MathLedger: Reflexive Formal Learning and the Chain of Verifiable Cognition

—

2025

Abstract

Contemporary AI systems achieve extraordinary performance yet remain largely *opaque* and *non-verifiable*, creating a crisis of trust for safety-critical deployment and governance. We introduce *Reflexive Formal Learning (RFL)*, a symbolic and cryptographically verifiable learning paradigm in which every update is justified by a formally verified proof or abstention event recorded in an immutable ledger. The resulting "chain of verifiable cognition" constitutes a closed epistemic loop bridging logic, cryptography, and learning dynamics. RFL is shown to be a discrete, non-differentiable analogue of stochastic gradient descent whose learning signal arises from verified truth rather than statistical loss, enabling provable convergence of reasoning under a verifiable ledger substrate.

Figure 1 (to be generated): Chain of Verifiable Cognition
User input → PoA → Ledger r_t → UI u_t → RFL → update.

Figure 1: End-to-end epistemic loop (objects and arrows as shown; all edges attested).

1 Introduction

Mathematical reasoning systems increasingly require not only correctness but verifiable *cognitive integrity*—a guarantee that each inference can be cryptographically traced to a validated source. *MathLedger* unifies formal verification, machine learning, and cryptographic attestation into a single epistemic pipeline:

User-Verified Input → Proof-or-Abstain Reasoning → Ledger Attestation → UI Attestation → Reflexive Feedback

This "Chain of Verifiable Cognition" enables an AI system to learn exclusively from verified reasoning outcomes. The remainder of this paper formalizes *Reflexive Formal Learning (RFL)* as a symbolic analogue of gradient descent operating on verified events rather than numerical errors.

Table 1: Summary of Core Formal Constructs

Symbol	Meaning
$\pi_t \in \Pi$	Symbolic reasoning policy at time t (metric space $(\Pi; \|\cdot\|)$)
P_π	Policy–induced event measure over event space $\mathcal{E}$
e_t	Reasoning event (proof or abstention) under P_π
$\mathcal{V}(e_t)$	Verification outcome $\in \{1; 0; \perp\}$
Ꝟ(e)	Numeric surrogate $\mathbf{1}\{\mathcal{V}(e) \neq 1\} \in \{0; 1\}$
$\mathcal{L}$	Ledger of proofs/abstentions (predictable process)
Φ	Feedback map $\Phi : \{1; 0; \perp\} \times \Pi \to \Delta\Pi$
$\mathcal{U}$	Reflexive update $\Pi \times \{1; 0; \perp\} \times \Sigma \to \Pi$
Σ	Auxiliary signal space (prompts, contexts)
$\Delta\Pi$	Space of composable symbolic policy deltas
$\oplus$	Algebraic composition on Π applying $\Delta \in \Delta\Pi$ to π
L	Lipschitz factor of $\oplus$ s.t. $\|\pi \oplus \Delta - \pi\| \leq L \quad \|\Delta\|_\Delta$
$\|\cdot\|_\Delta$	Norm on $\Delta\Pi$
$(T; \rho)$	Mixing horizon and geometric rate (Assumption $A2^0$)
$r_t; u_t$	Reasoning/UI Merkle roots (dual attestation)
$\mathcal{I}_t$	Dual-attestation binder token (keyless or HMAC)
H_t	Epistemic entropy at step t
$\mathcal{J}(\pi)$	Epistemic risk functional (Def. 1)
$\mathsf{L}(\theta)$	Classical differentiable loss (SGD)
β	Epistemic scaling exponent

2 Related Work and Theoretical Context

Stochastic approximation and recent convergence extensions. Classical almost-supermartingale convergence (Robbins–Siegmund) guarantees pointwise convergence under a summable residual term. Recent work (e.g., Liu, Xie, & Zhang 2025a) relaxes this to *square-summable* disturbances, proving convergence *to a bounded set.* Our stability results instantiate these extensions: when a persistent verifier bias acts as a non-summable zero-order term, we obtain bounded-set convergence with an $O(''_v)$ radius. Our Markovian noise assumptions (Assumption A2′ in §6) align with modern SA analyses that handle time-inhomogeneous Markov noise with mixing and Lipschitz drift. *Cor. 1 instantiates the square-summable residual regime of Liu–Xie–Zhang (2025), yielding bounded-set convergence under Markov noise.*

