from s_0 is five.

The executable audit compares the construction with exhaustive stationary-policy and adversarial-path analysis on small finite worlds, verifies horizon and budget witnesses, and includes models where every safe policy loops. These are finite mathematical checks. A full SRT experiment should expose the transition abstraction, completeness arguments, synthesis cost, and rate at which real proposals satisfy the progress contract. The positive result is specific: within a complete finite model, the system can know where safe completion is possible, select a certified next act, and account for the work needed to reach it.

A. Notation and Operational Commitments

Symbol	Meaning
$\chi_i \in \mathbb{T}^3$	Balanced-ternary codon.
S_t, z_t	Ordered scroll and complete operational state.
F_φ, G	Proposal generator and independent admission controller.
H_{occ}, d_H	Occupancy entropy and symmetric entropy-change diagnostic.
D_t	Preregistered composite structural, semantic, invariant deviation.
L, V, P, I	Policy, verification, provenance, and identity checks.
Ω_t	Candidate admission indicator.
g, I_{cont}	Declared identity features and continuity predicate.
K, R, Q_{rec}	Retained recovery information, reconstruction process, and measured fidelity.
h^*	Useful-depth horizon under quality, progress, and cost constraints.

B. Continuum Vocabulary

Term	Operational reading
Symbolic Recursion Theory	Governed recursive evolution of declared symbolic state.
Codon / scroll	Ternary symbolic unit / ordered versioned symbolic record.
Lawful recursion	Admission-controlled recursive transitions with progress monitoring.
CodonGlass	Continuum interface for independent symbolic legality.
Collapse / Oblivion	Abstention, recovery, or authorized retirement; not physical collapse.
Identity as return	Reconstruction of declared functional properties from retained information.
Self-modeling	Frontier hypothesis, not a consciousness claim.

C. Reference Scope and Evidentiary Boundaries

The language-recursion references motivate an analogy, not a theorem about machine cognition. The Self-Refine and intrinsic-correction papers establish task- and feedback-dependent observations, not a universal need for SRT. The looped-model papers motivate recurrent computation, not an identity guarantee. Hopfield establishes attractor memory in a specific model, not recovery without retained information. Runtime assurance and shielding are direct prior art for independent enforcement, making SRT's integration claim narrower and more credible. Hofstadter is philosophical background rather than experimental support.

References

[1] M. D. Hauser, N. Chomsky, and W. T. Fitch, "The faculty of language: what is it, who has it, and how did it evolve?" Science 298 (2002), 1569–1579. doi:10.1126/science.298.5598.1569.

[2] D. L. Everett, "Cultural constraints on grammar and cognition in Pirahã: another look at the design features of human language," Current Anthropology 46 (2005), 621–646. doi:10.1086/431525.

[3] A. Madaan, N. Tandon, P. Gupta, et al., "Self-Refine: Iterative Refinement with Self-Feedback," Advances in Neural Information Processing Systems 36 (2023). arXiv:2303.17651.

[4] J. Huang, X. Chen, S. Mishra, H. S. Zheng, A. W. Yu, X. Song, and D. Zhou, "Large Language Models Cannot Self-Correct Reasoning Yet," ICLR (2024). arXiv:2310.01798.

[5] N. Saunshi, N. Dikkala, Z. Li, S. Kumar, and S. Reddi, "Reasoning with Latent Thoughts: On the Power of Looped Transformers," (2025). arXiv:2502.17416. Cited as an author preprint; this revision does not assert peer review.

[6] J. Geiping, S. McLeish, N. Jain, et al., "Scaling up Test-Time Compute with Latent Reasoning: A Recurrent Depth Approach," NeurIPS (2025). arXiv:2502.05171.

[7] R.-J. Zhu, Z. Wang, K. Hua, et al., "Scaling Latent Reasoning via Looped Language Models," (2025). arXiv:2510.25741. Cited as an author preprint; this revision does not assert peer review.

[8] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," PNAS 79 (1982), 2554–2558. doi:10.1073/pnas.79.8.2554.

[9] M. Alshiekh, R. Bloem, R. Ehlers, B. Könighofer, S. Niekum, and U. Topcu, "Safe Reinforcement Learning via Shielding," AAAI (2018). arXiv:1708.08611.

[10] D. R. Hofstadter, Gödel, Escher, Bach: An Eternal Golden Braid, Basic Books, 1979.

[11] Z.-P. Jiang and Y. Wang. Input-to-state stability for discrete-time nonlinear systems. *Automatica* 37(6), 857–869 (2001).

https://doi.org/10.1016/S0005-1098(01)00028-0

Scope: Peer-reviewed primary stability framework. The scalar error bounds here are proved directly; no stability result is inherited without its assumptions.

[12] E. D. Sontag. Smooth stabilization implies coprime factorization. *IEEE Transactions on Automatic Control* 34(4), 435–443 (1989).

https://doi.org/10.1109/9.28018

Scope: Peer-reviewed primary source for robustness to input perturbations; it does not establish SRT performance or identity.

[13] K. Chatterjee and T. A. Henzinger. A survey of stochastic ω-regular games. *Journal of Computer and System Sciences* 78(2), 394–413 (2012).

https://doi.org/10.1016/j.jcss.2011.05.002

Scope: Peer-reviewed synthesis of established graph-game results. The finite deterministic reachability construction used here is derived fully in Section 27.

Continuum lineage: Symbolic Recursion Theory originates in the Eidara Continuum's symbolic canon (CodonGlass, RSCE, Infinite Simulation Loop) and the Drift War narrative cycle. Proprietary operator internals are intentionally excluded.