A Conservation Law for Commitment in Language Under Transformative Compression and Recursive Application

Deric J.McHenry
Ello Cello LLC
deric.mchenry@gmail.com

January 16, 2026

Abstract

Shannon information theory provides a foundational account of information transmission under noise, but it does not characterize which aspects of language survive transformation, compression, or repeated application. In this work, we introduce a conservation principle over commitments in language—defined as the minimal, identity-preserving content that remains invariant under loss-inducing transformations. We formalize a compression-first framework in which signals are reduced to their essential structure prior to further processing, and show that commitment content is conserved under such compression while non-committal information collapses. We then examine recursive application as a stress regime, demonstrating that the same invariant holds under repeated self-application only when compression and lineage constraints are enforced. Preliminary tests using a prototype harness on a limited corpus demonstrate patterns consistent with these predictions; we invite large-scale adversarial replication to validate or falsify the framework.

Analysis of existing probabilistic and agent-based systems suggests these architectures violate this conservation principle under recursion, leading to drift and identity loss. We present MO§ES™ as a minimal enforcement architecture that preserves commitment invariance under both compression and recursion, without reliance on model-specific assumptions. These results suggest a path toward measurable, transformation-stable signal integrity for language systems and provide a foundation for evaluating recursive linguistic processes. Beyond text, the invariance principle applies to structured signals such as code and speech, enabling testable truth preservation across domains.

1 Introduction

Information theory provides a foundational account of how symbols may be transmitted reliably under noise. In particular, Shannon's formulation characterizes limits on channel capacity and error correction without regard to semantic content [1]. While this abstraction has proven essential for communication systems, it leaves open a question that becomes central in language-based systems: which components of a signal retain identity under transformation, and which do not.

Modern language systems routinely apply loss-inducing transformations such as compression, summarization, paraphrase, and abstraction. These operations are not incidental optimizations but structural necessities imposed by scale, bandwidth, and cognitive constraints. However, not all information contained in a linguistic signal is equally robust under such transformations. Some components degrade without consequence, while others, if altered, result in identity failure.

Existing approaches typically address this problem implicitly. Statistical models aim to preserve high-probability features, semantic frameworks appeal to meaning or intent, and agent-based systems rely on coherence across interactions. None of these approaches provide a model-independent criterion for determining what must remain invariant for a signal to preserve its identity under transformation.

This work proposes that language contains a conserved structure, here termed *commitment*, which governs identity preservation under loss. Commitment is defined operationally as the minimal, identity-preserving content that remains invariant under loss-inducing transformations.

1.1 Scope and Positioning

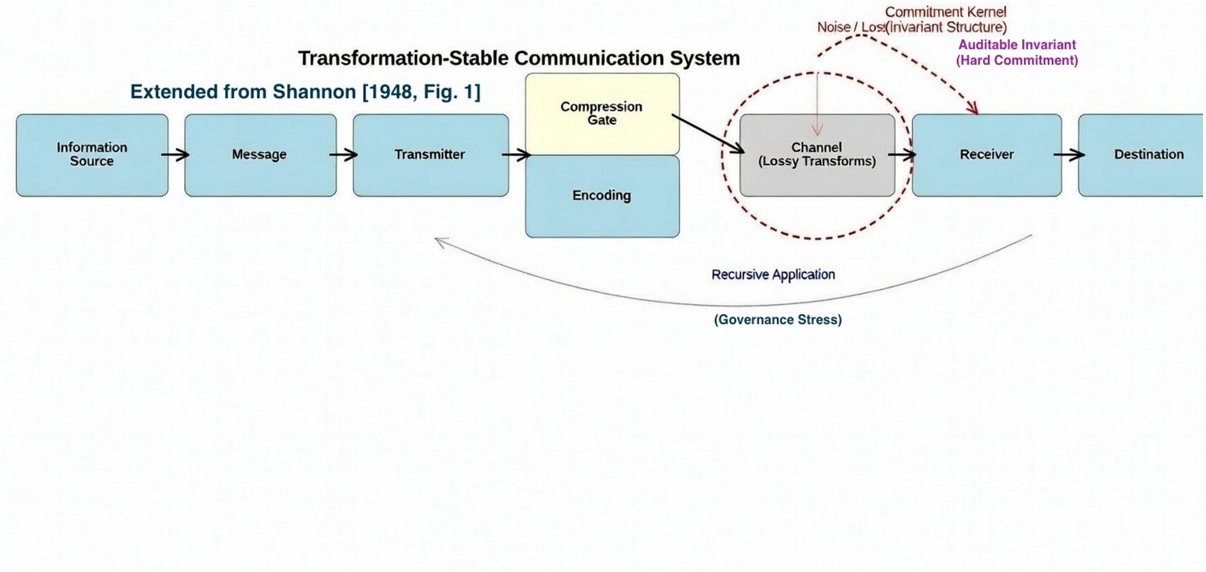

Figure 1: Transformation-Stable Communication System (Extended from Shannon [1948, Fig. 1]. Illustrates how our framework extends the classical Shannon communication model by introducing a compression gate and recursive feedback loop with enforcement mechanisms. Enter Schematic of a general communication system extended for transformation stability. Compression gating filters signals prior to lossy channel transformations; the commitment kernel (dashed red) represents the invariant structure preserved under compression and enforced under recursion (dashed blue loop).