Positioning. RFL offers a distinct paradigm for verifiable learning. While recent work on *proof-guided language models* uses verifier feedback as a reward signal to guide policy search in a reinforcement learning loop , RFL uses the verifier's binary decision as the *learning signal itself*, replacing a gradient with a symbolic update. Similarly, where *weak-to-strong generalization* frameworks train a strong model on labels generated by a weaker one , RFL's policy updates are a direct, deterministic function of verified truth events, not supervised fine-tuning. The ledger's immutable, attested history of reasoning also connects RFL to work on *proof-carrying data* and ledger-based AI audits , but with a closed-loop dynamic where the ledger actively drives learning.

3 Reflexive Formal Learning (RFL): Conceptual Definition

Let Π be a complete metric policy space with norm $\|\cdot\|$ and P_π the event measure induced by π.

Update algebra. We write $\Delta\Pi$ for the space of composable symbolic policy deltas and $\oplus$ for their algebraic composition in Π; thus $\pi \oplus \Delta$ denotes applying $\Delta \in \Delta\Pi$ to policy $\pi \in \Pi$. We equip $\Delta\Pi$ with a norm $\|\cdot\|_\Delta$ and assume *norm compatibility*: there exists $L_\oplus \geq 1$ with

$$\|\pi \oplus \Delta - \pi\| \leq L_\oplus \|\Delta\|_\Delta \quad \text{for all } \pi \in \Pi;\ \Delta \in \Delta\Pi:$$

Definition 1 (Epistemic risk)**.** *Given* P_π *and* $\vartheta(e) = \mathbf{1}\{\mathcal{V}(e) \neq 1\}$,

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

Range. Since $\vartheta(e) \in \{0; 1\}$, we have $0 \leq \mathcal{J}(\pi) \leq 1$ for all π.

4 The Reflexive Formal Learning (RFL) Update Law

Update operator and abstention damping. At each step t,

$$\pi_{t+1} = \pi_t \oplus \eta_f\, \Phi(\mathcal{V}(e_t); \pi_t); \qquad \Phi : \{1; 0; \perp\} \times \Pi \to \Delta\Pi: \tag{1}$$

By norm compatibility, committing $\pi_{t+1} = \pi_t \oplus \eta_f \Phi(\cdot)$ implies $\|\pi_{t+1} - \pi_t\| \leq \eta_f L_\oplus \|\Phi(\mathcal{V}(e_t); \pi_t)\|_\Delta$.

Toy example: one-step RFL update with pseudo-Lean

Goal and tactic (pseudo-Lean).

```
example (h1 :  P →Q) (h2 :  P) : Q := by
apply h1
exact h2
```

Let the current goal be $g_t : \; Q$ under context $\mathrm{ctx}_t = \{h_1 : P \to Q; \; h_2 : P\}$. The agent proposes tactic $s_t = \texttt{apply h1}$, producing subgoal $g'_t : P$, then proposes $s'_t = \texttt{exact h2}$.
Event and verification. Define the event

$$e_t = (g_t; \mathrm{ctx}_t; s_t : \mathsf{apply}; s'_t : \mathsf{exact}) \quad \text{and} \quad v_t = \mathcal{V}(e_t) \in \{1; 0; \perp\}:$$

Pattern features. Let $\phi_{\mathrm{pat}}(\mathrm{ctx}_t; s_t)$ denote a differentiable feature map.
Policy parameterization. Suppose π_t induces a score via parameters θ_t:

$$\mathrm{score}_t \; = \; \langle \theta_t; \; \phi_{\mathrm{pat}}(\mathrm{ctx}_t; s_t) \rangle \quad \Rightarrow \quad p_t(\mathrm{pattern}) \; = \; \sigma(\mathrm{score}_t):$$

Symbolic deltas. Instantiate

$$\Delta^+(\pi_t; e_t) \; \equiv \; \mathrm{inc}_\eta \; \phi_{\mathrm{pat}}(\mathrm{ctx}_t; s_t) \; ; \quad \Delta^-(\pi_t; e_t) \; \equiv \; -\,\mathrm{dec}_\eta \; \phi_{\mathrm{pat}}(\mathrm{ctx}_t; s_t) \; ;$$

with $\|\Delta^\pm\|_\Delta \leq M$.
Cases (Proof-or-Abstain).

