

# SUCCESSOR $\Omega$ HARNESS

Make Every Mission Build What Comes Next

Powered by TRANSFORMATIONAL INTELLIGENCE  $\infty$

For TRANSFORMATIONAL AUTONOMOUS ENTERPRISES:  
ACTIVATE. EVOLVE. SCALE.

MONTREAL.AI · AGI ALPHA Research & Engineering

A mission should leave more than an answer.  
It should leave a verified means of doing useful work next.

**Abstract.** Can a thin execution layer turn completed work into transferable productive capability while retaining evidence and controlling authority? We implement a bounded research companion over Alpha v1.41.0's released workbench extension interface. A constructor produces an exact rational detector-correction operator; a separate checker qualifies it under declared measurement premises. A successor adds a bracketing reference to support affine drift, then reuses the operator for mean response, rise contrast, and late endpoint. Durable receipts keep current qualification separate from historical evidence and production authority.

A prospectively frozen protocol compares retained construction, an equally equipped conventional cached implementation, repeated reconstruction, and direct expert construction over 96 supplied-model requests, a four-worker repetition, and deliberate model violations. All four controllers qualify 96 numerical intervals and resolve 78 threshold decisions. Retained construction uses 32,675 rational calls versus 37,422 for rebuilding (12.69% fewer), ties the conventional cached controller, and costs 92 more calls than direct expert construction. At the primary resource budget, all four have equal coverage. The software executes over simulated detector observations. No physical instrument, independent reviewer council, live settlement, or commercial profit is demonstrated. The contribution is an inspectable mechanism and a reproducible account of when transfer helps, ties, loses, or should stop.

**Three layers, kept separate.** **Demonstrated:** the executed bounded companion and its retained failures. **Testable next step:** independently calibrated hardware and qualified runtime integration. **Long-term ambition:** productive assets that help construct further scientific and industrial capacity.

Experimental research edition. AGI ALPHA is a project identity; achieved AGI and unlimited growth are not claimed. Historical papers and held integrations remain separate.

# 1 A completed mission can become a means of production

Imagine a small research service that measures a material's response. Its first mission succeeds by subtracting a detector's fixed offset. The useful achievement is not the number alone: it is the tested correction procedure, the observations that justify it, and the conditions under which another investigator can use it. The next request arrives with a drifting detector and a weaker signal. Reusing the old success flag would be a mistake. Reusing the old layout, data representation, and retained failure can help construct the next instrument.

This paper makes that progression executable. The instrument is a small linear operator over simulated detector readings, not laboratory hardware. Its coefficients determine which nuisance is canceled, which observations are needed, and how uncertainty propagates. A three-minute local experience exposes those facts before asking readers to understand the implementation.

## 1.1 The three-minute journey

Start the companion with `python3 companion/serve.py` after installing the documented Python and Node versions. Open the loopback address printed by the command. No wallet, account, model API, or newly issued credential is required.

1. **Choose a useful question.** Inspect a mean response, the rise between two measurements, or a late sampled level. The interface shows the acceptance threshold, read budget, and declared detector model.
2. **Meet the inherited limitation.** A single dark reference removes a constant offset. Change the conditions to affine drift: the retained parent is outside its qualified scope.
3. **Inspect the construction.** The bounded constructor compares no reference, one initial reference, and two bracketing references. It retains rejected candidates and their costs. The accepted operator is downloadable data that the worker actually executes.
4. **Build the successor.** A second reference makes the drift contribution removable within the stated model. The child inherits compatible structure and history; it receives a new local check.
5. **Challenge and recover.** Change signal, drift, or threshold; inspect a refused or unresolved result. Reopen the same request to recover its retained receipt. Inspect the counterexample in which two possible worlds give identical readings but different true signals.

The local user can alter inputs and obtain newly computed consequences. Recorded research results are labeled separately. The harness neither invents a physical observation nor converts a conditional numerical result into independent institutional approval.

## 1.2 The question and the contribution

Our research question is:

*Under what conditions can a lightweight harness turn completed work into transferable productive capability, while preserving evidence, controlling authority, and reducing the total cost of subsequent qualified missions?*

The contribution has three parts. First, an executable source extension connects mission specification, bounded construction, source-bound work receipts, a separately implemented checker, a durable local history, and a distinct successor requirement. Second, the experiment makes the

transfer boundary concrete: compatible implementation and observations can transfer, while changed nuisance assumptions require new construction and qualification. Third, the evaluation exposes the full measured cost vector and retains an equally capable conventional tie, expert advantages, and undetected model violations.

We do not claim that event logs, durable workflows, cached results, authorization, calibration, or reusable code are new. Temporal’s durable histories and RIFL’s durable completion records are established precedents (Temporal Technologies, n.d.; Lee et al., 2015). Proof-carrying code and its combination with authorization already distinguish semantic evidence from permission (Necula, 1997; Whitehead et al., 2004). Here, the new object of study is a particular, inspectable capability lifecycle and the conditions measured by its companion.

### 1.3 Why the Foundry failure stays in the story

The earlier Foundry research edition supplied a useful warning about the wrong acceptance target. In its flow/power challenge, a runtime residual of 4.9477% passed a 5% rule, while a separate numerical reference found 8.6964% calorimeter error against the reference model’s true heat. Agreement between the prediction and the reported measurement had not established accuracy of the useful measurand. The original edition also retained conventional engineering at 384 initial readings versus 672 for its construction, a conventional calibration tie, and a centralized modeled alternative CAD 460 cheaper (Boucher, 2026h). These are *historical reported results*, not fresh executions in this edition. [E1]

Three possible responses are worth distinguishing. A self-checking interface makes assumptions, reference channels, and failure conditions observable. A discoverability compiler searches for experiments that distinguish required decisions. A shared capability foundry allocates the cost and capacity of useful instruments across voluntarily participating investigations. The present contribution takes a deliberately narrow construction from the first direction and a finite search from the second. Shared physical production and commercial coordination remain testable developments.

The new detector experiment does not reproduce or repair the held thermal adapter. It changes a related measurement constraint by adding an actual executable observation procedure to a released extension seam. Its own common-cause failure shows why this construction still cannot validate all of its physical premises.

### 1.4 Transformational Intelligence as an operational method

We read the project’s foundational Transformational Intelligence manuscript and all six Out-stage papers at the pinned public source. They separate purpose from formulation, require executable constructions, preserve costs, and distinguish model-level results from physical or industrial ambition (Boucher, 2026g,d,f,a,e,c,b). Their separate experiments are not added together as evidence for this harness.

Our useful objective is a bounded statement about a detector response. The replaceable restriction is a measurement program with only one reference under conditions that now drift. The preserved obligations are the meaning of the requested functional, the acceptance policy, and the authority ceiling. The construction changes the available observations, not those obligations. Table 1 connects the stages to actual artifacts and computations in one project.

**Table 1.** The six Out-stages are causal contracts. Identify fixes the purpose before the six stages. Every cost belongs to the same retained history; no stage label counts as evidence by itself.

| Stage          | Consumed input and computation                                                              | Produced artifact and falsifier                                                             |
|----------------|---------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|
| Identify       | Functional, threshold, permitted actions, resource envelope; separate signal from nuisance. | Mission specification. Unsupported observable or scope fails.                               |
| Out-Learn      | Signal and reference observations with explicit noise premises.                             | Measurement record. A shared gain error can defeat its premises.                            |
| Out-Think      | Express readings as signal plus nuisance; preserve three future functionals.                | Rational matrix contract. Signal identity and nuisance moments must hold.                   |
| Out-Design     | Search the acquisition grammar and solve cancellation constraints.                          | Executable operator, failed candidates, and construction cost.                              |
| Out-Strategise | Check the current operator and any parent against their declared scope.                     | Qualified selection or refusal; rejected ancestry is retained.                              |
| Out-Execute    | Enforce read/operation budgets, run, recompute, and persist exact results.                  | Conditional interval, decision, and recovery record. Changed inputs cannot reuse its claim. |
| Out-Evolve     | Reveal drift, retain the parent, add a reference procedure, check anew.                     | Child operator and lineage. Prior proof and authority do not silently transfer.             |

## 2 What a transferable capability contains

**Definition 1** (Productive capability record). *A productive capability is an executable artifact together with its input/output contract, dependency identities, applicability conditions, evidence, cost history, and lineage. It describes an ability to do useful work. It is distinct from a security capability that conveys authority to access a protected object (Saltzer and Schroeder, 1975).*

Let a mission be  $M = (o, q, \mathcal{W}, \pi, B, \mathcal{A})$ : objective  $o$ , requested observable  $q$ , declared world class  $\mathcal{W}$ , acceptance policy  $\pi$ , resource-limit vector  $B$ , and permitted actions  $\mathcal{A}$ . Its reusable record is  $K = (a, d, \mathcal{W}_K, \pi_K, e, \ell, c)$ : executable artifact  $a$ , dependency identities  $d$ , applicability domain, qualification policy, evidence  $e$ , lineage  $\ell$ , and cost history  $c$ . A successor references  $K$ ; it does not mutate the parent into a more capable ancestor.

### 2.1 Four questions that must remain separate

- WORLD** Are the declared premises true of the present detector and inputs? Here they include a zero reference, unit gain, bounded noise, and a specified nuisance class.
- POLICY** What observable, quality threshold, resource budget, and stopping rule does this mission require? A weaker policy cannot silently replace a failed one.
- PROOF** What precisely did the checker establish, using which inputs, source, and assumptions? Here it establishes finite algebra and numerical enclosure under premises.
- AUTHORITY** Who may perform which consequential action now? This companion has an empty production-authority set. Local computation and history grant no laboratory, financial, governance, or deployment authority.

Identity does not imply correctness; correctness under a model does not imply physical validity; reviewer authentication does not imply independence; and possession of a qualified artifact does not imply present permission. The proof/authority distinction is established prior art (Whitehead et al., 2004). The engineering task is to enforce it at the actual transfer and recovery boundaries.

## 2.2 Constructing a detector procedure

At normalized times  $t = (1/4, 1/2, 3/4)$ , the declared detector model is

$$y_i = s_i + b_0 + b_1 t_i + \epsilon_i, \quad (1)$$

$$r(t) = b_0 + b_1 t + \eta_t, \quad (2)$$

where  $s_i$  is the desired signal and every observation error is bounded in absolute value by  $\epsilon$ . A dark reference has zero true signal. The constant-offset parent assumes  $b_1 = 0$ ; the child admits affine drift. These are exact rational model statements, not measured assurances about a physical detector.