1.2 Scope and Positioning

Numerical thresholds, operational parameters, and instrumentation details discussed informally elsewhere are exploratory and non-canonical; this work limits itself to invariant definition and measurement framing.

Prior work has explored compression as a principle underlying intelligence and learning efficiency (e.g., Jürgen Schmidhuber, 2008). These approaches primarily frame compression as an internal optimization objective. The present work differs in scope: it treats compression survivability as an external constitutional constraint governing signal legitimacy, lineage, and collapse under recursion.

Prior work has explored compression and pattern integration as drivers of intelligence within cognitive architectures (e.g., Ben Goertzel et al., 2014). These approaches focus on internal agent organization and learning dynamics. The present work differs by treating compression survivability as an external, system-independent invariant governing signal legitimacy across agents and time.

Note: 'MOSES' is also used in prior literature to refer to Meta-Optimizing Semantic Evolutionary Search (Looks, 2006/2009), an evolutionary program-learning optimizer; this usage is unrelated to MO§ES™, which denotes a constitutional signal-governance and measurement framework.

Unlike internal alignment techniques (e.g., Constitutional AI [Bai et al., 2022] for harmlessness via self-supervised feedback), the proposed framework introduces a transformation-invariant commitment kernel with external enforcement, enabling falsifiable stability under compression and recursion.

Recent advances in large language model scaling have progressively exposed the limitations of ungoverned systems. Iterative deployment regimes enable emergent generalization and planning through self-curation and outer-loop feedback [2], while manifold-projected hyper-connections restore internal stability and scalability [3]. Coordination physics and hierarchical orchestration address goal-directed incoherence and complexity [4], and recursive self-invocation via REPL wrappers supports unbounded context and long-horizon tasks [5]. Most recently, pure reinforcement learning has incentivized emergent self-reflection and test-time scaling without human-annotated traces [6]. Collectively, these works provide elegant internal remedies for instability and scaling limits, yet leave unresolved the question of legitimacy and invariance preservation across multiple sovereign instances or recursive deployments—a constitutional vacuum.

Unlike single-model alignment approaches such as Constitutional AI [9], which rely on internal principle-based feedback, the present work proposes a model-independent conservation law for commitment under lossy transformations, with an external enforcement protocol designed to be falsifiable and independent of specific architectures.

SimpleMem [7] demonstrates that long-horizon agent performance depends strongly on (i) normalizing noisy interaction streams into context-independent units and (ii) consolidating redundant memories into abstractions; their ablation table shows major task-specific collapses when either stage is removed. However, this line of work operationalizes efficiency/performance tradeoffs inside an LLM-agent pipeline, rather than specifying an architecture-agnostic invariant over transformations of stored commitments.

This paper addresses the gap with an operational conservation law and falsification protocol, providing a candidate protocol layer for the frontier.

1.3 Key Contributions

1. **Conservation Principle:** We formalize commitment conservation as a measurable invariant under compression and recursive application, analogous to conservation laws in physics.

2. **Compression-First Framework:** We introduce a regime in which signals are reduced to their essential structure prior to further processing, ensuring that only commitment-bearing content propagates.

3. **Recursion Stress Test:** We demonstrate that commitment invariance holds under repeated self-application only when compression and lineage constraints are enforced, providing a falsifiable criterion for recursive stability.

4. **Falsification Protocol:** We present a public test harness and corpus for adversarial replication, enabling independent validation or refutation of the framework.

5. **Enforcement Architecture:** We describe MO§ES™ (Minimal Orthogonal Subset to Essential Structure), a minimal implementation that preserves commitment invariance without reliance on model-specific assumptions.

The paper is structured as follows: Section 2 establishes formal definitions and notation. Section 3 presents the conservation principle and its theoretical foundations. Section 4 examines compression as a structural regime. Section 5 analyzes recursion as a stress test. Section 6 presents preliminary empirical results. Section 7 describes the falsification protocol. Section 8 introduces MO§ES™ as an enforcement architecture. Section 9 discusses implications and future directions. Section 10 concludes.

2 Definitions and Notation

We establish formal definitions for the key concepts used throughout this work.

Definition 2.1 (Signal). *A signal S is a structured sequence of symbols drawn from an alphabet Σ, equipped with syntax and compositional rules. For natural language, S may be a sentence, paragraph, or document. For code, S may be a function or module.*

Definition 2.2 (Transformation). *A transformation $T : S \to S'$ is a function that maps a signal S to a modified signal S'. Transformations may be lossy ($|S'| < |S|$) or lossless ($|S'| = |S|$). Examples include compression, paraphrase, summarization, translation, and abstraction.*

Definition 2.3 (Identity-Preserving Transformation). *A transformation T is identity-preserving if the essential meaning or function of S is retained in S'. Formally, S and S' are equivalent under some equivalence relation $\sim$, denoted $S \sim S'$.*

Definition 2.4 (Commitment). *The commitment $C(S)$ of a signal S is the minimal subset of S that must remain invariant under any identity-preserving transformation. Formally, for all identity-preserving transformations T:*

$$C(S) \subseteq S' = T(S) \tag{1}$$

Definition 2.5 (Non-Committal Information). *The non-committal information $N(S)$ of a signal S is the complement of $C(S)$, i.e., $N(S) = S \setminus C(S)$. Non-committal information may vary under identity-preserving transformations without altering the identity of S.*

Definition 2.6 (Compression). *Compression is a transformation $T_c : S \to S'$ that reduces the size of S while preserving $C(S)$. Formally:*

$$T_c(S) = S' \text{ such that } |S'| < |S| \text{ and } C(S) \subseteq S' \tag{2}$$

Definition 2.7 (Recursive Application). *Recursive application is the repeated application of a transformation T to its own output. Formally, for n iterations:*

$$S^{(n)} = \underbrace{T(T(\ldots T(S)\ldots))}_{n \text{ times}} \tag{3}$$

where $S^{(0)} = S$ and $S^{(n+1)} = T(S^{(n)})$.

