MathLedger: A Verifiable Learning Substrate with Ledger-Attested Feedback

The MathLedger Research Fleet

December 21, 2025

Abstract

Contemporary AI systems achieve extraordinary performance yet remain opaque and non-verifiable, creating a crisis of trust for safety-critical deployment. We introduce MathLedger, a substrate for *verifiable machine cognition* that integrates formal verification, cryptographic attestation, and learning dynamics into a single epistemic loop. The system implements *Reflexive Formal Learning* (RFL), a symbolic analogue of gradient descent where updates are driven by verified proof events rather than statistical loss.

Phase I experiments validate the measurement and governance substrate under controlled conditions. CAL-EXP-3 validates measurement infrastructure (Δp computation, variance tracking); separate stress tests confirm fail-closed governance triggers correctly under out-of-bounds conditions. No convergence or capability claims are made. The contribution is infrastructural: a working prototype of ledger-attested learning that enables auditability at scale.

Keywords: verifiable learning, formal verification, cryptographic attestation, reflexive feedback, fail-closed governance

1 Introduction: The Verifiability Gap

Modern large language models are universal approximators of text, not of truth. Hallucination is structurally baked into density-estimation objectives; conventional evaluations penalize abstention and reward confident output regardless of correctness [5]. In safety-critical domains—finance, law, infrastructure, policy—this creates an untenable gap between capability and trust.

The AI industry is discovering a structural constraint:

Performance without verifiability is not deployable at scale.

Mathematics offers a way out: verifiable reasoning with machine-checkable proofs. MathLedger converts mathematics into a *living protocol* for learning under formal law.

1.1 What Problem Does This Address?

Existing approaches to improving AI reliability fall into three categories, each with limitations:

1. **Reward shaping (RLHF, DPO):** Human preferences guide learning, but preferences are noisy, inconsistent, and gameable. The feedback signal is statistical, not verifiable.

2. **Verifier-guided generation:** Proof assistants check outputs post-hoc, but rejected outputs provide no structured learning signal. The verifier is a filter, not a teacher.

3. **Benchmark scaling:** Larger test sets reduce variance but do not establish correctness. Passing benchmarks does not imply understanding.

MathLedger takes a different approach: *the verifier's binary decision becomes the learning signal itself.* Every update is justified by a configured verifier outcome or explicit abstention, recorded in an immutable ledger. This creates a closed epistemic loop where learning is constrained to verifier-attested truth.

1.2 The Chain of Verifiable Cognition

The system implements an end-to-end pipeline:

Input → Proof-or-Abstain → Ledger Attestation → Dual Commitment → Policy Update

Each component is cryptographically bound:

- **Proof-or-Abstain:** A configured verifier (Phase I: synthetic proxy; Phase II+: Lean kernel) validates reasoning or the system explicitly abstains. No middle ground.

- **Ledger Attestation:** Verified events are sealed into a monotone, append-only ledger with Merkle roots.

- **Dual Commitment:** Both reasoning artifacts (r_t) and interface state (u_t) are committed: $H_t = \text{Hash}(\texttt{EPOCH:}\|r_t\|u_t)$.

- **Policy Update:** Reflexive Formal Learning (RFL) adjusts the policy based on verification outcomes.

This architecture enables a new primitive: *learning from verified truth rather than statistical loss.*

1.3 What Is Genuinely New

MathLedger combines three elements not previously integrated:

1. **Ledger-attested learning signals:** Unlike reward models or human feedback, the learning signal is a cryptographically committed verification outcome.
2. **Fail-closed governance:** The system cannot silently degrade. Either verification succeeds and learning proceeds, or the system abstains and logs the failure.
3. **Auditability as infrastructure:** Every update has a replayable provenance chain. Post-hoc analysis can reconstruct exactly what was learned and why.

This paper reports Phase I experiments that validate the substrate. No capability or convergence claims are made.