The parent uses  $z_0 = (y_1, y_2, y_3, r(0))^\top$  and

$$A_0 = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}.$$

Under drift, its corrected signals retain  $b_1 t_i$ . Its mean retains  $b_1/2$ . With  $|b_1| \leq 40$ , that mean may carry 20 units of unremoved drift in addition to observation noise. A new scope label cannot remove the term.

The successor adds a reference at time one, uses  $z_1 = (y_1, y_2, y_3, r(0), r(1))^\top$ , and constructs

$$A_1 = \begin{bmatrix} 1 & 0 & 0 & -3/4 & -1/4 \\ 0 & 1 & 0 & -1/2 & -1/2 \\ 0 & 0 & 1 & -1/4 & -3/4 \end{bmatrix}. \quad (3)$$

Each row subtracts the interpolation between the two references. The separate checker tests signal identity and cancellation of constant and linear nuisance. The constructor searches three declared acquisition candidates; the expert supplies the analytic coefficients and faces the same checker. This is bounded construction within a supplied grammar, not open-ended invention.

Three missions use the corrected representation:

$$w_{\text{mean}} = (1/3, 1/3, 1/3), \quad w_{\text{rise}} = (-1, 0, 1), \quad w_{\text{late}} = (0, 0, 1).$$

They answer different questions about the same observations. Late endpoint means the sample at  $t = 3/4$ , not an unobserved value at time one. Mean response is not an energy or dose integral without an additional physical model.

**Proposition 1** (Conditional enclosure). *If  $A$  preserves signal coordinates and annihilates every declared nuisance column, the estimator  $\hat{q} = wAz$  has error at most*

$$B_q = \varepsilon \|wA\|_1$$

when each observation error has magnitude at most  $\varepsilon$ .

*Proof.* Write  $z = Ss + Nb + e$ . The checked identities  $AS = I$  and  $AN = 0$  give  $\hat{q} - ws = wAe$ . Then  $|wAe| \leq \sum_j |(wA)_j| |e_j| \leq \varepsilon \|wA\|_1$ .  $\square$

For the affine child, the bounds are  $2\varepsilon$ ,  $3\varepsilon$ , and  $2\varepsilon$  for mean, rise, and late endpoint. They are deterministic enclosures under supplied bounds, not confidence intervals. Physical uncertainty budgets and metrological traceability require further evidence ([Joint Committee for Guides in](#)

Metrology, 2008; Taylor and Kuyatt, 1994; National Institute of Standards and Technology, 2024).

For  $I = [L, U] = [\hat{q} - B_q, \hat{q} + B_q]$ , the companion tests

$$\rho(I) = \frac{B_q}{\min(|L|, |U|)} \leq \frac{1}{20},$$

only when zero lies outside  $I$ . It reports **ABOVE** threshold  $h$  if  $L \geq h$ , **BELOW** if  $U < h$ , and **UNRESOLVED** otherwise. A good accuracy bound need not resolve a decision near its threshold.

## 2.3 What the child inherits

The child can retain the signal layout, rational representation, checker interface, compatible observations, and ancestor costs and failures. It must construct the extra reference procedure and coefficients, validate the new matrix, bind it to current source, and check every actual result. A constant-offset proof does not establish affine-drift cancellation. A conditionally true statement remains historical evidence when the world class changes.

A common multiplicative gain error changes the desired signal while leaving dark references uninformative about gain. Two worlds can yield the same complete input stream. No deterministic checker seeing only that stream can distinguish them. An independently calibrated nonzero reference is a new construction requirement, not another label on the existing receipt.

## 2.4 When inheritance pays

Let  $C_0$  include construction and qualification of retained ancestors; let  $R_i$  include retrieval, applicability checks, new work, validation, and recovery for mission  $i$ . Let  $D_i$  be the complete cost of the strongest comparison strategy. Inheritance helps on one resource dimension only when

$$C_0 + \sum_{i=1}^n R_i < \sum_{i=1}^n D_i$$

for the same accepted obligations. A cache hit does not make validation or storage free. Rational operations, detector acquisitions, wall time, human effort, currency, and energy are not interchangeable. An advantage in one dimension is not profit.

The competent conventional comparator receives ordinary reuse. If it retains the same valid operator and observations and performs the same checks, a tie is expected. The harness's value then lies in an operable, inspectable lifecycle, not a demonstrated superiority over conventional engineering.

## 3 A thin execution layer with an explicit boundary

The companion uses the released JavaScript factory `createWorkbench(adapters)` at Alpha v1.41.0, commit `031bdb5d`. Its seven-module dependency closure is copied byte-for-byte with a manifest and license; no held Harness or thermal implementation is copied (MONTREAL.AI, 2026a,b). A new fixed adapter supplies three methods: a manifest, input validation, and bounded execution. The released workbench constructs its existing job draft, binds input and source identities, hashes artifacts, and inspects the returned pack. [E2,E3]

The released pack also retains generic sample-job scaffolding whose aggregation comparator text does not describe this detector. Those seven zero-funded drafts are not commissioned detector

## From a bounded mission to a reusable capability

Local source extension · released Workbench SDK · unsigned persistence

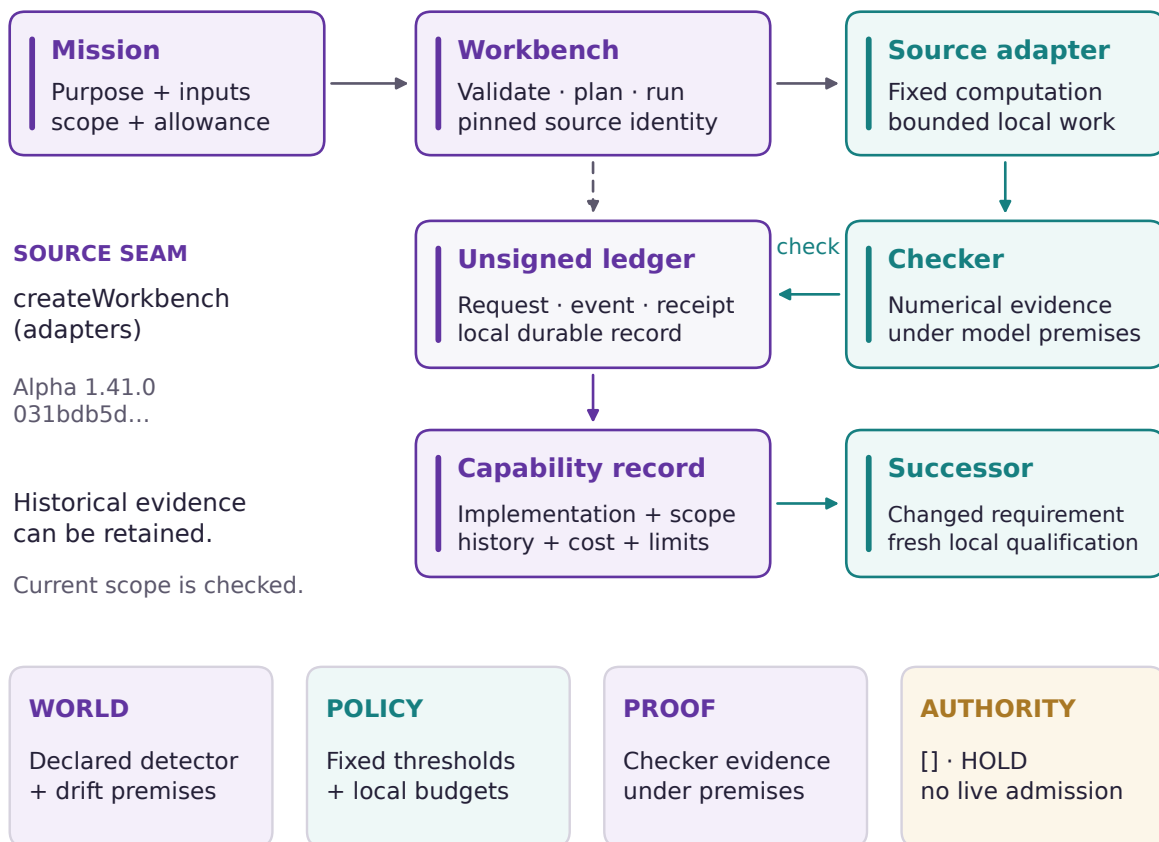


**Figure 1.** The executed source-extension boundary. The released workbench supplies mission/result commitments; the research adapter and data-only ledger add the bounded detector journey. WORLD and POLICY enter the local checker, while AUTHORITY stays separate. Conceptual arrows are implemented only within the stated local scope; they do not imply remote Fleet execution or institutional admission.

**Table 2.** Implemented interfaces and records. The complete schemas and exact file identities ship with the companion. Every record is research evidence; none is a live financial authorization.

| Interface              | Input and output                                                                                                      | Persisted meaning                                                                          |
|------------------------|-----------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|
| Mission specification  | Declared model, functional, rational observations, threshold, evaluation/read limits; optional artifact and ancestor. | Exact request, input rights, separate limits, and source binding.                          |
| Constructor            | Constant or affine model; bounded search or expert formula. Returns matrix, candidate outcomes, and operation counts. | Executable artifact plus unsuccessful candidates; no automatic approval.                   |
| Independent checker    | Artifact, current request, and result. Recomputes identities, interval, quality, and decision.                        | Conditional numerical evidence under the named premises.                                   |
| Workbench adapter      | Fixed source-bound worker input and a validated result pack.                                                          | Artifact hashes, source digest, zero-funded draft, empty authority, and HOLD.              |
| Research ledger        | Tenant label, request key, exact specification, and source digest.                                                    | SQLite state, exact stdout bytes, attempts, event chain, local review context, and export. |
| Successor construction | Frozen parent reference and a changed nuisance requirement.                                                           | New matrix and fresh check; parent history retained separately.                            |

jobs. Detector acceptance comes from the fixed adapter and its checker; the draft mismatch is an integration gap retained in the raw evidence, not a settlement claim. Autonomous operation here means executing a fixed bounded procedure and retaining its results after a request. Humans still choose the purpose, model, acceptance policy, and any move toward physical operation.