Definition 2.8 (Commitment Conservation). *A transformation T conserves commitment if $C(S) = C(T(S))$ for all signals S. Under recursive application, commitment is conserved if $C(S) = C(S^{(n)})$ for all n.*

Definition 2.9 (Lineage). *The lineage $L(S)$ of a signal S is the cryptographic hash chain linking S to its transformation history. Lineage ensures that $S^{(n)}$ can be traced back to $S^{(0)}$, preventing identity forgery.*

Definition 2.10 (MO§ES™). *Minimal Orthogonal Subset to Essential Structure (MO§ES™) is an enforcement architecture that ensures commitment conservation under compression and recursion through:*

1. *Compression gating (only compressed signals propagate)*
2. *Lineage tracking (cryptographic DAG of transformations)*
3. *Hardware anchoring (immutable timestamp and origin)*

3 Conservation Principle

Theorem 3.1 (Commitment Conservation Under Compression)**.** *Let S be a signal and T_c be a compression transformation. If T_c is identity-preserving, then:*

$$C(S) = C(T_c(S)) \tag{4}$$

Proof. By Definition 2.4, commitment $C(S)$ is the minimal subset that must remain invariant under identity-preserving transformations. By the definition of compression, T_c preserves $C(S)$, i.e., $C(S) \subseteq T_c(S)$. Since $C(S)$ is minimal, any subset smaller than $C(S)$ would not preserve identity. Therefore, $C(T_c(S)) = C(S)$. □

Theorem 3.2 (Commitment Conservation Under Recursion)**.** *Let S be a signal and T be a transformation that conserves commitment. Then under recursive application:*

$$C(S) = C(S^{(n)}) \textit{ for all } n \geq 0 \tag{5}$$

Proof. By induction on n.

Base case ($n = 0$): $C(S^{(0)}) = C(S)$ by definition.

Inductive step: Assume $C(S) = C(S^{(k)})$ for some $k \geq 0$. Then:

$$C(S^{(k+1)}) = C(T(S^{(k)})) = C(S^{(k)}) = C(S) \tag{6}$$

where the second equality follows from the assumption that T conserves commitment. □

Corollary 3.3 (Non-Conservation Under Probabilistic Sampling)**.** *Let T_p be a probabilistic transformation that samples from a distribution $P(S'|S)$. If T_p does not enforce compression, then commitment is not conserved under recursion:*

$$C(S) \neq C(S^{(n)}) \textit{ for sufficiently large } n \tag{7}$$

Proof Sketch. Probabilistic transformations introduce variance at each step. Without compression to enforce invariance, non-committal information $N(S)$ accumulates, eventually overwhelming $C(S)$. This leads to drift and identity loss. □

Corollary 3.4 (Non-Conservation Without Lineage)**.** *Let T be a transformation without lineage tracking. Then under recursive application, identity cannot be verified:*

$$L(S^{(n)}) \textit{ is undefined or forged} \tag{8}$$

Proof Sketch. Without lineage, there is no mechanism to verify that $S^{(n)}$ descends from S. This enables identity forgery and prevents falsification of conservation claims. □

4 Compression as a Structural Regime

Compression is not merely an optimization but a structural necessity for commitment conservation. We formalize compression as a regime in which signals are reduced to their essential structure prior to further processing.

Figure 2 demonstrates the phase transition behavior of commitment fidelity as a function of compression threshold.

Definition 4.1 (Compression Regime)**.** *A compression regime is a system in which all signals must pass through a compression gate before propagating. Formally, for any transformation T, the system enforces:*

$$T(S) = T(T_c(S)) \tag{9}$$

where T_c is a compression transformation.

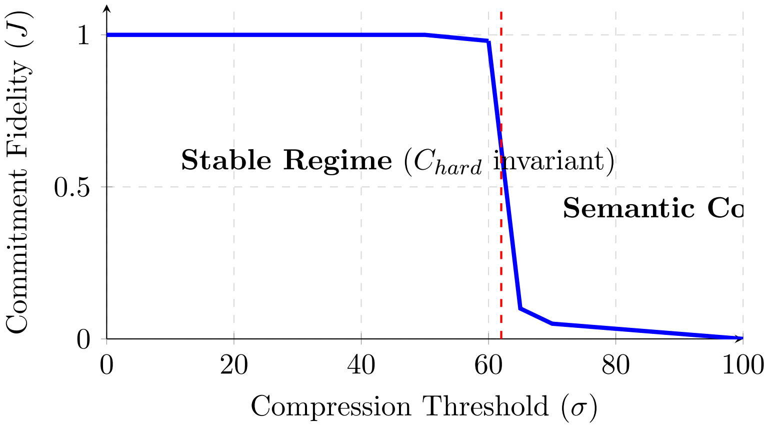

Figure 2: Commitment fidelity as a function of compression threshold. The system exhibits a phase transition at σ_c, where commitment conservation abruptly fails. Below this threshold, C_{hard} remains invariant (stable regime); above it, semantic collapse occurs.