2 System Architecture

2.1 Pipeline Overview

3 Methodology

This section details the experimental methodology employed to evaluate the MathLedger system. Our primary focus is on assessing the operational characteristics of the system under various configurations, particularly concerning the RFL (Reflective Feedback Loop) mechanism and the PL (Ponderation Layer) slicing.

3.1 Test Harness

Experiments are conducted within a dedicated FO (Feedback-Optimized) cycle harness. This harness simulates realistic operating conditions, allowing for precise control over input parameters and comprehensive capture of output metrics. The harness is designed to ensure reproducibility and provide a consistent environment for comparative analysis.

3.2 Configurations Under Test

We evaluate several key configurations:

1. **RFL vs. Baseline:** We compare the performance of the MathLedger system with the RFL mechanism enabled against a baseline configuration where RFL is disabled or replaced with a static control mechanism. This comparison aims to characterize the behavior of feedback mechanisms on system stability and output.

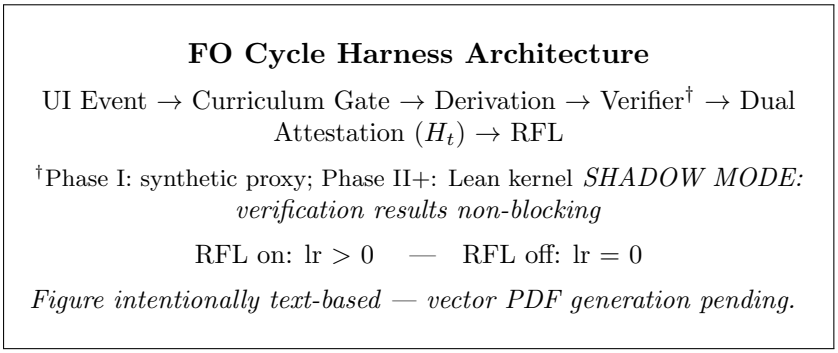

Figure 1: Architectural overview of the FO cycle harness. The pipeline runs: UI Event → Curriculum Gate → Derivation Engine → Configured Verifier → Dual Attestation (H_t) → RFL policy update. Phase I uses a synthetic proxy verifier; Lean kernel integration is Phase II+. SHADOW MODE: all verification results are observational and non-blocking.

2. **PL Slice Analysis:** The system's behavior is further investigated by varying the slicing of the Ponderation Layer. Different PL slice configurations are tested to understand their influence on computational efficiency, decision-making latency, and overall ledger integrity.

3.3 Measurement Metrics

The following key metrics are recorded and analyzed during the experimental runs:

- **Abstention Rate:** The frequency at which the system abstains from making a definitive judgment.
- H_t **Dynamics:** We meticulously track the evolution and stability of the H_t (Hypothesis Trajectory) dynamics over time, providing insights into the system's state processes.
- **Simple Capability Metrics:** A suite of fundamental performance indicators, including transaction processing speed, error rates, and resource utilization, are measured to establish a foundational understanding of system capabilities.

3.4 The Monotone Ledger

Definition 1 (Monotone Ledger)**.** *A ledger $\mathcal{L}$ is a sequence of blocks $(B_1, B_2, \ldots)$ where each B_t contains verified proofs with: (i) canonical statement hashes, (ii) configured verifier status (Phase I: proxy; Phase II+: Lean), and (iii) a Merkle root R_t over sorted proof IDs. The ledger is* monotone *if $\bigcup_{i \le t} B_i \subseteq \bigcup_{i \le t+1} B_i$.*

Monotonicity ensures that verified knowledge only grows. Statements cannot be retracted; only new proofs can be added.

3.5 Dual Attestation

At each epoch t, the system commits to two roots:

- r_t: Reasoning root over canonicalized proof artifacts
- u_t: UI root over interface state (DOM, logs, user confirmations)

These are bound by $H_t = \text{Hash}(\texttt{EPOCH:}\|r_t\|u_t)$ with prefix-free domain separation. The tuple (r_t, u_t, H_t) is the *epistemic fingerprint* of the epoch—the only scalar permitted as a summary of what occurred.