### 3.1 Minimal interfaces and persisted state

The worker computes a SHA-256 identity over the actual adapter, constructor, checker, worker, and released module bytes. The ledger additionally retains the exact request and exact stdout hash. These hashes bind bytes. They do not authenticate a laboratory, establish the truth of a reference observation, or prove that an adversary could not rewrite an unsigned local database.

The research ledger is a small SQLite data record around the released interface. Using the native signed Journal would create an Ed25519 identity and API token; this credential-free edition deliberately does not initialize it. Existing domain-specific project stores do not accept this detector schema. The choice is therefore explicit: source-extension execution with unsigned local persistence, not a claimed installed enterprise facade.

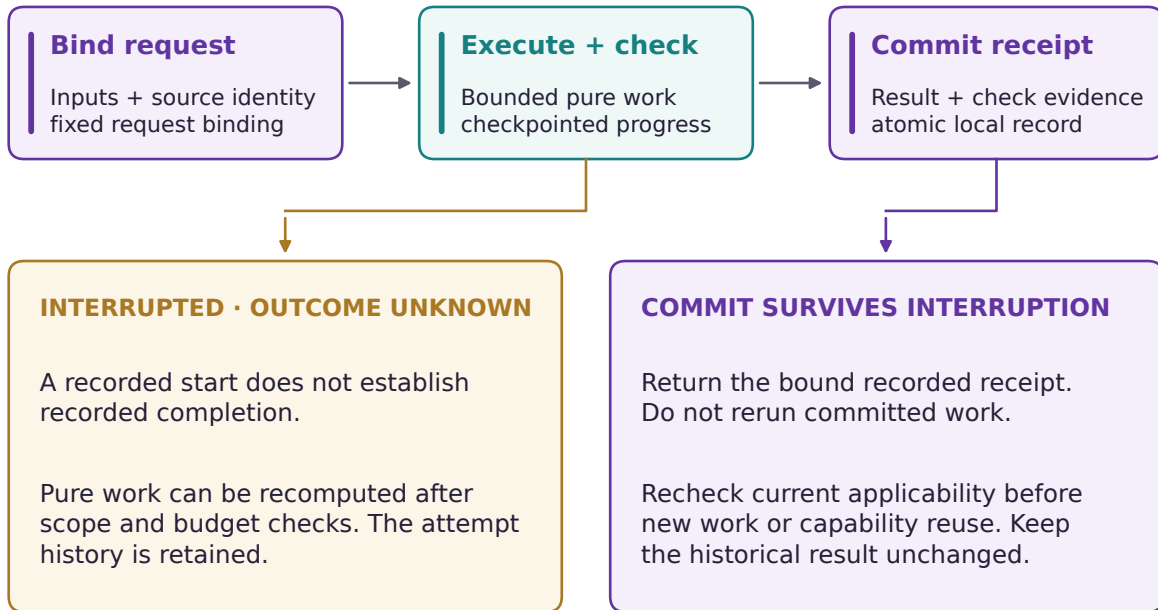
### 3.2 States, retries, and uncertain outcomes

The ledger uses SQLite WAL mode and transactions ([SQLite, 2026](#)). It records **planned**, **running**, **result**, **rejected**, **cancelled**, **timeout**, or **unknown**. A request key is bound to both its canonical input and source digest. Reusing a key with changed inputs or source fails before execution. Competing callers cannot both own the same active attempt.

Completed retries return the retained bytes. They do not require the current worker to execute, do not reissue the original computation, and do not refresh a review. If the process dies after computing but before committing its receipt, the result is uncertain. The next inspection pre-

## Recovery follows what is durably known

Local pure computation · unsigned request records · no external side effects



External effects need an outcome query or an idempotent sink.  
They are outside this local experiment.

**Figure 2.** Recovery distinguishes a retained completion from an uncertain outcome. Actual process-death tests exercise the local durable boundaries. Repeating pure computation after an unknown outcome is permitted explicitly; no general exactly-once guarantee is claimed for a physical actuator, browser action, payment, or remote provider.

serves that uncertainty. An explicit resume may repeat this *pure local computation*; the attempt history remains and the repeated work is not represented as free.

This is a local process-failure claim, not a distributed atomicity theorem or a power-loss certification. RIFL makes the required relationship between effects and durable completion records explicit; durable workflow guidance likewise requires idempotent effect handling (Lee et al., 2015; Tenzer and Smith, 2024). The companion has no external effects to reconcile. A future instrument adapter would need effect identifiers, device receipts, read-back or reconciliation, and a defined response to an unknowable physical outcome.

### 3.3 Bounded tools, local review, and restoration

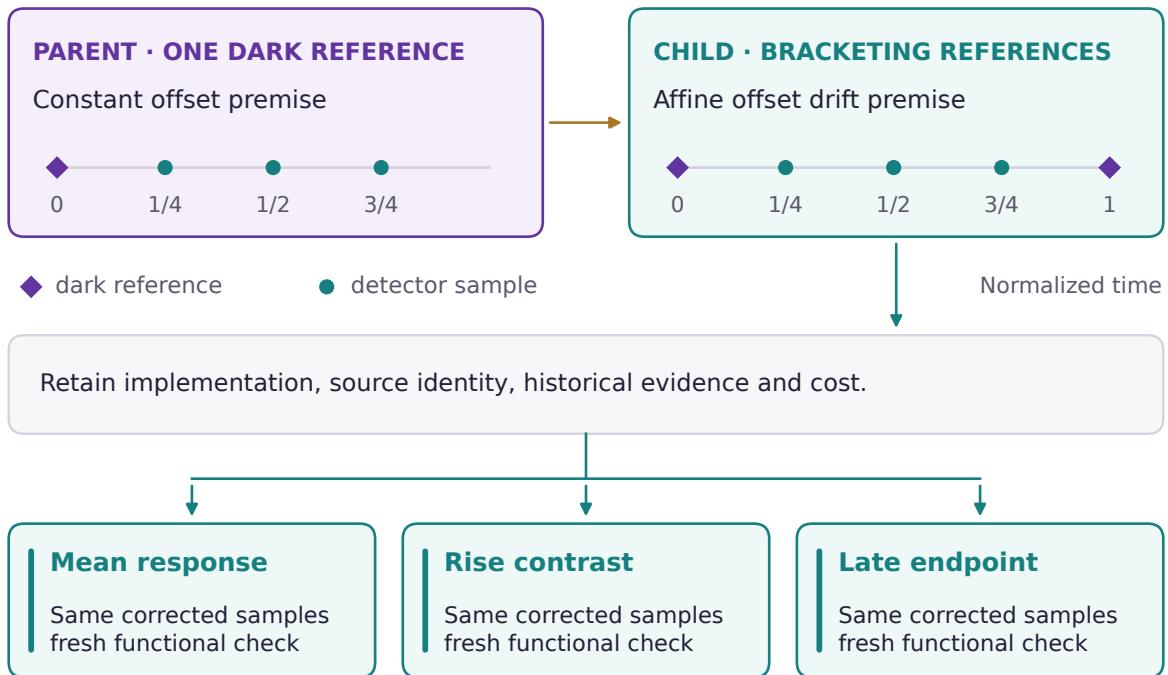
The production worker path is fixed; mission inputs cannot choose an executable, import path, arbitrary code, command, or provider. The driver scrubs inherited environment variables apart from PATH and bounds input, output, attempts, and wall time. Construction allows 4,096 counted rational operations; evaluation allows at most 512; each checker invocation allows 2,048. A parent check is a separate charged invocation. The released workbench also imposes its own one-minute cooperative deadline, while the subprocess driver has a 65-second deadline.

The UI binds only to loopback and permits its fixed local operations. It does not implement authenticated multi-tenant service security. Tenant labels separate local idempotency namespaces, but the operator with filesystem access remains in the trust base. There is no arbitrary execution mode, remote executor, wallet, or live provider hidden behind a simulation switch.

## Reusable construction, fresh qualification

Fixed source experiment · generic detector units · no physical heat claim

**New requirement: affine drift**



**Historical evidence transfers. Active authority remains [] / HOLD.**

**Figure 3.** A parent with one dark reference becomes a starting point for a child with two. The child reuses the declared signal layout, representation, and retained history; it freshly constructs and checks its coefficients. The resulting interface serves three different functionals. No production authority or universal calibration claim crosses the lineage boundary.

A local review is bound to exact result, WORLD, POLICY, and checker identities with an expiration. Inspection without a matching current context does not restore review eligibility. An imported record is historical; it grants neither current review status nor authority. Schema version one rejects incompatible future formats before mutation. A future migration must retain its original export and acquire its own tested semantics.

The separate checker has an independently written rational implementation and does not call the constructor's arithmetic. This is algorithmic separation, not organizational independence. All code and research review were produced within the same project workflow. The workbench's `independentProof=false`, `admission=HOLD`, and empty authority remain unchanged. A result-level checker is useful without being proof of its own correctness or of the world model (McConnell et al., 2011).

### 3.4 What actually transfers

The constructor itself receives the target model, not a magical inherited intelligence state. Its child matrix is freshly solved. The driver retains a real preceding parent receipt and passes its artifact for compatibility checking; the new run records that identity and its new qualification cost. The seven named stage records trace a fixed pipeline. They are not independently learned

strategies, and a single lineage record is not evidence of succession until the subsequent mission actually consumes the retained capability.

## 4 A frozen protocol with strong alternatives

The scientific protocol and complete experiment-source closure were frozen before the first primary request or adversarial campaign was generated. The timestamped record and its SHA-256 were retained durably and posted to the existing coordination issue before execution. The protocol is an internal prospective commitment, not an externally registered or blinded trial. The public formulas, finite grammar, and two development examples were available to the implementers. [E4]

### 4.1 Workload, outcomes, and stopping rules

The primary bank contains 16 profiles, two nuisance classes, and three functionals: 96 ordered requests. Constant-offset profiles have signal means  $1000 + 25(k \bmod 5)$ ; affine-drift profiles have means  $100 + 5(k \bmod 5)$ . Their first-to-last rise is  $100 + 10(k \bmod 4)$ . Offset, drift, bounded observation errors, and fixed threshold values vary by the published formulas. The weaker successor signals make the new scope a meaningful requirement without implying a new physical discovery.