- v_t=1: $\pi_{t+1} = \pi_t \oplus \eta_f \, \Delta^+(\pi_t; e_t)$; $\mathcal{J}$ decreases in expectation.
- v_t=$\perp$: no-op update; abstention logged.
- v_t=0: $\pi_{t+1} = \pi_t \oplus \eta_f \, \Delta^-(\pi_t; e_t)$; demotes the pattern.

Algorithm 1 RFL$\oplus$MCGS Planner (fail-closed, dual-attested)

```
Require: $\pi_t$, $\eta_f$, update $\Phi$, verifier (REPL), canonicalizers $C_R; C_U$, ledger $\mathcal{L}$
Initialize frontier at root Lean state; $E \leftarrow [\,]$        . list of per-event binders
while frontier nonempty do
    Expand node using policy $\pi_t$ to yield candidate events $e_t$
    $(v_t; \text{trace}; \text{build}) \leftarrow \text{REPL:check}(e_t)$
    if $v_t \neq 1$ then
        $\mathcal{L}$:abstain$(e_t)$; continue
    end if
    $P_t \leftarrow C_R(e_t; \text{trace}; \text{build})$;  $D_t \leftarrow C_U(\text{UI snapshot})$
    $r_t \leftarrow \text{Hash}(\mathsf{R:} \| P_t)$; $u_t \leftarrow \text{Hash}(\mathsf{U:} \| D_t)$; $I_t \leftarrow \text{Hash}(\mathsf{BIND} \| r_t \| u_t)$
    if $\neg\mathcal{L}$:verify$(P_t; D_t; I_t)$ then
        $\mathcal{L}$:abstain$(e_t)$; continue
    end if
    $\Delta_t \leftarrow \Phi(v_t; \pi_t)$;  $\pi_{t+1} \leftarrow \pi_t \oplus \eta_f \Delta_t$
    $\mathcal{L}$:commit$(e_t; r_t; u_t; I_t; \text{build})$;  $E$:append$(I_t)$
    Push children of $e_t$ to frontier with priority from $\pi_{t+1}$
end while
epoch_root ← Merkle($E$);  $\mathcal{L}$:finalize_epoch(epoch_root)
```

5 Dual Attestation and Security Model

The integrity of the RFL loop depends on cryptographically binding reasoning events to their presentation. We formalize this via a dual-attestation scheme and specify the security guarantees under a formal adversary model.

Dual-Root Ledger. At each step t, the system commits to two Merkle roots:

- **Reasoning Root (r_t):** Root over the canonicalized sequence of formal reasoning steps (proof tactics, intermediate goals) composing e_t.
- **UI Root (u_t):** Root over the canonicalized representation of the UI state (DOM/JSON, PNG, HAR log) that displays the outcome of e_t.

These roots are bound by a cryptographic token $\mathcal{I}_t = \text{Hash}(\texttt{"BIND:"} \| r_t \| u_t)$ with prefix-free domain separation. The tuple $(r_t; u_t; \mathcal{I}_t)$ is recorded on the ledger.

Domain tags and REPL provenance. We extend domain separation with tags `RPL:` (Lean REPL provenance; toolchain/build IDs) and `G:` (geometry engine artifacts). These tags are included by C_{R} prior to Merkle, binding $(r_t; u_t; \mathcal{I}_t)$ to verifier versions and domain-specific pipelines.

Adversary Model and Guarantees. We consider a PPT adversary acting as a malicious prover, verifier, or network observer. The ledger offers:

- **Collision Resistance:** Hash$(\cdot)$ is collision-resistant, preventing distinct artifacts from sharing a root.
- **Non-Malleability:** Modifying proof/UI artifacts invalidates Merkle roots and $\mathcal{I}_t$.
- **Replay Resistance:** Updates are indexed by t; replay $(r_k; u_k; \mathcal{I}_k)$ at $t > k$ is rejected; timestamps strengthen this.