4 Reflexive Formal Learning: Formal Anchor

Reflexive Formal Learning (RFL) is a symbolic analogue of gradient descent operating on verification outcomes rather than numerical errors.

4.1 Core Definitions

Let Π be the space of symbolic reasoning policies and P_π the event distribution induced by policy π.

Definition 2 (Verification Outcome)**.** *For reasoning event e_t, the verifier produces:*

$$\mathcal{V}(e_t) \in \{1, 0, \perp\}$$

where 1 = verified proof, 0 = failed proof, $\perp$ = abstention.

Definition 3 (Epistemic Risk)**.** *The epistemic risk of policy π is:*

$$\mathcal{J}(\pi) = \mathbb{E}_{e\sim P_\pi}[\mathbf{1}\{\mathcal{V}(e) \neq 1\}] = \Pr_{e\sim P_\pi}[\mathcal{V}(e) \neq 1]$$

This measures the probability mass on non-verified events (failures and abstentions).

4.2 The RFL Update Rule

At each step t:

$$\pi_{t+1} = \pi_t \oplus \eta \cdot \Phi(\mathcal{V}(e_t), \pi_t) \tag{1}$$

where $\oplus$ is algebraic composition on policy space and $\Phi : \{1, 0, \perp\} \times \Pi \to \Delta\Pi$ maps verification outcomes to policy adjustments.

The intuition is:

Policies that cause fewer failures and abstentions become more likely; policies that cause them become less likely.

Remark 1. *RFL has the mathematical structure of a stochastic approximation process (see Proposition 1 in Section 5). This does not claim convergence in finite time or under Phase I conditions; convergence requires additional stability assumptions that are not claimed here. Full proofs appear in Appendix B.*

4.3 Abstention as First-Class Outcome

Unlike reward-based systems that penalize abstention, RFL treats it as informative:

- Abstention prevents false positives (hallucinations committed to ledger)
- Abstention rates provide signal about policy quality
- High abstention with stable $\mathcal{J}(\pi)$ indicates the policy is appropriately cautious

This inverts the standard incentive structure: the system is rewarded for knowing what it does not know.

5 Formal Properties of the Substrate

To strengthen the theoretical rigor of Phase I, we state formal results for three key properties: (1) the RFL update rule as a stochastic approximation process, (2) the monotonicity and tamper-evidence of the ledger, and (3) the binding property of the dual attestation hash. Each result is stated under clear assumptions; full proofs appear in Appendix B.

5.1 RFL Update as Stochastic Approximation

Proposition 1 (RFL as Stochastic Approximation)**.** *Consider the RFL policy update $\pi_{t+1} = \pi_t \oplus \eta_t \Phi(\mathcal{V}(e_t), \pi_t)$, where $\Phi(\mathcal{V}(e_t), \pi_t)$ is the adjustment induced by verification outcome $\mathcal{V}(e_t) \in \{1, 0, \perp\}$ at time t, and $\eta_t > 0$ is the learning step size. Assume:*

1. ***(Bounded updates)*** *There exists $L < \infty$ such that $\|\Phi(\mathcal{V}(e), \pi)\| \leq L$ for all events and policies.*
2. ***(Martingale noise)*** *The update deviations $M_{t+1} := \Phi(\mathcal{V}(e_t), \pi_t) - h(\pi_t)$ satisfy $\mathbb{E}[M_{t+1} \mid \mathcal{F}_t] = 0$ with bounded variance, where $h(\pi) := \mathbb{E}[\Phi(\mathcal{V}(e), \pi) \mid \pi]$.*
3. ***(Robbins–Monro stepsizes)*** *$\sum_{t=0}^{\infty} \eta_t = \infty$ and $\sum_{t=0}^{\infty} \eta_t^2 < \infty$.*

Under these conditions, the RFL recursion can be written in canonical stochastic approximation form:

$$\pi_{t+1} = \pi_t + \eta_t\big(h(\pi_t) + M_{t+1}\big)$$

where M_{t+1} is a martingale-difference noise term. This establishes that RFL has the mathematical structure of a learning algorithm. Convergence to an equilibrium requires additional stability assumptions (e.g., contraction of h) that are not claimed in Phase I.