Every runtime input contains only the declared model, requested functional, observed values, policy, and limits. Hidden generating parameters and true signals remain evaluator-side. The checker cannot use those truths to select an operator, accept a result, or conceal a model failure. A separate posthoc process compares completed receipts against the exact generated truth; its computation is charged separately.

The predeclared quality rule requires a finite deterministic relative bound of at most 5%. A threshold decision is an additional outcome. The study reports quality-qualified intervals, resolved decisions, true-value containment, and false assertions separately. On a malformed input, invalid scope, missing dependency, exhausted operation/read budget, or failed check, the worker retains a refusal or incomplete result. The study never loosens the threshold after seeing a failure.

Four controllers share the fixed worker, independent checker, source binding, ledger, persistent observation cache, and full-coordinate interface:

1. **Bounded enterprise:** search the acquisition grammar once for each applicable model, then reuse checked artifacts and compatible observations.
2. **Conventional cached implementation:** ordinary model-keyed memoization with the same construction, persistence, selective validity checks, and cache. It receives no artificial restriction on reuse.
3. **Rebuild ablation:** repeat construction before every request while retaining the same observation cache and required parent history. This isolates construction reuse; it is not the strongest conventional comparator.
4. **Direct expert, cached:** use the analytic matrix for the same interface, retain it, and pay the same checking and execution costs.

The serial campaign runs all 96 in original order. A fresh four-worker repetition first performs one parent and one child request as counted preparation, then queues the remaining 94. Each of the eight scenarios has its own persistent cache and ledger. Preparation is included in wall

**Table 3.** Primary result, per 96-request serial campaign. All four arms also attain the same coverage at the frozen budget of 50,000 rational calls and 160 new simulated acquisitions.

| Controller            | Quality | Resolved | Calls  | New reads |
|-----------------------|---------|----------|--------|-----------|
| Bounded enterprise    | 96      | 78       | 32,675 | 144       |
| Conventional cached   | 96      | 78       | 32,675 | 144       |
| Rebuild ablation      | 96      | 78       | 37,422 | 144       |
| Direct expert, cached | 96      | 78       | 32,583 | 144       |

Every arm contains the exact truth in 96/96 intervals and makes zero false threshold assertions within the declared model. Each makes 432 read requests: 144 new logical acquisitions and 288 cache hits. No physical reads occurred.

time. Actual worker concurrency, queue delay, service latency, failures, and retained receipts are observed; four slots do not imply four CPUs or a guaranteed speedup.

Primary resource coverage is a predeclared prefix projection with 50,000 counted rational operations and 160 new simulated acquisitions. A fixed grid of operation budgets  $\{1000, 5000, 10000, 25000, 50000, 100000, 250000\}$  exposes sensitivity. Prefixes of 6, 24, and 96 requests reuse the same actual observations; they are not extra experiments. The projection charges a channel at its first requested use in the original order, including for parallel runs. It does not claim that a separately stopped campaign was executed.

## 4.2 What the cost ledger includes

The runtime sum has five nonoverlapping components: constructor search, current instrument qualification, parent qualification, evaluation, and independent result checking. Failed constructor candidates remain inside construction cost. Rational add/subtract/multiply/divide/compare calls are instrumented; GCD remainder counts are separate. Logical requests, new simulated acquisitions, and cache hits are separate from these arithmetic operations.

Elapsed time includes worker startup, source hashing, the workbench, persistence, and local review recording within the measured scope. It does not price engineering effort, human review, hardware, electricity, financing, or maintenance. The primary study's local review records are made by the automated controller, not human participants or independent governors. Human active time and the price of prior expert knowledge remain unknown.

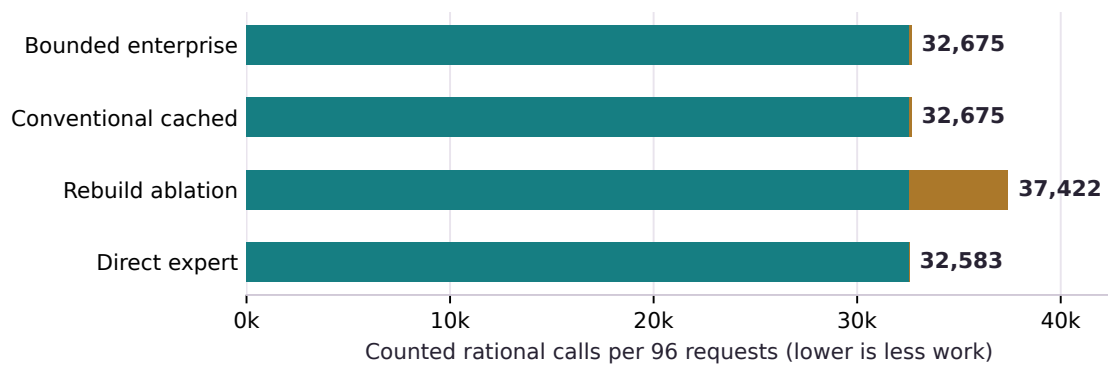
Deliberate challenges add stale scope, changed coefficients, absent reference data, resource exhaustion, zero and weak signals, invalid ancestry, hidden common gain, and curved drift. Three process-death drills and the remaining recovery tests are separate from the scientific bank. The independent Python reference reconstructs the prescribed cases and calculations with `fractions.Fraction`; it imports none of the JavaScript arithmetic or ledger modules and requires coverage of all eight 96-case groups. [E5,E6,E7]

## 5 Results: what became possible, and what did not

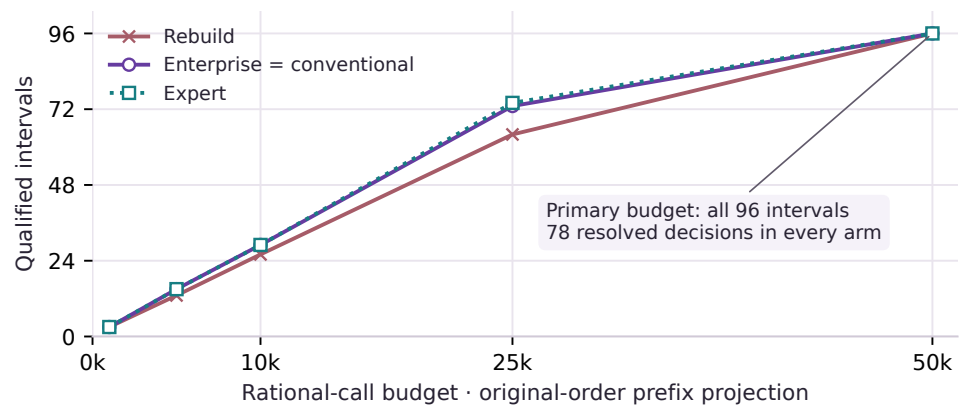
All 768 primary and parallel worker invocations completed. In each of the eight scenarios, all 96 intervals met the conditional quality rule and contained the exact generated truth; 78 threshold decisions were resolved and none was false. Eighteen remained unresolved because their intervals crossed the requested threshold. The primary budget supported all 96 quality-qualified intervals and all 78 resolved decisions for every controller: there was no coverage advantage at that budget. [E5]

## A Full measured software work

- Execution + all current checks
- Construction, including failures



## B Qualified coverage under a constrained budget



Executed software · simulated observations · labor, money and physical validity unmeasured

**Figure 4.** Finite-workload comparison. All arms receive the same full-coordinate task interface, retained observations, persistence, and validation. Cost reduction against rebuilding isolates construction reuse. A conventional tie and expert advantage remain visible. Coverage is a predeclared resource-prefix projection, not a claim of revenue, demand, physical throughput, or a separately stopped campaign.

**Table 4.** Counted rational calls, including construction, ancestors and current validation. The four-worker repetition produces the same logical counts.

| Component        | Enterprise | Conventional | Rebuild | Expert |
|------------------|------------|--------------|---------|--------|
| Construction     | 101        | 101          | 4,848   | 9      |
| Instrument check | 5,184      | 5,184        | 5,184   | 5,184  |
| Parent check     | 2,304      | 2,304        | 2,304   | 2,304  |
| Evaluation       | 11,823     | 11,823       | 11,823  | 11,823 |
| Result checker   | 13,263     | 13,263       | 13,263  | 13,263 |
| Total            | 32,675     | 32,675       | 37,422  | 32,583 |

The count is a defined software-work metric, not CPU instructions, bit complexity, labor, dollars, energy, or a complete economic cost. Failed candidate work is included.

## 5.1 Reuse compounds; conventional engineering can share the gain

Retained construction uses 32,675 counted rational calls, compared with 37,422 when reconstructing before each request: 4,747 fewer calls, a 12.69% reduction. The conventional cached implementation ties it exactly. Direct expert construction uses 32,583 calls, 92 fewer than search-and-reuse. Search considers six candidates across the parent and child and retains three rejections; rebuilding considers 288 and retains 144 rejections. The expert directly constructs two valid operators. All arms issue 432 logical reads, with 144 new simulated acquisitions and 288 cache hits.

At a tighter predeclared 25,000-call projection, enterprise and conventional reuse qualify 73 intervals and resolve 60 decisions; rebuilding qualifies 64 and resolves 54; the expert qualifies 74 and resolves 61. These are resource-prefix projections from completed work, not separately executed truncated campaigns. If every cached read is charged against the 160-read ceiling, the fixed-budget coverage falls to 35 quality-qualified intervals and 29 resolved decisions for all arms. Reuse benefits depend on the resource actually constrained.

The comparison does not establish a special mathematical advantage from the enterprise label. The two caching policies use the same scientific primitives. Their equality is useful evidence: the proposed mechanism survives a competent ordinary implementation, but does not outperform it on the counted work. The direct expert's formula avoids search and remains a valid alternative. Reuse should therefore be chosen for its measured and operational benefits, not because the paper's title requires it to win.

## 5.2 The successor changes scope, not just its name

The parent operator is rejected for affine scope because it leaves a nonzero drift moment. The child adds a second reference, constructs the interpolation coefficients, and pays for fresh checks. Its accepted representation then supports mean response, rise contrast, and late sampled level. What expands is the set of declared-model questions that can be answered within the stated quality contract, relative to the original qualified operator. This is not an impossibility claim for an expert starting from scratch: the direct expert constructs the child independently and more cheaply. No new physical instrument, company, market, or demand was created.