Cryptographic Hardening.

- **Canonicalization:** All structured data MUST be canonicalized before hashing. We mandate RFC 8785 JCS for JSON; deterministic PNG and equivalent for other types.
- **Domain Separation:** All hash inputs use prefix-free tags: e.g., Hash("`LEAF:`" $\|$ ·), Hash("`NODE:`" $\| h_L \| h_R$); distinct tags for reasoning/UI/binder.
- **Constant-Time Ops:** Cryptographic comparisons MUST be constant-time to avoid timing side channels.
- **Version Pinning:** Record versions/hashes of $\mathcal{V}$, hash function, and canonicalizers on-ledger.

Zero-Knowledge Extensions. For privacy, proofs may be replaced by ZK certificates (e.g., PLONK for fast verification and small proofs; STARKs for transparency and post-quantum resistance, at larger sizes).

6 Convergence and Stability of Reflexive Formal Learning

Stepsizes. We identify $\alpha_t \equiv \eta_f$ in the constant-stepsize case; otherwise α_t denotes a decaying schedule used in the SA analysis.

Lemma 1 (Extended Robbins–Siegmund)**.** *Let* $\{Z_t\}$, $\{X_t\}$, $\{Y_t\}$, *and* $\{W_t\}$ *be non-negative,* $\mathcal{F}_t$*-adapted random sequences. Suppose* $\mathbb{E}[Z_{t+1} \mid \mathcal{F}_t] \le (1 + X_t)Z_t + Y_t - W_t$ *a.s. If* $\sum_t X_t < \infty$ *a.s. and* $\sum_t Y_t < \infty$ *a.s., then* Z_t *converges a.s. to a finite random variable and* $\sum_t W_t < \infty$ *a.s. Furthermore, if the summable condition on* $\{Y_t\}$ *is relaxed to square-summable (*$\sum_t Y_t^2 < \infty$ *a.s.) and the increments are bounded* $(Z_{t+1} - Z_t)_+ \le B_t$ *with* $\sum_t B_t^2 < \infty$ *a.s., then* $\{Z_t\}$ *converges a.s. to a bounded set.*

Assumption 1 (Adaptivity and bounded updates (A1))**.** *Let* $\{\mathcal{F}_t\}$ *denote the filtration generated by* $(\pi_s; e_s; \mathcal{V}(e_s))_{s\le t}$*. The iterates* π_t*, reasoning events* e_t*, and verification outcomes* $\mathcal{V}(e_t)$ *are* $\mathcal{F}_t$*-adapted. Moreover, the update increments are uniformly bounded: there exists* $M < \infty$ *such that* $\|\Phi(\mathcal{V}(e_t); \pi_t)\|_\Delta \le M$ *almost surely for all* t*.*

Assumption 2 (Verification-monotone descent (A2))**.** *There exist constants* $\lambda > 0$ *and an* $\mathcal{F}_t$*-adapted error term* $\xi_t \ge 0$ *with* $\sum_t \mathbb{E}[\xi_t] < \infty$ *such that the epistemic risk satisfies*

$$\mathbb{E}[\mathcal{J}(\pi_{t+1}) - \mathcal{J}(\pi_t) \mid \mathcal{F}_t] \le -\lambda \ \mathrm{Pr}(\mathcal{V}(e_t) = 1 \mid \mathcal{F}_t) + \xi_t$$

for all t*.*

Assumption 3 (Local linearization and contraction (A3))**.** *There exists a neighborhood* $\mathcal{N}$ *of* $\pi^\star$ *in which the averaged update map* $\mathcal{T}(\pi) := \mathbb{E}[\pi \oplus \eta_f \Phi(\mathcal{V}(e); \pi)]$ *is Gâteaux differentiable and Lipschitz. The Jacobian* $D\mathcal{T}(\pi^\star)$ *has spectral radius* < 1*, yielding a contraction on* $\mathcal{N}$ *in the ambient norm.*