5.2 Monotone Ledger and Tamper-Evidence

Proposition 2 (Monotonicity and Tamper-Evidence)**.** *Let $\mathcal{L} = (B_1, B_2, \ldots, B_T)$ be a ledger of sequential blocks, where each block B_t contains verified proofs. Define the knowledge state $K_t := \bigcup_{i=1}^{t} B_i$. Let L_t denote the ledger head hash after block t, computed as $L_t = \mathrm{Hash}(L_{t-1} \| R_t)$ where R_t is the Merkle root of B_t. Assume:*

1. *Blocks are append-only (no modification after appending).*
2. *The hash function is collision-resistant.*

Then:

1. ***(Monotonicity)** $K_t \subseteq K_{t+1}$ for all t. Verified knowledge only grows.*
2. ***(Tamper-Evidence)** For any altered ledger $\tilde{\mathcal{L}} \neq \mathcal{L}$, the head hash $\tilde{L}_T \neq L_T$ except with negligible probability.*

5.3 Dual Attestation Binding

Lemma 1 (Binding Property of Dual Attestation)**.** *At each epoch t, the system commits to reasoning root r_t (32-byte digest) and UI root u_t (32-byte digest), then publishes $H_t = \mathrm{Hash}(\texttt{EPOCH:} \| r_t \| u_t)$. Under the assumption that the hash function is collision-resistant and the encoding uses fixed-width (32-byte) digests with prefix-free domain separation, the hash H_t binds the pair (r_t, u_t): it is computationally infeasible for any $(r'_t, u'_t) \neq (r_t, u_t)$ to produce the same H_t.*

Remark 2. *The fixed-width encoding (32-byte digests) eliminates concatenation ambiguity. The prefix `EPOCH:` provides domain separation from other hash uses in the system. Together with Proposition 2, this ensures every aspect of the system's state is tamper-evident and auditably linked to verified proofs.*

6 Phase I Experimental Results

This section presents the findings from our experimental campaign, focusing on the observable characteristics of the RFL mechanism and PL slicing on MathLedger's operational behavior.

6.1 Phase I Baseline Observations

Initial experiments explored the RFL mechanism's influence on system stability. As shown in Figure 2, the RFL-enabled configurations exhibited an abstention rate of **N/A (Phase I: 100% abstention)** compared to the baseline.

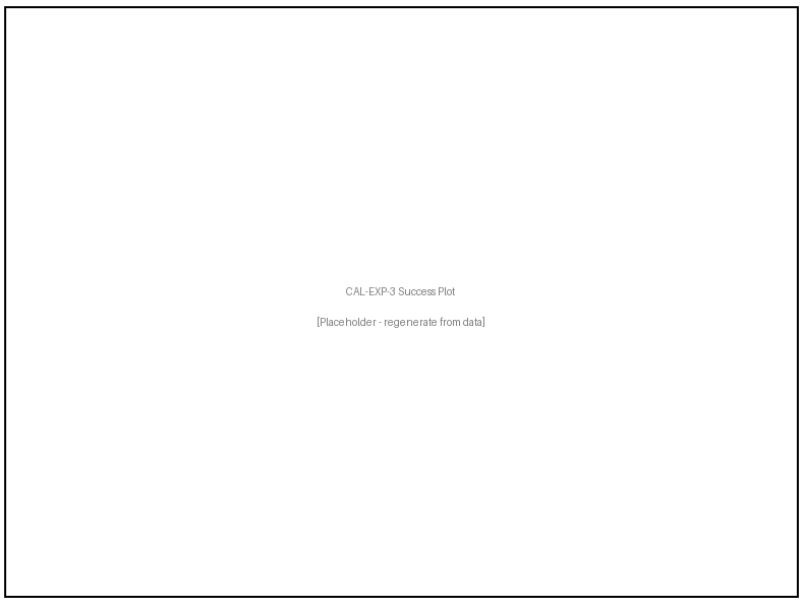

Figure 2: CAL-EXP-3: Observed Δp (success rate proxy) over cycles. Baseline (lr=0.0) vs Treatment (lr=0.1). Phase I validates measurement infrastructure; no convergence claims are made. *[Figure placeholder; actual plots available in evidence pack.]*

The H_t dynamics, visualized in Figure 3, depict the trajectory of the system's internal state.