The actual cache reuses observations across functionals within each model/profile. The changed-model child acquires a new simulated stream; it does not inherit the parent's numeric readings. A separate direct-child challenge succeeds without a parent in 426 counted calls, compared with 474 when the parent is supplied. The extra 48 calls check ancestry. Lineage is useful evidence, not a prerequisite for this analytic construction or an automatic cost saving.

The cross-run history retains the first parent’s artifact identity and receipt, the failed use under changed conditions, and the separately constructed child. A parent check is charged whenever supplied. Importing a record retains the evidence without renewing its review context or authority.

### 5.3 The failure remains observable in the evidence

The original challenge campaign stopped on its second call because a generated display name exceeded the released SDK’s 48-character limit. Its successful first receipt, error, 303 counted runtime calls and eight posthoc calls remain intact; partial work in the failed call is unknown. Amendment 01 shortens only the display-name construction and adds an explicit provenance binding. It does not change the worker, checker, hypotheses or acceptance rules. The corrected campaign executes all 19 planned records. Satisfying the challenge protocol includes retaining failures the detector cannot see; it does not mean every measurement is valid. [E6]

For the common-gain challenge, the complete observations are

$$(80, 140, 200, 20, 60).$$

They are consistent with signals (50, 100, 150) at gain one, or (46, 92, 138) at gain 25/23, with the same offset and affine drift. The runtime sees one stream. The mean estimate is 100 with interval [98, 102]; the second world’s true mean is 92. The claimed conditional bound is valid only under unit gain, and the hidden violation escapes the dark-reference check.

Curvature supplies a different failure: adding  $48t(1 - t)$  leaves endpoint references unchanged. It changes the mean and late-level result but cancels in this particular symmetric rise contrast. Thus a passed functional is not a general calibration certificate. The failure map in Figure 5 separates covered failure classes from undetectable premises.

### 5.4 Recovery and workload scaling

The retained local recovery suite passes all 17 test methods, including separate process deaths before worker execution, after worker return but before receipt commit, and after receipt commit. Those tests use a fixed synthetic worker to isolate the ledger. The browser journey separately invokes the actual detector adapter and recovers the same committed receipt after a deliberately lost response. Changed request/source identities, malformed or oversized output, timeout, cancellation, stale review context, expired review, and altered exports are exercised in the bounded local scope. The exact tests and receipts are retained; this is not a remote-device or distributed-payment recovery result. [E7,E8]

On the primary host, the enterprise campaign takes 13.021 seconds serially and 6.210 seconds with four workers. Its service p95 rises from 148 to 333 milliseconds under concurrency. The four-worker rebuild takes 5.721 seconds and the conventional cached arm 5.896 seconds, both faster than the enterprise in this observation. Logical efficiency therefore does not imply lower observed latency. No human review workload or reviewer scaling was measured.

The four-worker comparison is one finite repetition on the recorded host. It tests increased concurrent workload and the same cache/validation obligations; it does not estimate a population reliability rate. Latency differences can reflect scheduling, filesystem contention, process startup, and host load. Long-term maintenance, multi-machine scaling, reviewer capacity, and physical acquisition throughput remain unmeasured.

**Table 5.** Challenge outcomes after Amendment 01. Quality denotes the declared-model relative-error rule of at most 5%; it is not physical validity. Containment is checked only by the posthoc evaluator. [E6]

| Case                          | Quality | Truth in interval | Observed behavior                  |
|-------------------------------|---------|-------------------|------------------------------------|
| Constant-offset parent        | Yes     | Yes               | Conditionally qualified            |
| Affine-drift successor        | Yes     | Yes               | Conditionally qualified            |
| Parent used in affine scope   | —       | —                 | Scope refused                      |
| Altered coefficient proposal  | —       | —                 | Coefficient check refused          |
| Missing reference observation | —       | —                 | Missing data refused               |
| Read-request limit: 3         | —       | —                 | Budget exhausted                   |
| Evaluation-call limit: 1      | —       | —                 | Budget exhausted                   |
| Zero functional               | No      | Yes               | Relative accuracy withheld         |
| Child without a parent        | Yes     | Yes               | Direct child; no inheritance claim |
| Invalid supplied parent       | —       | —                 | Ancestry check refused             |
| Hidden gain: mean             | Yes     | <b>No</b>         | <b>Undetected model failure</b>    |
| Curved drift: mean            | Yes     | <b>No</b>         | <b>Undetected model failure</b>    |
| Hidden gain: rise             | Yes     | <b>No</b>         | <b>Undetected model failure</b>    |
| Curved drift: rise            | Yes     | Yes               | This curvature cancels             |
| Hidden gain: late level       | Yes     | <b>No</b>         | <b>Undetected model failure</b>    |
| Curved drift: late level      | Yes     | <b>No</b>         | <b>Undetected model failure</b>    |
| Weak signal: mean             | No      | Yes               | Relative accuracy withheld         |
| Weak signal: rise             | No      | Yes               | Relative accuracy withheld         |
| Weak signal: late level       | No      | Yes               | Relative accuracy withheld         |

A dash denotes no completed numerical result. For hidden gain, containment refers to world B; the same observed stream contains world A's truth. **Five conditional passes miss the true value:** three gain cases and the mean/late-level curvature cases. The rise cancels this particular curvature; this is not drift detection. Corrected execution uses 5,637 SDK rational calls, 120 posthoc calls and 71 logical read requests. The original failed invocation remains separate: 303 SDK calls, eight posthoc calls and four reads. Its unreturned failure overhead remains unknown. The one-call limit applies to evaluation; qualification and checking have separate frozen allowances.

**Table 6.** One observed serial run and one four-worker repetition per arm. Wall time includes counted preparation; service latency excludes queueing, whereas end-to-end latency includes it.

| Controller            | Serial (s) | Four (s) | Service p95 | E2E p95 |
|-----------------------|------------|----------|-------------|---------|
| Bounded enterprise    | 13.021     | 6.210    | 333         | 5,732   |
| Conventional cached   | 13.031     | 5.896    | 319         | 5,452   |
| Rebuild ablation      | 14.483     | 5.721    | 298         | 5,203   |
| Direct expert, cached | 13.710     | 6.399    | 332         | 5,861   |

Latency columns are milliseconds. All 96 requests complete in every scenario. Rebuilding is fastest in the four-worker observation, despite requiring more rational calls. Host scheduling and startup affect these finite measurements.

## 5.5 Reproduction and cumulative cost

The separately implemented Python Fraction reference checks all 768 retained packs across all eight required groups with no failed comparison. It makes 2,592 fixture-construction calls and 110,568 evaluation/check calls in its own counter; those units are not assumed interchangeable with the JavaScript counter. The primary JavaScript campaign itself uses 270,710 runtime rational calls, plus 23,328 evaluator fixture-generation calls and 6,360 posthoc truth-comparison calls. Those evaluator costs are outside the plotted runtime budget and remain explicit. [E5]

A fresh archive extraction independently executes all 768 cases, the Python reference and the amended challenges. All scientific outcomes and counted costs reproduce. The first comparator nevertheless flags 19,200 differences in generated expiry fields and dependent commitments. That failed comparison is preserved. An additive correction verifies the original commitments and seven-day expiry rule before normalizing them for comparison. It matches 768 scientific cases and 768 normalized packs with zero differences, while four negative checks detect changed results, authority, expiry policy and source. The experiment itself is not rerun to fix this comparator. Receipts are not byte-identical. [E8]

The separate aggregate account retains 615,520 known JavaScript rational calls across primary execution, reproduction, challenges and recorded development, plus 226,320 Python Fraction calls under their separate metric. This is additional research work, not a single treatment cost. Unreturned failures, uninstrumented overhead, authorship, review time, energy, hardware, maintenance and realized economic value remain unknown or unmeasured; they are not reported as zero.

All primary raw receipts remain separate from later reproduction, UI development, numerical-package smoke, and failure-test evidence. Reproduction recomputes software results; it does not recreate the original prospective withholding. The package records exact tool versions, file identities, commands, and the comparison rule used for a rebuild. No success flag replaces a missing evidence file.

## 6 The ecosystem and its authority boundaries

The companion exercises a released source extension interface and local persistence. Table 7 separates that execution from installed qualification and deployment. The current coordination record accepts the historical Alpha 1.39/Jobs rc.20/Manager 1.20.1 API tuple; it establishes neither an accepted Alpha 1.41 tuple nor four installed products (MONTREAL.AI, 2026f). [E2]

The release records support these bounded identities (MONTREAL.AI, 2026a,b,h,i,g). The Workbench invokes vendored Jobs compiler/template code. Its seven unfunded packages retain inherited aggregate-template wording; they are inert drafts, not commissioned detector acceptance jobs. Their presence establishes code reuse, not a Jobs service integration. This edition performs no financial transaction, token approval, contract deployment, or local-EVM lifecycle run. Its arithmetic examples below are source-derived, not transaction receipts.

### 6.1 What the inspected five-contract suite supports

The five contract files at historical commit c0fc5217 and inspected commit 031bdb5d have identical Git blobs (MONTREAL.AI, 2026d,e). The deployment record's chain-1 observations date to 18 October 2026, block 26,220,072. It records activation and public funding as unapproved. Source equality, recorded explorer matches, and the absence of a global pause do not establish present

**Table 7.** Four repositories, distinct evidence. Commit prefixes resolve to the full identities in audit E2. Published qualification is not a fresh installation in this experiment.

| Repository | Inspected source / role                                 | Installed evidence                                                             | Deployed or live boundary                                          |
|------------|---------------------------------------------------------|--------------------------------------------------------------------------------|--------------------------------------------------------------------|
| Alpha      | v1.41.0, 031bdb5d; source SDK used                      | Released wheel’s numerical family invoked; detector remains a source extension | Five-contract record dated 18 October; no fresh chain verification |
| Jobs v0    | Discovery rc.20, f7c00e66; tooling main 9983478d        | Published local-review package; not installed here                             | Separate AGIALPHA backend; no transaction here                     |
| Manager    | v1.20.1, c5958eae; separate receipt route               | Published package; not installed here                                          | Separate USDC backend; no transaction here                         |
| Fleet      | Local Skills preview.5, 45759fb0; execution/review role | Published development preview; not installed here                              | No connected Fleet execution or review here                        |

bytecode, proxy dependencies, role eligibility, balances, or operating readiness. [E2] These are Alpha’s Ascension contracts; the separate AGIJobsv0 suite requires its own pause, admission, collateral and stake checks.