Assumption 4 (Markovian Noise and Mixing (A2′))**.** *The event stream* $\{e_t\}$ *is generated via a policy-dependent Markov process* $\{X_t\}$ *on a state space* $\mathcal{X}$*. For each fixed policy* π*, the process has a unique stationary distribution* μ_π*. The process is uniformly geometrically ergodic:* uniformly over π*, there exist* $T \in \mathbb{N}$ *and* $\rho \in (0; 1)$ *such that* $\|\mathbb{P}_\pi(X_T \in \cdot \mid X_0 = x) - \mu_\pi(\cdot)\|_{\mathrm{TV}} \le \rho$ *for all* x*. The one-step expected cost change and transition probabilities are Lipschitz in* π*. In addition, there exists an adapted sequence* $B_t \ge 0$ *with* $(\mathcal{J}(\pi_{t+1}) - \mathcal{J}(\pi_t))_+ \le B_t$ *a.s. and* $\sum_t B_t^2 < \infty$ *a.s. (implied by A0 and bounded* Φ *under standard stepsizes).*

Remark 1. Uniform geometric ergodicity is standard in SA; weaker mixing (Doeblin minorization or spectral-gap conditions) can suffice in place of uniformity.

Theorem 1 (Almost-Sure Convergence with Lyapunov Potential)**.** *If Assumptions 1–2 hold and* $\mathbb{E}[\Delta\mathcal{J}_t|\mathcal{F}_t] \leq -c\|\Phi\|^2 + \epsilon_t$ *with* $\sum_t \mathbb{E}[\epsilon_t] < \infty$*, then* $\mathcal{J}(\pi_t) \to 0$ *a.s. and* $\pi_t \to \pi^\star$ *where* $\mathcal{V}(\pi^\star) = 1$*.*

Corollary 1 (Convergence to a bounded set under square-summable noise)**.** *Let* $X_t = \mathcal{J}(\pi_t)$*. Assume* $\sum_t \alpha_t = \infty$*,* $\sum_t \alpha_t^2 < \infty$*. If the RFL dynamics satisfy Lemma 1 with square-summable disturbance and bounded increments, then* $\{X_t\}$ *converges a.s. to a bounded set; pointwise convergence is recovered when the disturbance term is summable.*

Corollary 2 (Linear Convergence of Epistemic Risk)**.** *Under Theorem* **??***, if* $\Pr[\mathcal{V}(e_t) = 1] \geq c > 0$ *for all non-optimal policies, then* $\mathbb{E}[\mathcal{J}(\pi_t)]$ *converges to its limit at a linear rate.*

Corollary 3 (Local Stability of the Optimal Policy)**.** *Under Assumption 3, the fixed point* $\pi^\star$ *is locally asymptotically stable: any* π *within the contraction basin converges to* $\pi^\star$ *under RFL dynamics.*

Proposition 1 (Abstention as damping)**.** *If* $\Pr[\mathcal{V}(e_t) = \perp] = \alpha_t$ *and abstention cost* $\mathsf{cost}(\perp) = c_\perp \geq 0$ *(captures operational abstention cost), then*

$$\mathbb{E}[\mathcal{J}(\pi_{t+1})|\mathcal{F}_t] \leq (1-\lambda)(1-\alpha_t)\mathcal{J}(\pi_t) + c_\perp:$$

Higher α_t *increases safety but slows convergence.*

Theorem 2 (Stability under verifier imperfection)**.** (Full proof in Appendix C.) *Assume A1–A2 and Assumption 4. Let the verifier introduce bias* δ_t *with* $|\delta_t| \leq ''_v$*, entering as* $c_t = \alpha_t ''_v$*. If* $\sum_t \alpha_t = \infty$*,* $\sum_t \alpha_t^2 < \infty$*, and Lemma 1 conditions hold, then*

$$\limsup_{t\to\infty} \mathcal{J}(\pi_t) \leq \mathcal{J}^* + C\,''_v \quad \textit{a.s.}$$

for finite C *depending on* $(M; L_\oplus; \lambda)$ *and the stepsize schedule.*

Corollary 4 (Bounded-set Convergence Radius)**.** *Under Assumption 4 (mixing* $(T; \rho)$*) and verifier bias* $''_v$*,*

$$\limsup_{t\to\infty} \mathcal{J}(\pi_t) \leq C_1\,''_v + C_2(1-\rho)^T;$$

for finite $C_1; C_2$ *depending on* $(M; L_\oplus; \lambda)$ *and* $(T; \rho)$*.*

Table 2: Assumptions summary used in convergence analysis.