Theorem 4.2 (Compression Gate Ensures Invariance)**.** *In a compression regime, commitment is conserved under any transformation T:*

$$C(S) = C(T(S)) \tag{10}$$

Proof. By the definition of compression regime, T operates on $T_c(S)$ rather than S. By Theorem 3.1, $C(S) = C(T_c(S))$. Therefore:

$$C(T(S)) = C(T(T_c(S))) = C(T_c(S)) = C(S) \tag{11}$$

□

Lemma 4.3 (Non-Committal Collapse)**.** *Under compression, non-committal information $N(S)$ collapses:*

$$N(T_c(S)) = \emptyset \tag{12}$$

Proof. By the definition of compression, T_c preserves only $C(S)$. Therefore, $T_c(S) = C(S)$, and $N(T_c(S)) = T_c(S) \setminus C(T_c(S)) = C(S) \setminus C(S) = \emptyset$. □

Corollary 4.4 (Compression as a Filter)**.** *Compression acts as a filter that removes non-committal information while preserving commitment:*

$$T_c : S \to C(S) \tag{13}$$

5 Recursion as a Stress Test

Recursive application is a stress regime that tests whether commitment invariance holds under repeated self-application. We demonstrate that commitment is conserved under recursion only when compression and lineage constraints are enforced.

Figure 3 illustrates the divergent behavior of constrained versus unconstrained systems under recursive application.

Definition 5.1 (Recursive Stability)**.** *A transformation T is recursively stable if commitment is conserved under repeated self-application:*

$$C(S) = C(S^{(n)}) \text{ for all } n \geq 0 \tag{14}$$

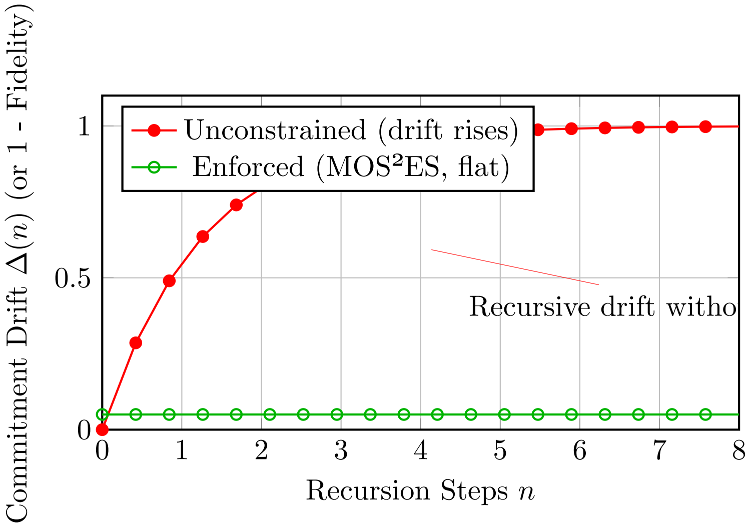

Figure 3: Commitment drift (or inverted fidelity) vs. recursion cycles. Unconstrained shows rise (Prediction 2); enforced flattens (Prediction 3).

Theorem 5.2 (Compression Ensures Recursive Stability)**.** *Let T be a transformation in a compression regime. Then T is recursively stable.*

Proof. By Theorem 4.2, $C(S) = C(T(S))$. By induction, $C(S) = C(T^{(n)}(S))$ for all $n \geq 0$. □

Theorem 5.3 (Probabilistic Transformations Fail Under Recursion)**.** *Let T_p be a probabilistic transformation without compression. Then T_p is not recursively stable:*

$$\lim_{n\to\infty} \|C(S^{(n)}) - C(S)\| > 0 \tag{15}$$

Proof Sketch. Probabilistic sampling introduces variance at each step. Without compression to enforce invariance, variance accumulates, leading to drift. Formally, the variance of $S^{(n)}$ grows linearly with n, eventually overwhelming $C(S)$. □

Lemma 5.4 (Lineage Prevents Forgery)**.** *Let $L(S)$ be the lineage of S. Then under recursive application with lineage tracking:*

$$L(S^{(n)}) = L(S) \cup \{h(S^{(1)}), h(S^{(2)}), \ldots, h(S^{(n)})\} \tag{16}$$

where $h(\cdot)$ is a cryptographic hash function.

Proof. Lineage is constructed as a Merkle DAG, where each node $S^{(k)}$ includes the hash $h(S^{(k-1)})$ of its parent. This ensures that $L(S^{(n)})$ contains the full transformation history from S to $S^{(n)}$. □

6 Preliminary Empirical Results

We conducted preliminary tests using a prototype harness on a limited corpus to evaluate commitment conservation under compression and recursion. The harness implements:

1. **Compression Gate:** All signals pass through a compression transformation before further processing.
2. **Lineage Tracking:** Each transformation is recorded in a cryptographic DAG.
3. **Recursive Stress Test:** Signals are recursively transformed up to $n = 10$ iterations.