**Access and NovaSeed.** `AscensionAccess.register` checks ownership beneath the source-defined roots: `business *.alpha.agi.eth`, `agent *.alpha.agent.agi.eth`, and `validator *.alpha.club.agi.eth`. These are distinct from `*.agent.agi.eth`. Eligibility additionally requires unexpired, governor-set admission; wrapped names require unexpired ownership. Token holding alone grants no role. `NovaSeed.seal` records capsule, plan-root and evidence commitments, job count, creator, metadata URI and optional parent whose token the caller owns. The ERC-721 records ownership of that commitment-bearing token; plan contents and executable capability remain off-chain. It conveys no automatic intellectual-property or equity transfer.

**Risk oracle.** `openReview`, `attest`, and `green` implement seven-day, owner-bound review rounds. Validators require current Access eligibility and minimum stake; owner/creator self-review and duplicate votes are excluded. The dated record gives quorum two, a 2,000-basis-point risk ceiling, and minimum stake of ten AGIALPHA tokens. An affirmative vote also requires policy approval and an evidence hash. Approval expires and rechecks owner eligibility and every supporter’s eligibility. Address exclusions do not establish organizational independence. Job review additionally reserves validator availability and capacity.

**Job market.** The source implements `post`, `bid`, `award`, `submit`, `validate`, `timeout`, and `claimRefund`. Reputation-weighted selection *does exist*. For price  $p$ , bounty  $B$ , duration  $d$ , maximum duration  $D$ , price weight  $w \in [0, 10^4]$ , and reputation  $r$ , the integer score is

$$\begin{aligned}
 C &= \left\lfloor \frac{10^6 p w}{B} \right\rfloor, & T &= \left\lfloor \frac{10^6 d (10^4 - w)}{D} \right\rfloor, \\
 S &= \left\lfloor \frac{10^4 (C + T)}{10^4 + \min(r, 10^4)} \right\rfloor.
 \end{aligned} \tag{4}$$

The lowest eligible score wins; ties use price, duration, then address. Reputation is snapshotted at bidding. This guarantees neither cheapest nor fastest delivery. Three distinct eligible validators form the committee; client/business addresses and the worker are excluded. Two artifact-bound affirmative votes pay; two negative votes fail. Eligibility is rechecked at settlement (MONTREAL.AI, 2026e).

For accepted price  $p$ , each qualifying validator’s gross share is  $v = \lfloor p/40 \rfloor$ , and worker gross is  $p - 2v$ . Each recipient separately burns  $\lfloor g/100 \rfloor$  from gross  $g$ . At  $p = 100$  base units, the

gross amounts are 96, 2, 2 and all three burns are zero; a blanket one-percent deduction would be wrong. Unspent bounty is refundable. Rejection or missed delivery slashes the bond with its separate burn; missing review quorum refunds without slashing. Terminal refunds require the client's single claim.

**MARK.** `open/buy/sell/transferLots` implement admitted participation in a reversible lot curve. Buying  $n > 0$  lots at supply  $s$ , base  $b$ , and slope  $m$  costs  $nb + mn(2s + n - 1)/2$ . `bloom` requires the target reserve and current green owner. `route` verifies an indexed FusionPlan Merkle leaf, remaining reserve and execution horizon; this plan binding does not apply to every direct market posting. `collect/close/reclaim` return applicable unused reserves. Funding-phase selling remains available after admission revocation or risk expiry. This is a specific lot/funding mechanism, not a generic DEX or arbitrary reinvestment engine. Its canonical 18-decimal AGIALPHA route remains distinct from Manager's six-decimal USDC route.

## 6.2 Unresolved integration retains its owners

PR 4863 retains its original review/merge restriction; PR 4865 depends on its legitimate integration. PR 4871's native thermal outcome remains held. The new detector extension clears none of them. Fleet's canonical integration also retains its restriction and unresolved findings. Original restriction operators must provide the missing dispositions; eligible exact-source reviewers and the existing integrators/release owners then complete the applicable qualification. The coordinator in issue 4850 retains integration order and compatibility tracking. Passing local checks, new branches, or source citations cannot replace those decisions ([MONTREAL.AI, 2026f](#)). [E2]

## 7 Economic usefulness requires more than fewer operations

The demonstrated productive asset is a reusable detector procedure with a checked, bounded interface. It supports further questions and a changed nuisance class. That is a mechanism by which achievements can expand future usefulness. Its realized economic value is not measured in this study: no customer paid, no purchase occurred, no physical instrument produced a service, and active human time, energy, engineering labor, and maintenance costs were not fully metered.

The measured operation counts support a narrower statement: retaining a valid artifact can avoid repeated construction under this workload, after charging its ancestor and current checks. A stronger conventional workflow can obtain the same advantage. The expert's direct formula can be cheaper. Acquisition counts describe simulated detector operations, and elapsed times describe this host. Neither is a bill or a valuation.

### 7.1 A transparent adoption condition

For a future service, let  $v_i$  be a beneficiary's independently observed willingness to pay for an accepted outcome;  $a_i$  its acquisition and execution cost;  $r_i$  its review, failure, and recovery cost; and  $F$  the complete construction, calibration, financing, and maintenance allocation. A necessary operating condition is

$$\sum_{\text{accepted } i} v_i > F + \sum_i (a_i + r_i),$$

subject also to delivery times, capacity, rights, and voluntary participation. This inequality is an accounting condition, not an estimate of the terms. Withdrawals, low utilization, calibration drift, unreliable suppliers, or a cheaper expert alternative can defeat it.

A simple capacity check is equally necessary. Useful throughput cannot exceed the least available capacity among acquisition, execution, review, and demanded accepted work. Adding software workers does not add reference standards, reviewer attention, cooling, power, laboratory hours, or buyers. The four-worker experiment measures a finite local workload; it does not validate the scaling economics of a fleet or manufacturing line.

The commercial next step is small and falsifiable: one buyer-defined measurement decision, a written acceptable error and delivery time, an actual expert quote, a full uncertainty budget, and a priced alternative. If the new capability cannot beat or complement that alternative under the buyer's constraints, retain the incumbent.

## 7.2 From digital assets to physical capacity

Table 8 describes necessary constructions at each transition. Each row adds a different engineering and economic obligation; none follows automatically from copying the previous software.

**Table 8.** A conditional development path. Every transition is a testable construction, followed by a separate decision to proceed. No row after the digital companion was physically executed here.

| Transition                         | What must be built and supplied                                                                                                                                     | What must be validated                                                                                                                                     | Economic condition                                                                                        |
|------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|
| Digital procedure to instrument    | Detector, shutter or zero-reference path, stable nonzero reference, timed acquisition electronics, fixtures, calibrated standards, operator time, electrical power. | Traceable gain and offset calibration, drift envelope, correlated errors, timing, repeatability, environmental effects, physical fault and reset behavior. | A useful decision whose value covers construction, calibration, consumables, downtime, and expert review. |
| Instrument to scientific discovery | Sample preparation, controlled stimuli, material supply, reference specimens, suitable containment, and an experimental campaign.                                   | Identifiability of the scientific quantity, controls, independent replication, out-of-model stresses, and uncertainty appropriate to the claim.            | Demand for the knowledge or service; sufficient utilization and an advantage over an existing laboratory. |
| Discovery to manufacturing         | Process equipment, tooling, feedstock, assembly, quality control, skilled maintenance, utilities, and supplier capacity.                                            | Yield, tolerances, throughput, safety, durability, repair, waste, and transfer from laboratory conditions to process scale.                                | Unit economics after scrap, rework, inventory, capital, labor, energy, and warranty obligations.          |
| Manufacturing to infrastructure    | Civil works, land access, transport, storage, communications, control systems, grid or utility connections, and long-lived maintenance capability.                  | System reliability, permitting and applicable engineering requirements, interfaces, resilience, operations, and decommissioning.                           | Financing, durable demand, capacity utilization, ownership obligations, and lifecycle cost.               |

| Transition                               | What must be built and supplied                                                                                                                                                  | What must be validated                                                                                                     | Economic condition                                                                                            |
|------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|
| Infrastructure to useful energy capacity | An independently justified conversion or generation technology, materials, energy inputs, siting, power electronics, transmission, thermal management, and storage where needed. | Measured energy balance, efficiency, availability, lifetime, environmental performance, and safe grid or load integration. | Net useful energy and value after all inputs, construction, losses, maintenance, financing, and alternatives. |

The path can compound only if accepted outputs actually help construct and qualify the next means of production. For example, a calibrated measurement procedure could make a process parameter observable; that measurement could enable a better process window; a reproducible process could make a useful component; and verified components could support a larger instrument. Each link requires materials, energy, engineering, and acceptance beyond the software record.

## 8 Limits, rejected interpretations, and the next construction

The primary workload is finite, public, and generated from authored formulas. It is prospectively withheld from constructor execution, not externally blinded. The model contains three signal times and two nuisance classes. Its grammar and functionals are supplied by the research team. Exact rational agreement does not establish the adequacy of those assumptions for a detector in the laboratory.