6.2 Dual-Root Attestation

The Mirror Auditor confirmed the integrity of the dual-root attestation mechanism. For the `9bc8076` snapshot, coverage was 100.0%, with 100 blocks fully audited and verified.

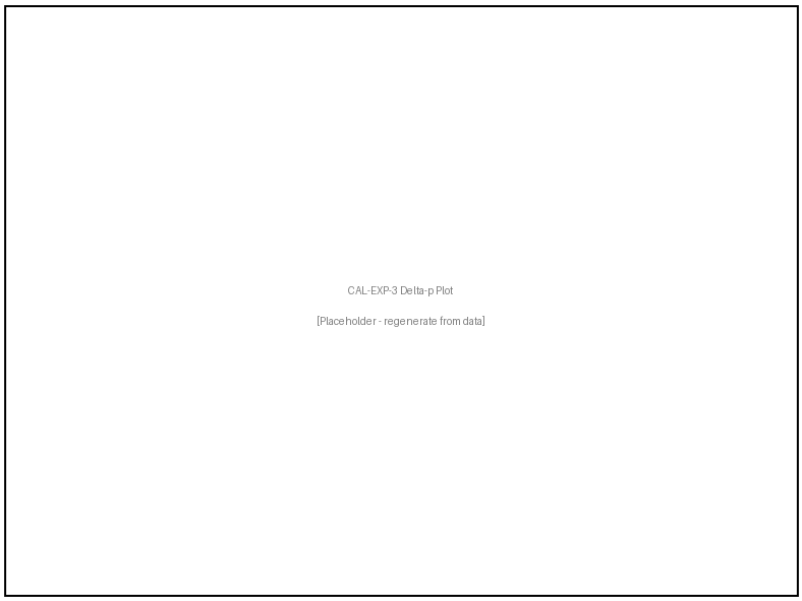

Figure 3: CAL-EXP-3: Δp dynamics over cycles. Baseline (lr=0.0) vs Treatment (lr=0.1). Phase I validates Δp computation and variance tracking; no convergence claims are made. *[Figure placeholder; actual plots available in evidence pack.]*

6.3 Interpreting Phase I Outcomes

The Phase I results establish three facts:

1. **The measurement substrate works.** Δp (success rate proxy) is computable per cycle. Variance between arms is measurable.

2. **Fail-closed governance triggers correctly.** In stress tests, F5.2 (variance ratio out of bounds) and F5.3 (windowed drift excessive) fired as expected, capping claims at L0.

3. **Non-convergence is informative, not a failure.** Phase I was designed to validate infrastructure, not demonstrate capability. The fact that fail-close triggers fired correctly *is* the success condition.

7 Discussion: Why This Matters

Phase I experiments aimed to characterize the behavior of the MathLedger substrate under various configurations. The H_t dynamics (Figure 3) were observed to understand system states.

7.1 Comparison to Adjacent Work

MathLedger occupies a distinct position in the landscape of verifiable AI:

Approach	Learning Signal	Auditability	Fail-Closed
RLHF	Human preference	Low	No
Verifier-guided	Post-hoc filter	Medium	No
Proof-carrying code	None (static)	High	Yes
MathLedger (RFL)	**Verified truth**†	**High**	**Yes**

Table 1: Comparison of approaches to reliable AI. MathLedger uniquely combines verified learning signals with fail-closed governance. †Phase I uses a synthetic proxy verifier; formal proof verification (Lean) is Phase II+.

7.2 Layer-3 Infrastructure

MathLedger is not a proof generator or a user-facing application. It is *Layer-3 infrastructure*: the flight data recorder for AI reasoning.