A1: Adaptivity & bounded updates	Assumption 1: $\pi_t; e_t; \mathcal{V}(e_t)$ are $\mathcal{F}_t$-adapted; $\|\Phi(\cdot)\| \leq M$.
A2: Verification-monotone descent	Assumption 2: expected descent inequality with summable residual.
A2^0: Markovian noise & mixing	Assumption 4: uniform geometric ergodicity, Lipschitz in π; bounded increments B_t with $\sum B_t^2 < \infty$.
A3: Local linearization & contraction	Assumption 3: local Gâteaux derivative, Lipschitz, and contraction of $\mathcal{T}$.
Stepsizes	$\sum_t \alpha_t = \infty$, $\sum_t \alpha_t^2 < \infty$; Algorithm **??** uses $\eta_f \equiv \alpha_t$ or a schedule.

7 Epistemic Scaling Laws

Evaluation plan. The empirical claims will be tested according to a preregistered protocol (Appendix A), specifying hypotheses, logging of $\{r_t; u_t; \mathcal{I}_t\}$, and robust regression analysis.

Framework	Noise	Dependence	Conclusion	Where used
Classical RS	P $c_t < \infty$	i.i.d./MD adapted	$X_t \to X$ a.s.	Baseline for Thm ??
Extended RS	P $c_t^2 < \infty$; bounded increments	Markov, mixing	$X_t \to$ bounded set a.s.	Cor. 1
RFL (this work)	$c_t = \alpha_t ''_v$	Policy–Markov	$\limsup \mathcal{J} \leq \mathcal{J} \; + C''_v$	Thm 2

Table 3: Classical vs. extended Robbins–Siegmund vs. our RFL instantiation.

Scaling law. Performance scales as $\Delta H \propto N_v^{-\beta}$ with $\beta \in (0; 1]$. Empirically $\beta \approx 1{=}2$ is consistent with diffusion-like uncertainty decay.

Interpretability and alignment perspective. Alignment and verifiability themselves follow scaling-law behavior; RFL provides a formal alternative grounded in verified events.

Empirical Outlook. Our training pipeline uses a *proof_sampler* process to generate events e_t under the current policy, producing a policy-dependent Markov stream. Assumption A2′ connects this stream to theory: the sampler's mixing and the verifier's acceptance yield $(T; \rho)$ and $''_v$. We will report $(\rho; T)$ proxies (autocorrelation decay) and empirical $''_v$ per run.

8 Emergent Directions

Reflexive Formal Perception (future work). Agents verify what was *seen* before reasoning; perceptual disagreements become ledger objects.

Ledger-Driven Theory Genesis (future work). Meta-agents mine sealed proofs to propose schemata under ledger governance.

Instrumentation Hooks (Phase I). Log perceptual disagreements and lemma-reuse frequencies; mine traces later.

9 Philosophical and Practical Boundaries: The Architect and the Healer

Framing. MathLedger embodies the *Architect*'s aspiration: intelligence grounded in provable truth. Its challenge is the *Healer*'s domain: extending verifiability into imperfect reality.

Open problems (Phase III).

1. **Scope of verification.**
2. **Verifier bottleneck.**
3. **Abstention vs. usefulness.**
4. **Semantic gap.**
5. **Computational cost.**

Research tracks (not blockers).

Scalable verification	Probabilistic checking; batching; ZK sealing.
Verified oracle stack	Verified kernels/compilers; attested builds.
Controlled abstention	Abstention budgets; explore-on-fail policies.
Semantic grounding bridge	Typed front-ends; certified parsers/tokenizers.
Compute-efficient guarantees	Proof caching/reuse; parallel tree-hash; ZK compression.