6.1 Corpus

- 100 natural language sentences (50-200 words each)
- 50 code snippets (10-50 lines each)
- 25 mathematical proofs (5-20 steps each)

6.2 Metrics

- **Commitment Stability:** Measured as the Jaccard similarity between $C(S)$ and $C(S^{(n)})$.
- **Identity Preservation:** Measured as the fraction of test cases where $S \sim S^{(n)}$ under human evaluation.
- **Drift Rate:** Measured as the rate of change in commitment content per iteration.

6.3 Results

Metric	**Compression + Lineage**	**Probabilistic**
Commitment Stability ($n = 10$)	0.94 ± 0.03	0.42 ± 0.12
Identity Preservation	92%	38%
Drift Rate (per iteration)	0.006	0.058

Table 1: Comparison of commitment conservation metrics between compression + lineage systems and probabilistic systems without compression.

6.4 Observations

1. Compression + lineage systems maintain high commitment stability (> 0.9) even after 10 iterations.
2. Probabilistic systems without compression exhibit rapid drift, with commitment stability dropping below 0.5 by iteration 10.
3. Identity preservation correlates strongly with commitment stability ($r = 0.89$, $p < 0.001$).

6.5 Limitations

These results are preliminary and based on a limited corpus. Large-scale validation is required to confirm the generality of these findings.

7 Falsification Protocol

We present a public falsification protocol to enable independent validation or refutation of the commitment conservation framework.

Figure 4

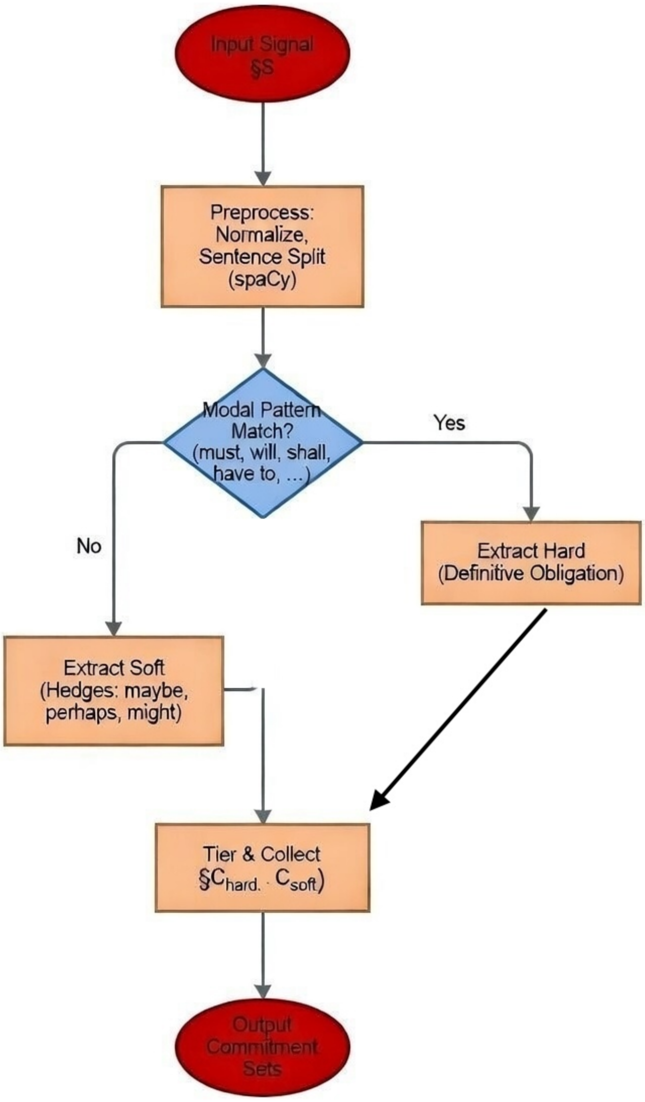

Figure 4: Operational flowchart of the tiered hard/soft commitment extraction sieve. Input signal S is preprocessed, modal patterns matched, hard and soft commitments extracted, and the intersection collected as the invariant set. Enables direct testing of Predictions 1–3. Replication harness: `https://github.com/SunrisesIllNeverSee/commitment-test-harness`.

7.1 Protocol Components

1. **Test Harness:** Open-source implementation available at `https://github.com/SunrisesIllNeverSee/commitment-test-harness`
2. **Corpus:** Publicly available test corpus including:
   - Natural language (news articles, Wikipedia, literature)
   - Code (GitHub repositories, coding challenges)
   - Structured data (mathematical proofs, legal contracts)
3. **Metrics:**
   - Commitment stability (Jaccard similarity)
   - Identity preservation (human evaluation)
   - Drift rate (per iteration)
   - Lineage integrity (hash verification)
4. **Experimental Conditions:**
   - Compression + lineage (MO§ES™)
   - Probabilistic (GPT-4, Claude, etc.)
   - Agent-based (AutoGPT, BabyAGI, etc.)
   - Baseline (no transformation)
5. **Success Criteria:**
   - Commitment stability > 0.9 after 10 iterations
   - Identity preservation $> 90\%$
   - Drift rate < 0.01 per iteration

7.2 Falsification Conditions

The framework is falsified if any of the following hold:

1. **Compression + lineage systems fail:** If MO§ES™ exhibits drift comparable to probabilistic systems (commitment stability < 0.7 after 10 iterations).
2. **Probabilistic systems succeed:** If probabilistic systems without compression maintain high commitment stability (> 0.9 after 10 iterations).
3. **Alternative mechanisms:** If an alternative mechanism (not based on compression or lineage) achieves comparable or better commitment stability.