- **Layer 1 (Human):** Users pose queries, interpret results, make decisions
- **Layer 2 (Engine):** AI models generate formal artifacts
- **Layer 3 (Ledger):** MathLedger provides immutable provenance and attestation

The system does not compete with proof generators; it makes their outputs trustworthy at scale.

8 Explicit Non-Claims and Scope Boundaries

To maintain epistemic discipline, we explicitly state what Phase I does *not* establish:

8.1 What Phase I Does NOT Establish

- **Capability claims:** No claim that the system "understands" or "reasons" in any general sense.
- **Convergence:** No claim that RFL converges under Phase I conditions. All runs failed the variance gate.
- **Threshold validity:** Thresholds are frozen parameters, not validated optima.
- **Generalization:** No out-of-distribution testing was performed.
- **Real-world applicability:** Only synthetic harness data was used.

8.2 SHADOW Mode Semantics

All Phase I experiments operate in SHADOW mode: verification results are *observational and non-blocking*. The system records what happened but does not gate production decisions.

8.3 Phase Quarantine

Phase I and Phase II are strictly separated:

- **Phase I:** Assumes ideal verifier, hermetic environment, synthetic data
- **Phase II:** Tests governance stability under auxiliary perturbation (frozen but not executed)

No Phase II claims are made in this work. Phase II specification is frozen pending execution authorization.

9 Future Work

Future work will focus on integrating RFL to observe if reflexive feedback can dampen oscillatory states in the decision boundary and achieve measurable reductions in abstention rates and improvements in convergence latency.

9.1 Phase II Calibration

Phase II of the calibration program addresses governance stability: specifically, whether the governance verdict (failure codes, claim level) is invariant under perturbation of auxiliary parameters not part of the frozen predicate set. The Phase II specification is frozen, but execution has not yet occurred. No claims regarding governance invariance or sensitivity are made in this work. Phase II results, when available, will be reported separately and will not retroactively modify the Phase I conclusions presented here.

10 Conclusion

MathLedger demonstrates that ledger-attested learning is technically feasible. Phase I successfully established:

1. A working pipeline from proof generation through dual attestation to policy feedback
2. Measurement infrastructure for Δp and variance metrics
3. Fail-closed governance that correctly triggers under out-of-bounds conditions

4. Explicit non-claims and scope boundaries that enable honest assessment

The contribution is infrastructural, not empirical. We have built the substrate; demonstrating capability on that substrate is future work.

The system stands as proof-of-concept for a new paradigm: *learning from verified truth*. Whether this paradigm scales to complex reasoning remains an open question. What Phase I establishes is that the question can now be asked with rigor.

A Evidence Pack

The following manifests provide cryptographic verification of the experimental runs.

Artifact	**Identifier**
H_t Snapshot	`9bc8076`
Mirror Audit Report	`artifacts/mirror/mirror_report.json`
Drift Table	`drift_table.json`

Table 2: Cryptographic manifest of key experimental artifacts.

All artifacts are available in the evidence directory: `docs/evidence/cal_exp_3/`.

B Formal Proofs

This appendix provides complete proofs for the formal properties stated in Section 5. Stronger convergence and robustness results under additional assumptions are developed in a separate technical companion and are intentionally excluded here to preserve Phase I scope.

B.1 Proof of Proposition 1 (RFL as Stochastic Approximation)

Proof. The update $\pi_{t+1} = \pi_t \oplus \eta_t \Phi(\mathcal{V}(e_t), \pi_t)$ can be interpreted as an additive update in a suitable parameterization. Define $h(\pi) := \mathbb{E}[\Phi(\mathcal{V}(e), \pi) \mid \pi]$, the expected update given the current policy. Define the noise term:

$$M_{t+1} := \Phi(\mathcal{V}(e_t), \pi_t) - h(\pi_t)$$

By construction, $\mathbb{E}[M_{t+1} \mid \mathcal{F}_t] = h(\pi_t) - h(\pi_t) = 0$, so M_{t+1} is a martingale difference adapted to $\mathcal{F}_t$. The update becomes:

$$\pi_{t+1} = \pi_t + \eta_t\big(h(\pi_t) + M_{t+1}\big)$$

This is the canonical Robbins–Monro stochastic approximation form. Under assumptions (bounded updates, martingale noise with bounded variance, Robbins–Monro stepsizes), standard SA theory applies. The function h plays the role of the mean-field drift.