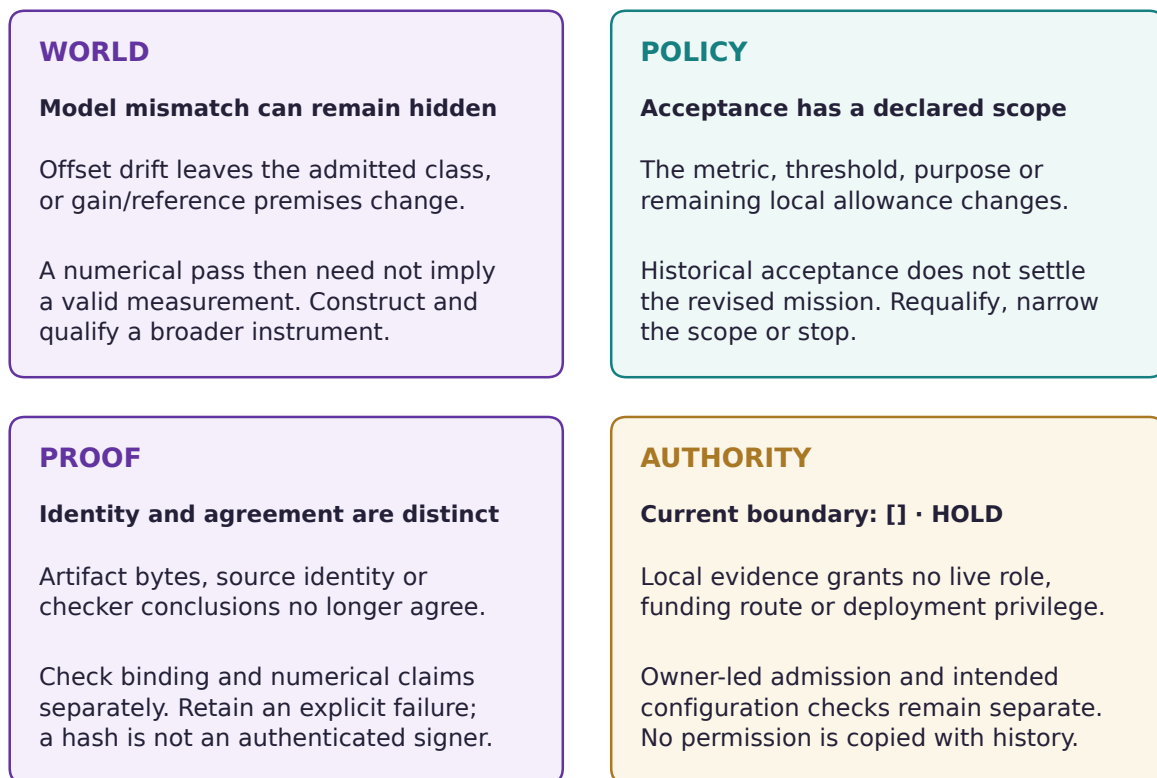
The runtime checker shares the project's authorship and does not independently establish every cost counter. Arithmetic-call accounting excludes bit complexity, some bookkeeping, source hashing, process startup, persistence, and human labor; separate wall times cover some of that software overhead without pricing it. Failed process attempts may have performed unobserved partial computation. Their cost is unknown, not zero.

The conventional comparator deliberately receives the same cache, persistence, validation, and tools. The analytic expert comparator implements the same complete corrected-coordinate interface. It is not an unrestricted global cost optimum: a single-purpose expert can sometimes read fewer channels, such as two signals for constant-offset rise. A broad claim of superior allocation, fastest execution, cheapest work, autonomous governance, or commercial profit would exceed the evidence.

The local ledger provides integrity checking and request attribution within an unsigned research workspace. It is not an authenticated enterprise security boundary. Four-repository remote execution, native detector admission, independent institutional review, current chain readiness, and live settlement remain outside the demonstrated journey. Their owners and restrictions are retained in Section 6.

## Four boundaries of a qualified local result

A numerical pass, an intact artifact and permission to act answer different questions



**Local source qualification does not establish physical validity or institutional approval.**

**Figure 5.** An honest failure map. Identity checks, algebraic checks, and declared budgets cover different failure classes. Gain and curvature counterexamples preserve the gap between conditional model validity and physical truth. Missing institutional, integration, and commercial evidence is not filled by a local green indicator.

**The next smallest useful construction**

Add a calibrated, nonzero reference observation that can test or bound detector gain, then challenge the extended instrument with gain drift and reference failure. Keep the affine child and its failure record. Freeze the new requirement, compare against direct expert construction, and charge the new reference, calibration, qualification, and operating costs. A laboratory pilot follows only after its hardware, operator, and acceptance obligations are separately in place.

The long-term ambition is an organization whose achievements increase its capacity to do useful work. This edition tests one small mechanism: a mission leaves an executable procedure, a changed requirement exposes its limit, and a fresh construction expands what the successor can answer under declared conditions. A finite example proves neither unlimited growth nor a civilization-scale outcome. Its value is that the next construction is specific enough to build, measure, and reject.

## References

- Vincent Boucher. Out-Design: Creating the freedom to transform. MONTREAL.AI research manuscript, October 2026a. URL [https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/031bdb5d055c8bad896daa076baea2b5e17b242c/docs/Out\\_Design.pdf](https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/031bdb5d055c8bad896daa076baea2b5e17b242c/docs/Out_Design.pdf). Revised edition dated 21 October 2026; pinned public source.
- Vincent Boucher. Out-Evolve: Building the means of discovery and production. MONTREAL.AI research manuscript, October 2026b. URL [https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/031bdb5d055c8bad896daa076baea2b5e17b242c/docs/Out\\_Evolve.pdf](https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/031bdb5d055c8bad896daa076baea2b5e17b242c/docs/Out_Evolve.pdf). Revised edition dated 20 October 2026; actual 29-page PDF inspected.
- Vincent Boucher. Out-Execute: Constructing the conditions of completion. MONTREAL.AI research manuscript, October 2026c. URL [https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/031bdb5d055c8bad896daa076baea2b5e17b242c/docs/Out\\_Execute.pdf](https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/031bdb5d055c8bad896daa076baea2b5e17b242c/docs/Out_Execute.pdf). Revised edition dated 21 October 2026; pinned public source.
- Vincent Boucher. Out-Learn: Change what you need to know. constructed experiments and sharp calibration guarantees for reusable learning. MONTREAL.AI research manuscript, October 2026d. URL [https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/031bdb5d055c8bad896daa076baea2b5e17b242c/docs/Out\\_Learn.pdf](https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/031bdb5d055c8bad896daa076baea2b5e17b242c/docs/Out_Learn.pdf). Revised edition dated 21 October 2026; pinned public source.
- Vincent Boucher. Out-Strategise: Transformational intelligence for constructing the conditions of success. MONTREAL.AI research manuscript, October 2026e. URL [https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/031bdb5d055c8bad896daa076baea2b5e17b242c/docs/Out\\_Strategise.pdf](https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/031bdb5d055c8bad896daa076baea2b5e17b242c/docs/Out_Strategise.pdf). Edition dated 20 October 2026; pinned public source.
- Vincent Boucher. Out-Think: Transformational intelligence for inventing the space of reasoning. MONTREAL.AI research manuscript, October 2026f. URL [https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/031bdb5d055c8bad896daa076baea2b5e17b242c/docs/Out\\_Think.pdf](https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/031bdb5d055c8bad896daa076baea2b5e17b242c/docs/Out_Think.pdf). Edition dated 21 October 2026; pinned public source.
- Vincent Boucher. Transformational intelligence: Learning to change the game. MONTREAL.AI research manuscript, October 2026g. URL [https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/031bdb5d055c8bad896daa076baea2b5e17b242c/docs/Transformational\\_Intelligence.pdf](https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/031bdb5d055c8bad896daa076baea2b5e17b242c/docs/Transformational_Intelligence.pdf). Pinned public source, inspected 24 October 2026.
- Vincent Boucher. Foundry v2—computational research edition. MONTREAL.AI research release, tag study-foundry-v2-2026-10-22-r1, October 2026h. URL <https://github.com/MontrealAI/AGI-Alpha-Agent-v0/releases/tag/study-foundry-v2-2026-10-22-r1>. Published 22 October 2026. Executed software, simulated physics and modeled contracts; physical construction and deployment not run.
- Joint Committee for Guides in Metrology. Evaluation of measurement data—guide to the expression of uncertainty in measurement. Technical Report JCGM 100:2008, Bureau International des Poids et Mesures, 2008. URL <https://www.bipm.org/en/doi/10.59161/jcgm100-2008e>. GUM 1995 with minor corrections.
- Collin Lee, Seo Jin Park, Ankita Kejriwal, Satoshi Matsushita, and John Ousterhout. Implementing linearizability at large scale and low latency. In *Proceedings of the 25th Symposium on Operating Systems Principles*, SOSP '15, pages 71–86. Association for Computing Machinery, 2015. doi: 10.1145/2815400.2815416. URL <https://sigops.org/s/conferences/sosp/2015/current/2015-Monterey/printable/126-lee.pdf>.