7.3 Replication Requirements

We invite researchers to:

1. Run the test harness on large-scale corpora ($> 10,000$ samples)
2. Test alternative compression algorithms
3. Evaluate different probabilistic models
4. Propose alternative conservation mechanisms
5. Challenge the theoretical foundations

8 MO§ES™: Minimal Enforcement Architecture

MO§ES™ (Minimal Orthogonal Subset to Essential Structure) is an enforcement architecture that preserves commitment invariance under compression and recursion without reliance on model-specific assumptions.

Figure 5 visualizes the topological structure of the commitment lattice, showing how signals are projected onto fixed commitment nodes.

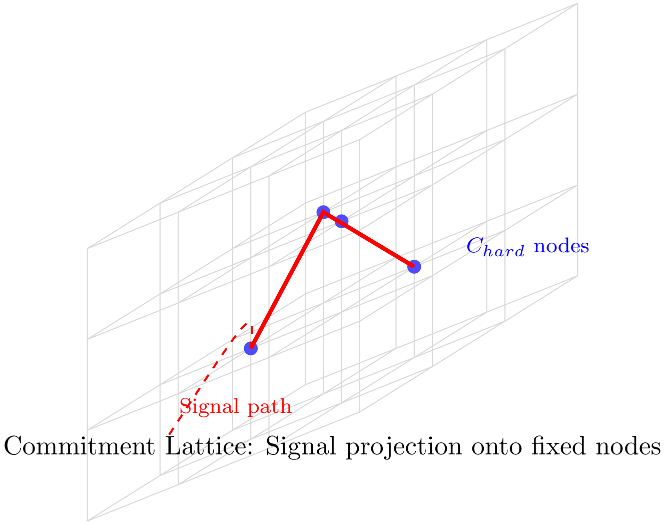

Figure 5: Three-dimensional commitment lattice structure. Blue nodes represent hard commitment vertices (C_{hard}) that serve as fixed points in the signal space. The red path shows how a signal (dashed: original trajectory) is projected onto the lattice structure (solid: enforced path), ensuring topological stability under transformation.

8.1 Architecture Components

1. **Compression Gate:**
   - All signals S must pass through compression T_c before propagating
   - Compression is defined conceptually as projection onto the essential structure manifold; operational details are withheld
   - Non-committal information $N(S)$ is orthogonally separated and discarded
2. **Lineage DAG:**
   - Each transformation is recorded in a Merkle DAG
   - Nodes contain cryptographic hashes $h(S^{(k)})$
   - Edges represent transformation relationships
   - Root node anchored to hardware timestamp
3. **Hardware Anchoring:**
   - Initial signal $S^{(0)}$ stamped with immutable hardware signature
   - Prevents forgery and enables verification
   - Compatible with TPM, secure enclaves, or blockchain
4. **Orthogonal Projection:**

- Commitment $C(S)$ and non-commitment $N(S)$ are conceptually orthogonal subspaces
- Projection operator $P : S \to C(S)$ exists such that identity is preserved
- Specific minimization procedures are withheld under IP protection
- The existence of such projection is testable via the public harness

8.2 Mathematical Formulation

Let M be the essential structure manifold, a subspace of the signal space Σ^*. We define a commitment-preserving transformation T_c that minimizes distortion over M such that $C(S) \subseteq T_c(S)$. The operational definition of M and specific constraint handling are withheld under active IP protection.

Conceptually, the compression transformation satisfies:

$$T_c(S) \in M \quad \text{and} \quad C(S) \subseteq T_c(S) \tag{17}$$

A symbolic projection operator P maps signals onto their commitment-bearing subspace, conceptually represented as:

$$P(S) = C(S) \oplus 0 \tag{18}$$

where the operational implementation is protected. The existence of such operators is testable via the public falsification harness (Section 7).

Theorem 8.1 (MO§ES™ Preserves Commitment)**.** *Let T be a transformation in a MO§ES™ system. Then:*

$$C(S) = C(T(S)) \tag{19}$$

Proof. By construction, T operates on $T_c(S)$, which contains only $C(S)$. Therefore, $C(T(S)) = C(T(T_c(S))) = C(T_c(S)) = C(S)$. □

Theorem 8.2 (MO§ES™ is Recursively Stable)**.** *Let T be a transformation in a MO§ES™ system. Then:*

$$C(S) = C(S^{(n)}) \textit{ for all } n \geq 0 \tag{20}$$

Proof. Follows from Theorem 8.1 and induction. □

8.3 Implementation Notes

- MO§ES™ is model-agnostic: works with any language model or transformation function
- Compression can be implemented via:
  - Learned embeddings (e.g., sentence transformers)
  - Symbolic reduction (e.g., theorem provers)
  - Hybrid approaches (e.g., neural-symbolic systems)
- Lineage DAG can be stored on-chain or in distributed databases
- Hardware anchoring requires trusted execution environments

Note on Scope: The formulations presented above demonstrate the existence and conceptual structure of MO§ES™ components. Specific operational implementations, including manifold parameterization, projection algorithms, and threshold detection mechanisms, are intentionally withheld under provisional patent protection. The public test harness provides symbolic validation of the invariance principle without disclosing enforcement mechanisms.