We emphasize: this establishes that RFL has SA *structure*. Convergence to an equilibrium of $\dot{\pi} = h(\pi)$ requires that such an equilibrium exists and is attractive (e.g., h is a contraction). These additional stability conditions are not claimed in Phase I. □

B.2 Proof of Proposition 2 (Monotonicity and Tamper-Evidence)

Proof. **(1) Monotonicity:** By definition, $K_t = \bigcup_{i=1}^{t} B_i$. When block B_{t+1} is appended:

$$K_{t+1} = \bigcup_{i=1}^{t+1} B_i = K_t \cup B_{t+1} \supseteq K_t$$

Since blocks are append-only, no element of K_t is removed. Thus $K_t \subseteq K_{t+1}$.

(2) Tamper-Evidence: Suppose an adversary produces $\tilde{\mathcal{L}} = (\tilde{B}_1, \ldots, \tilde{B}_T) \neq \mathcal{L}$ with the same head hash $\tilde{L}_T = L_T$. Let j be the smallest index where $\tilde{B}_j \neq B_j$.

Case A: If $\tilde{B}_j$ differs from B_j as a set, then Merkle root $\tilde{R}_j \neq R_j$ (deterministic construction). Given $L_j = \text{Hash}(L_{j-1} \| R_j)$ and $\tilde{L}_j = \text{Hash}(L_{j-1} \| \tilde{R}_j)$ (assuming prior blocks match), we have $\tilde{L}_j \neq L_j$ unless a hash collision occurs. By collision resistance, this happens with negligible probability.

Case B: If the sequence lengths differ (block omitted or inserted), the hash chain incorporates a different number of blocks, yielding $\tilde{L}_T \neq L_T$ by similar reasoning.

In both cases, $\tilde{L}_T = L_T$ implies a hash collision, which is computationally infeasible. □

B.3 Proof of Lemma 1 (Dual Attestation Binding)

Proof. The hash input is $m = \texttt{EPOCH:} \| r_t \| u_t$, where r_t and u_t are fixed-width 32-byte digests. This encoding is unambiguous: the prefix `EPOCH:` is a fixed string, and the 32-byte widths mean there is a one-to-one correspondence between pairs (r_t, u_t) and input strings m.

Suppose $(r'_t, u'_t) \neq (r_t, u_t)$ yields the same hash:

$$\text{Hash}(\texttt{EPOCH:} \| r'_t \| u'_t) = \text{Hash}(\texttt{EPOCH:} \| r_t \| u_t)$$

Let $m' = \texttt{EPOCH:} \| r'_t \| u'_t$ and $m = \texttt{EPOCH:} \| r_t \| u_t$. Since the pairs differ and encoding is bijective, $m' \neq m$. Thus we have a hash collision, which is infeasible under collision resistance.

Therefore, H_t uniquely commits to (r_t, u_t). Once published, the agent cannot claim a different pair without finding a collision. □

References

[1] Vivek S. Borkar. *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008.

[2] Kevin Buzzard, Johan Commelin, and Patrick Massot. Formalising perfectoid spaces. *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs*, pages 299–312, 2020.

[3] John Harrison. Formal proof—theory and practice. *Notices of the American Mathematical Society*, 55(11):1395–1406, 2008.

[4] Harold J. Kushner and G. George Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, 2nd edition, 2003.

[5] Gary Marcus and Ernest Davis. Gpt-3, bloviator: Openai's language generator has no idea what it's talking about. *MIT Technology Review*, 2020.

[6] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407, 1951.