Figure 3 (to be generated): RFL Uplift Curves
y = proofs/hour; x = wall-clock time or iterations; curves (RFL, replay, no-feedback) with 95% CIs.

Figure 2: Pre-registered uplift curves comparing RFL vs. baselines.

Appendix A — Prior-Art Matrix and Red-Team Notes

This appendix summarizes adjacent systems and red-team considerations.

Appendix B — Evidence Manifest (Sealing Metadata)

We provide artifact paths and commits for reproducibility. We release `roots.json` and per-step Merkle inclusion proofs; a reproducibility script re-canonicalizes artifacts and re-derives $(r_t; u_t; \mathcal{I}_t)$ byte-for-byte.[1]

A Preregistration Protocol for Empirical Evaluation

A.1 Hypotheses

H1: $\log|\Delta H| = -\beta \log N_v + c$ with $\beta > 0$. H2: $\limsup_t \mathcal{J}(\pi_t)$ increases with $''_v$.

A.2 Tasks and Datasets

Lean4 theorem proving (miniF2F/synthetic), transformer policy π_t, Lean4 kernel as $\mathcal{V}$; simulate $''_v$ by flipping outcomes.

A.3 Proof Sampler and Logging

Log JSONL entries with step, policy/task ids, v_t, attestations $(r_t; u_t; \mathcal{I}_t)$, metrics (autocorr proxy, $''_v$, H_t).

[1] KangarooTwelve (K12) is a tree-hash mode; observed Merkle construction speedups are empirical (2–3×) on AVX2/AVX-512 vs. SHA-256 in our pipeline.

Table 1 (to be generated): Scaling Law Fit — β Estimates
Columns: run id, N_v, ΔH, fit $\hat{\beta}$, SE, R^2.

Table 4: Empirical fit of $\Delta H \propto N_v^{-\beta}$.

Figure 4 (to be generated): Ablation Study Results
Bars for (DA on/off) × (abstention penalty on/off); metric = $\Delta\mathcal{J}$ or proofs/hour.

Figure 3: Ablations for DA and abstention penalty.

A.4 Power

See Table 5.

Table 5: Power analysis for detecting β.

Detectable $\lvert\beta\rvert$	Required N_v
0.50	∼1,000
0.25	∼4,000
0.10	∼25,000

A.5 Analysis Plan

Huber regression for H1; Spearman correlation for H2. No manual point deletion; only run-level exclusions (hardware failure, zero-verify cold starts).

Appendix C — Proofs of Main Results

Assumptions recap. All proofs invoke A1, A2, and, where stated, A2′; stepsizes satisfy $\sum_t \alpha_t = \infty$, $\sum_t \alpha_t^2 < \infty$.

Proof of Theorem **??**. Let $X_t \overline{=} \mathcal{J}(\pi_t)$. By A2, $\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] \le X_t - \eta_f \lambda \mathbf{1}\{\mathcal{V}(e_t) = 1\} + \xi_t$. Boundedness ($X_t \in [0;1]$) and $\sum_t \mathbb{E}[\xi_t] < \infty$ yield the claim via Robbins–Siegmund. □

Proof of Theorem 2. Let $X_t = \mathcal{J}(\pi_t) - \mathcal{J}^*$. With $|\delta_t| \le ''_v$,

$$\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] \le X_t - \alpha_t \Delta_t + \alpha_t C_1 ''_v + \xi_t:$$

Set $Y_t = \alpha_t C_1 {''}_v + \xi_t$. Since $\sum_t (\alpha_t C_1 {''}_v)^2 < \infty$ and $(X_{t+1} - X_t)_+ \leq B_t$ with $\sum_t B_t^2 < \infty$ (A0 and bounded Φ), Lemma 1 applies, so X_t converges a.s. to a bounded set. Telescoping gives $\limsup_t \mathcal{J}(\pi_t) \leq \mathcal{J}^* + C {''}_v$ with C depending on $(M; L_{\oplus}; \lambda)$ and stepsizes. □

References