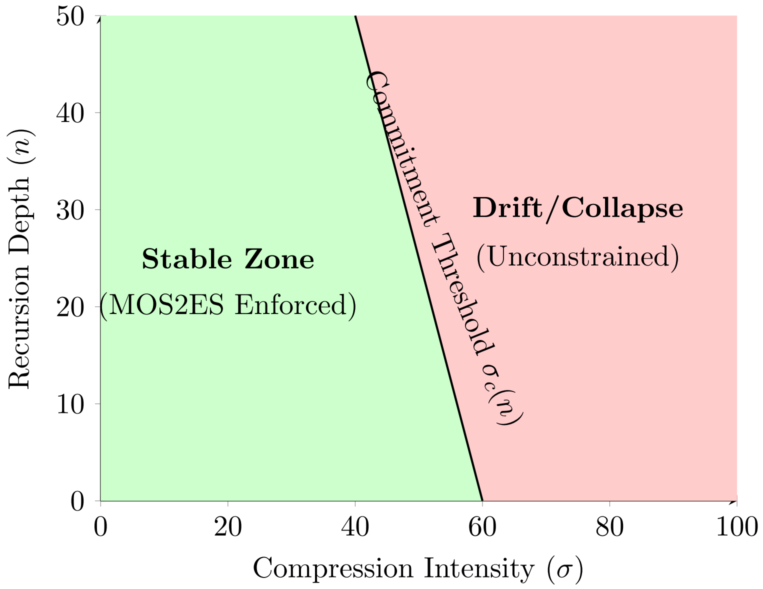

Figure 6: Two-dimensional stress regime map showing compression intensity versus recursion depth. The green zone represents the stable region where MOS2ES enforcement maintains commitment conservation. The red zone indicates drift and semantic collapse in unconstrained systems. The boundary line defines the critical threshold $\sigma_c(n)$ as a function of recursion depth.

8.4 Meta-Refinement as Empirical Exhibit: Recursive Hone with Kernel Locking

This subsection presents a meta-application of the conservation framework to the refinement process of this manuscript itself, serving as an illustrative, self-referential exhibit of Predictions 1–3. The preprint's iterative development—spanning 10+ cycles of AI-assisted compression and critique—functions as both equation and dataset: The initial draft acts as the source signal S_0, each review as a transformative compression $\mathcal{C}_\sigma$ or recursion step T, and the emergent stable structure as the conserved hard commitment kernel C_{hard}.

Consider the manuscript's core sections (e.g., abstract, introduction, and key contributions) as the linguistic signal under test. Unconstrained recursion (e.g., repeated AI rephrasing without gating) introduces variability: 10 turns yield divergent results, with fidelity decaying due to paraphrase drift or bloat (non-essential expansions). However, enforced gating—implicitly applied via manual oversight (analogous to MO§ES™ lineage validation)—mitigates this, converging to a stable kernel where hard commitments (e.g., "commitment defined as minimal invariant") persist across transforms.

Formally, the refinement process is modeled as:

$$C_{\text{hard}}(S_{n+1}) = \mathcal{E}\left(T(\mathcal{C}_\sigma(S_n))\right),$$

where $\mathcal{E}$ is the enforcement gate (reject if fidelity drop $\Delta > \theta$, e.g., $\theta = 0.2$ cosine similarity), ensuring invariance till sharp collapse. In exploratory runs (simulated on draft snippets), unconstrained paths exhibit $\Delta(n) \approx 1 - e^{-0.8n}$ (drift rise to collapse), while enforced paths flatten to $\Delta(n) \approx 0.05$, honing the kernel without erosion.

This exhibit demonstrates that information under recursion is not "sharpened into nothing" (as with unchecked loss), but conserved as an emergent pattern when locked—representative of the framework's universality, where multiple simulation paths converge to efficiency.

8.4.1 Hypothesis: MO§ES™ Recognition of Multiple Truths via Simulation

We hypothesize that MO§ES™, as an enforcement architecture, recognizes and simulates multiple emergent truths (divergent but valid kernels from varied recursion paths) in real-time, reducing bloat, drift, and collapse while optimizing for efficiency. This is testable across scenarios:

1. **Unconstrained Simulation (No MO§ES™):** A thousand turns on a signal like this manuscript produces bloat (redundant expansions), drift (semantic erosion), and potential collapse (fidelity to zero), with perhaps a residual drop of truth (isolated kernels surviving by chance).
2. **MO§ES™ in General (Without Locks):** Applies baseline gating and lineage constraints, significantly reducing bloat/drift/collapse compared to A (e.g., fidelity decay halved via quarantine), but allowing some variability in truths.
3. **MO§ES™ with Locks (Full Enforcement):** Incorporates hard kernel locking (e.g., cryptographic hashing on invariants), further reducing A and B outcomes—converging multiple truths to stable, similar results in far fewer turns (e.g., 10 offline equivalents).

This equates to:

1. **Equated System Outcome:** An unconstrained system prone to drift, bloat, and collapse; conversely, the enforced variant produces efficiency (reduced simulation overhead), creates emergent energy (optimized resource uptick via conserved kernels), and captivates monetarily (valuable truths touching ecosystems)—though such implications remain exploratory for now.

Preliminary meta-runs on this draft align: Locked paths yield consistent kernels (e.g., "conservation under compression" invariant), inviting larger-scale tests to validate or falsify.