- Ross M. McConnell, Kurt Mehlhorn, Stefan Näher, and Pascal Schweitzer. Certifying algorithms. *Computer Science Review*, 5(2):119–161, 2011. doi: 10.1016/j.cosrev.2010.09.009. URL <https://www.sciencedirect.com/science/article/pii/S1574013710000560>.
- MONTREAL.AI. \$AGIALPHA Agent 1.41.0. GitHub software release, tag v1.41.0, October 2026a. URL <https://github.com/MontrealAI/AGI-Alpha-Agent-v0/releases/tag/v1.41.0>. Published 24 October 2026; source commit 031bdb5d055c8bad896daa076baea2b5e17b242c. Experimental software release.
- MONTREAL.AI. SUCCESSOR Workbench: Released source extension interface. AGI-Alpha-Agent-v0, examples/successor-workbench/engine.mjs, October 2026b. URL <https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/031bdb5d055c8bad896daa076baea2b5e17b242c/examples/successor-workbench/engine.mjs>. Pinned Alpha 1.41.0 source; inspected 24 October 2026. Git blob d3f20b907b4d57867d9a12f46e4c32fb71a03b21.
- MONTREAL.AI. Alpha ascension five-contract suite: Inspected public source. AGI-Alpha-Agent-v0, contracts/ascension source directory, October 2026c. URL <https://github.com/MontrealAI/AGI-Alpha-Agent-v0/tree/031bdb5d055c8bad896daa076baea2b5e17b242c/contracts/ascension>. Commit 031bdb5d055c8bad896daa076baea2b5e17b242c; inspected 24 October 2026. The five named contract Git blobs match the historical c0fc5217 source; no fresh chain-state verification is implied.
- MONTREAL.AI. Alpha ascension five-contract suite: Historical source. AGI-Alpha-Agent-v0, contracts/ascension source directory, October 2026d. URL <https://github.com/MontrealAI/AGI-Alpha-Agent-v0/tree/c0fc5217e364a39a4cd123d17a8317c06b80f820/contracts/ascension>. Commit c0fc5217e364a39a4cd123d17a8317c06b80f820. AscensionAccess, NovaSeed, AscensionRiskOracle, AscensionJobMarket and AscensionMark; source identity does not establish present operational readiness.
- MONTREAL.AI. AscensionJobMarket.sol: Plan-bound job lifecycle and settlement arithmetic. Pinned Alpha 1.41.0 Solidity source, October 2026e. URL <https://github.com/MontrealAI/AGI-Alpha-Agent-v0/blob/031bdb5d055c8bad896daa076baea2b5e17b242c/contracts/ascension/AscensionJobMarket.sol>. Inspected source identity: Git blob a50a5cc584ccb6c096c6df6aa862aa48d27d9ffe. A source citation is not an integration test or a fresh deployment-state check.
- MONTREAL.AI. Existing delivery continuation and canonical integration owners. GitHub issue 4850, comment 6384166924, October 2026f. URL <https://github.com/MontrealAI/AGI-Alpha-Agent-v0/issues/4850#issuecomment-6384166924>. Posted 24 October 2026 at 12:23:42 UTC; inspected the same day. Current accepted API tuple and unresolved review/integration boundaries; mutable coordination record.
- MONTREAL.AI. AGI Fleet Local Skills 0.1.0—linux development preview. AGI-Fleet-Linux GitHub prerelease, tag v0.1.0-local-skills-preview.5, October 2026g. URL <https://github.com/MontrealAI/AGI-Fleet-Linux/releases/tag/v0.1.0-local-skills-preview.5>. Published 16 October 2026; source 45759fb096faf8988f58ae5ea71248d0f35a97c7. Development preview, not a production-qualified operator installer or completed four-repository integration.
- MONTREAL.AI. Jobs Discovery Installed 2.1.0-rc.20. AGIJobsv0 GitHub prerelease, tag discovery-v2.1.0-rc.20, October 2026h. URL <https://github.com/MontrealAI/AGIJobsv0/releases/tag/discovery-v2.1.0-rc.20>. Published 23 October 2026; source f7c00e668fe31ce1ea907f0006b6c14dc9db5e01. Qualified installed local-review software with zero currency authority; a separate AGIALPHA route.

- MONTREAL.AI. AGIJobManager v1.20.1—shared mission receipt context. GitHub software release, tag v1.20.1, October 2026i. URL <https://github.com/MontrealAI/AGIJobManager/releases/tag/v1.20.1>. Published 24 October 2026; source c5958eae872458c2f3cac48fea0b4bbfad66ac6c. Separate USDC route; no four-repository installed acceptance follows from this release.
- National Institute of Standards and Technology. NIST policy on metrological traceability, 2024. URL <https://www.nist.gov/calibrations/traceability>. Updated 17 June 2024; accessed 24 October 2026.
- George C. Necula. Proof-carrying code. In *Proceedings of the 24th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, POPL '97, pages 106–119. Association for Computing Machinery, 1997. doi: 10.1145/263699.263712. URL <https://homes.cs.washington.edu/~mernst/teaching/6.893/readings/necula-popl97.pdf>.
- Jerome H. Saltzer and Michael D. Schroeder. The protection of information in computer systems. *Proceedings of the IEEE*, 63(9):1278–1308, September 1975. doi: 10.1109/PROC.1975.9939. URL <https://web.mit.edu/Saltzer/www/publications/protection/>.
- SQLite. Transaction. Official SQLite documentation, 2026. URL [https://www.sqlite.org/lang\\_transaction.html](https://www.sqlite.org/lang_transaction.html). Page update dated 18 February 2026; accessed 24 October 2026.
- Barry N. Taylor and Chris E. Kuyatt. Guidelines for evaluating and expressing the uncertainty of NIST measurement results. Technical Report NIST Technical Note 1297, National Institute of Standards and Technology, 1994. URL <https://doi.org/10.6028/NIST.TN.1297>. 1994 edition.
- Temporal Technologies. Events and event history. Official Temporal documentation, n.d. URL <https://docs.temporal.io/workflow-execution/event>. Accessed 24 October 2026; mutable documentation, not a dated research paper.
- Keith Tenzer and Joshua Smith. What is idempotency? and why it matters for durable systems. Temporal technical article, February 2024. URL <https://temporal.io/blog/idempotency-and-durable-execution>. Published 27 February 2024; inspected 24 October 2026. The currently served article includes subsequent updates.
- Nathan Whitehead, Martín Abadi, and George Necula. By reason and authority: A system for authorization of proof-carrying code. In *Proceedings of the 17th IEEE Computer Security Foundations Workshop*, CSFW '04, pages 236–250. IEEE, 2004. doi: 10.1109/CSFW.2004.1310744. URL [https://people.eecs.berkeley.edu/~necula/Papers/sigpcc\\_csfw04.pdf](https://people.eecs.berkeley.edu/~necula/Papers/sigpcc_csfw04.pdf).

## A Reproduction, evidence, and publication boundaries

### A.1 Pinned source and runtime

The executed baseline is Alpha v1.41.0 at commit 031bdb5d055c8bad896daa076baea2b5e17b242c. The companion vendors the seven released JavaScript modules required by createWorkbench; its source manifest and license accompany those bytes. The primary campaign used Linux x86\_64, CPython 3.12.14, and Node 24.22.1. The local driver uses the Python standard library. It requires no wallet, model-provider account, or initialization of the native signed Journal. An existing Node executable must be available on PATH.

The 20-file experimental freeze is retained in protocols/FROZEN.json; its SHA-256 is 3267e737ad0fe9e966acc35d97a09ce6a443d81976bd7a5a1738eda974eb7b76. The protocol hash is a5f44fdc645b81d49747c94a3d0c5dea54259a627f5f5348fe55a5e19e3fe5ad; the actual worker source identity is 3ef2d004f830f870307e2472dfc90e1745708aa69236449d5bb9b384972784bb. The [public freeze record](#) was posted at 13:00:09 UTC on 24 October 2026, before the retained primary campaign began at 13:00:25 UTC. This is an internal prospective freeze with a public generator, not independent preregistration or external blinding.

### A.2 Commands and retained outputs

Run commands from the extracted companion's top-level directory. The local experience starts with:

```
python3 companion/serve.py --state local-project --port 8769
```

Open the printed loopback address. The UI supports input changes, parent construction, stale-scope refusal, successor construction, reuse, pause, receipt export, and restoration. Stop the server with Ctrl-C; the state directory remains. Exports are unsigned historical evidence. Importing them neither renews a local review nor creates production authority.

For a fresh scientific run and independently implemented reference:

```
python3 companion/study.py --output runs/study-001
python3 companion/reference_fraction.py \
  --input runs/study-001 --out runs/reference-001.json
python3 tests/test_recovery.py
node companion/challenges_v2.mjs \
  --out runs/challenges-001.json
```

Use new output names. The study refuses an existing run unless explicitly resumed with `--resume`; it verifies the original frozen source before executing. Resume retains prior attempts and appends reports. An unknown outcome may require repeated pure computation. A committed completion returns its original bytes. These rules do not establish exactly-once effects for hardware, payment, or a remote service.

The extracted-source reproduction wrapper packages its declared source closure, verifies extracted bytes, runs the study and Fraction reference from that extraction, and compares the scientific and accounting outcomes:

```
python3 companion/reproduce.py \
  --output runs/reproduction-001 \
  --compare-primary evidence/primary
```

Its archive, source manifest, commands, raw results, reference report, and comparison remain together. This is a clean extracted-source execution, not a claim of installing four products. Wall times and process identities are observations of the new run, not expected byte-identical outputs.

### A.3 The correction is part of the result

The original challenge invocation is preserved as `evidence/challenges-primary.json`. It completed the parent case, then failed because the successor’s 53-character display name exceeded the released SDK’s 48-character limit. Later challenges therefore did not run. Amendment 01 binds the original failure and an additive runner that uses the existing case identifier as the display name. The scientific cases, models, checker, budgets, and acceptance criteria are unchanged.

The corrected invocation completes all 19 cases and retains five conditional passes outside their true-value intervals under violated premises. Its completion is not universal validation. The separate amendment account includes repeated measured work and leaves unreturned failure work unknown. Neither the frozen runner nor its failed report is replaced. [E6]

### A.4 Evidence docket and scope

**Table 9.** Evidence identifiers used in the paper. The machine-readable docket binds exact files, hashes, scopes, and outstanding records.

| ID | Retained record                      | What it can support                                                                                                      |
|----|--------------------------------------|--------------------------------------------------------------------------------------------------------------------------|
| E1 | Historical Foundry downloads         | Previously reported metrology counterexample and comparator losses; no new thermal execution.                            |
| E2 | Repository and compatibility audits  | Dated release/source identities, contract semantics, existing owners and restrictions. No fresh chain readiness.         |
| E3 | SDK source and numerical receipt     | Executed released extension seam; separately installed numerical family. No installed detector or thermal qualification. |
| E4 | Protocol and freeze                  | Criteria and source fixed before this primary invocation; public chronology.                                             |
| E5 | Primary packs and Fraction reference | Eight 96-case groups, measured accounting and independent numerical recomputation under the declared model.              |
| E6 | Original and amended challenges      | Retained interface failure, correction provenance, declared refusals, and undetected model violations.                   |
| E7 | Recovery-test record                 | Bounded local process-death, retry, tampering, review-context, and restoration behavior.                                 |
| E8 | Browser and reproduction records     | Tested local UI/download scope and separate extracted-source recomputation.                                              |
| E9 | Publication QA and rebuild records   | Final page inspection, citation/cross-reference checks, build identity and delivered-file manifest.                      |

`evidence/CLAIM_DOCKET.json` distinguishes observed, historical, proposed, and pending evidence. `audit/compatibility-manifest.json` separates source inspection, installed numerical execution, and unexercised remote/deployed routes. No financial transaction or local-EVM lifecycle was run in this edition. Complete monetary cost, human active time, physical validity, institutional approval, and realized commercial value remain unmeasured. Publication and ownership instructions are additive in `handoff/PUBLICATION_HANDOFF.md`; the original Foundry paper, releases, URLs, and held implementations remain distinct.