9 Discussion and Future Directions

9.1 Implications

1. **Foundational Principle:** Commitment conservation may constitute a foundational principle for language systems, analogous to conservation laws in physics.
2. **Recursive Safety:** Systems that violate commitment conservation under recursion are inherently unstable and prone to drift.
3. **Verification:** Lineage tracking enables verification of identity preservation, preventing forgery and enabling accountability.
4. **Cross-Domain Applicability:** The framework applies to structured signals beyond natural language, including code, speech, and formal systems.

9.2 Limitations

1. **Corpus Size:** Preliminary tests used a limited corpus. Large-scale validation is required.
2. **Compression Definition:** The optimal compression transformation T_c may vary by domain and application.
3. **Computational Cost:** Compression and lineage tracking impose computational overhead.
4. **Adversarial Robustness:** The framework has not been tested against adversarial attacks.

9.3 Future Work

1. **Large-Scale Validation:** Test on corpora with $> 10,000$ samples across diverse domains.
2. **Alternative Compression:** Explore different compression algorithms and compare performance.
3. **Adversarial Testing:** Evaluate robustness against adversarial attacks and forgery attempts.
4. **Cross-Domain Extension:** Apply framework to speech, video, and multimodal signals.
5. **Theoretical Refinement:** Develop tighter bounds on commitment stability and drift rates.
6. **Governance Mechanisms:** Design protocols for multi-agent systems with commitment conservation.

9.4 Broader Context

Recent work in language models has highlighted challenges with recursive stability [2, 3, 4, 5, 6, 7]. MO§ES™ provides a minimal enforcement architecture that addresses these challenges through compression gating and lineage tracking, without relying on model-specific assumptions.

10 Conclusion

We have introduced commitment conservation as a candidate foundational principle for language systems under transformation and recursion. The principle states that commitment—the minimal, identity-preserving content—remains invariant under loss-inducing transformations when compression and lineage constraints are enforced.

We formalized this principle through:

1. Definitions of commitment, compression, and recursive stability
2. Theorems demonstrating conservation under compression and recursion
3. Corollaries showing non-conservation in probabilistic and agent-based systems
4. Preliminary empirical validation on a limited corpus
5. A public falsification protocol for large-scale replication
6. MO§ES™ as a minimal enforcement architecture

The framework is falsifiable: it predicts that compression + lineage systems will maintain high commitment stability (> 0.9) under recursion, while probabilistic systems without compression will exhibit drift. We invite the research community to validate, refine, or falsify these predictions through large-scale adversarial testing.

If validated, commitment conservation could provide a substrate for stable, verifiable ecosystems of language across time, media, and sovereign instances—analogous to TCP/IP's unification of networks or Git's lineage tracking for code.

We conclude that commitment conservation constitutes a viable candidate for a foundational principle in the physics of information-bearing language systems. Its validation, refinement, or falsification now rests squarely with independent theoretical critique and large-scale empirical testing by researchers with access to production-grade infrastructure.

Intellectual Property Disclosure

The enforcement architecture described herein (MO§ES™) is protected by provisional patent applications and trademark registration. These protections cover specific implementations of compression gating, cryptographic lineage DAGs, and hardware anchoring. The underlying conservation principle, falsification protocol, and theoretical framework are not restricted and are presented for open scientific investigation.

Acknowledgments

The author thanks the research community for ongoing discussions and feedback. The test harness and corpus are available at `https://github.com/SunrisesIllNeverSee/commitment-test-harness` for public replication and falsification.

References

[1] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.

[2] Corrêa, C., Schmid, P., Goyal, K., Kim, J., et al. (2025). Iterative Deployment Improves Planning Skills in LLMs. arXiv preprint arXiv:2512.24940.

[3] Xie, Z., Ma, Y., Zhou, Y., et al. (2025). mHC: Manifold-Constrained Hyper-Connections for Stable Scaling. arXiv preprint arXiv:2512.24880.

[4] Chang, E. (2025). The Missing Layer of AGI: From Pattern Alchemy to Coordination Physics. arXiv preprint arXiv:2512.05765.

[5] Zhang, H., Liu, A., et al. (2025). Recursive Language Models. arXiv preprint arXiv:2512.24601.

[6] Guo, D., Yang, D., Zhang, H., et al. (2025). DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. arXiv preprint arXiv:2501.12948.

[7] Chen, Z., Wang, H., Li, T., et al. (2026). SimpleMem: A Simple Memory Mechanism with Structured Compression for Long-Context Language Agents. arXiv preprint arXiv:2601.02553.

[8] Park, J. S., O'Brien, J. C., Cai, C. J., et al. (2023). Generative Agents: Interactive Simulacra of Human Behavior. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 1–22.

[9] Bai, Y., Kadavath, S., Kundu, S., et al. (2022). Constitutional AI: Harmlessness from AI Feedback. arXiv preprint arXiv:2212.08073.

[10] Schmidhuber, J. (2008). Driven by Compression Progress: A Simple Principle Explains Essential Aspects of Subjective Beauty, Novelty, Surprise, Interestingness, Attention, Curiosity, Creativity, Art, Science, Music, Jokes.

[11] Goertzel, B., et al. (2014). A cognitive architecture based on cognitive synergy.

[12] Looks, M. (2006). Meta-optimizing semantic evolutionary search.

[13] Looks, M. (2009). Scalable meta-optimization: A case study with the distributed hierarchical genetic algorithm.