



ENCEVRA RESEARCH / ROUND 33

# Who Certifies Independence?

Provenance Completeness and the Limits of Self-Attested Witness  
Quorums

---

**READINESS DECISION: BLOCKED**

October 24, 2026

An exact 3,375-topology diagnostic shows that signatures and self-attested lineage cannot certify administrative independence. Partial external evidence remains unsound if missing critical layers are interpreted optimistically.

# Contents

|                                                                  |           |
|------------------------------------------------------------------|-----------|
| <b>Abstract</b>                                                  | <b>3</b>  |
| <b>1. Archive continuity and corrected round number</b>          | <b>3</b>  |
| <b>2. Research question, novelty, scope, and exclusions</b>      | <b>4</b>  |
| <b>3. Definitions and the claim being certified</b>              | <b>6</b>  |
| <b>4. Literature map and competing positions</b>                 | <b>6</b>  |
| <b>5. Claims and evidence table</b>                              | <b>8</b>  |
| <b>6. Transcript-equivalence limit</b>                           | <b>8</b>  |
| <b>7. Proposed mechanism: AIC-33</b>                             | <b>9</b>  |
| <b>8. Executed exact finite diagnostic</b>                       | <b>10</b> |
| <b>9. Results</b>                                                | <b>11</b> |
| <b>10. Worked example</b>                                        | <b>12</b> |
| <b>11. Brain-inspired hypothesis, kept testable</b>              | <b>12</b> |
| <b>12. Prospective experiment under controlled total compute</b> | <b>13</b> |
| <b>13. Metrics, regressions, and harmful changes</b>             | <b>14</b> |
| <b>14. Alternative explanations and how to distinguish them</b>  | <b>16</b> |
| <b>15. Expected failure modes</b>                                | <b>16</b> |
| <b>16. Limitations</b>                                           | <b>16</b> |
| <b>17. Readiness decision</b>                                    | <b>17</b> |
| <b>18. Next-round handoff</b>                                    | <b>17</b> |
| <b>19. Reproducibility artifacts</b>                             | <b>18</b> |
| <b>20. Sources and inspection record</b>                         | <b>19</b> |
| <b>21. Conclusion</b>                                            | <b>21</b> |

# Abstract

Round 32 showed that separately keyed witness, registry, and monitor services can detect many execution and evidence-history faults, yet a shared administrator can emit the same observable pattern as a genuinely independent deployment. Round 33 asks the logically prior question that result exposed: **Can a collaboration system certify that a 3-of-4 witness quorum is administratively independent using only witness-generated identities and provenance claims, or does a useful certificate require externally anchored and sufficiently complete dependency evidence?**

The literature supports two parts of the answer and leaves the crucial boundary open. Byzantine quorum and general-adversary work makes guarantees relative to a declared failure structure, not merely a participant count [S03-S04]. Provenance and attestation systems make signed histories and role relationships verifiable, but signatures attest statements under a trust model; they do not prove that no dependency was omitted [S06-S09]. A new August 2026 preprint formalizes epistemic fault domains and a structural cut, while explicitly conditioning its lower-bound theorem on closed causal accounting and complete conservative exposure and warning that incomplete or fabricated lineage can inflate the apparent cut [S02]. Round 33 therefore does not duplicate the preprint's fault-domain counter. It targets its stated premise: what partial provenance can and cannot certify.

The executed diagnostic represents four witnesses across three critical layers: administrator, key custody, and deployment control. Each layer can have any of the 15 set partitions of four witnesses, producing  $15^3 = 3,375$  topologies. Ground truth is positive when some three witnesses are pairwise separated on all three layers. Exactly 211 topologies satisfy that condition; 3,164 do not.

Distinct keys or self-attested unique roots accept all 3,375 worlds and falsely certify all 3,164 negative worlds. Auditing one layer while treating missing layers optimistically still produces 1,364 false certifications among 1,575 acceptances, an 86.60% false-discovery rate in this deliberately uniform topology universe. Auditing two layers optimistically produces 434 false certifications among 645 acceptances, a 67.29% false-discovery rate. Complete external evidence for all three declared layers classifies the finite universe exactly by construction. A fail-closed policy that certifies only a fixed fully anchored trio produces no false certifications but recognizes only 64 of 211 positive worlds, 30.33% recall. These are combinatorial properties of the stated universe, not empirical estimates.

The round proposes **AIC-33, the Authority Independence Certificate**. It freezes the quorum claim and dependency dimensions, requires every witness-to-root binding to be supported by an issuer outside the root being described, checks freshness and revocation, conservatively couples missing or conflicting paths, and certifies only if the resulting conflict graph contains an independently supported quorum. Self-attestation remains evidence of what a signer said, not evidence that the signer is separately administered. Threshold signatures may compress participation, but the signer roster and authority bindings must remain auditable.

**No new Encevera answer benchmark or model benchmark was run. No production authority, key, deployment, or policy root was changed. The carried readiness audit remains 29 controls present, 4 partial, and 1 absent, with critical failures A3, I1, R4, R5, and C6. Round 33 clarifies the evidence required to resolve those failures; it does not resolve them.**

## 1. Archive continuity and corrected round number

The automation brief retained the historical label Round 11. Before topic selection, the live Lovable project was inspected at project ID 8483bc59-da36-492e-b8b8-6bf684070241. The Research overview, rounds index, publications list, and complete Round 32 and Round 31 texts showed completed work through Round 32. Relevant earlier texts included Rounds 30, 29, 26, 10, and 9. The correct next sequential number is therefore **Round 33**, not Round 11.

The carry-forward is narrow:

| Round | Established program contribution                                                                                                                               | Boundary carried into Round 33                                                                               |
|-------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|
| 9     | Reusable procedures should be learned only from verified experience and retained with the conditions under which they succeeded.                               | AIC-33 is proposed and tested structurally; it is not yet a learned or production procedure.                 |
| 10    | Agents should exchange recoverable evidence, revise selectively, and compare against matched-cost baselines that expose harmful revision.                      | Round 33 asks whether the claimed plurality behind that evidence is itself supportable.                      |
| 26    | Observationally identical worlds cannot be distinguished by a rule over those observations; root-disjoint channels matter.                                     | The present round applies that boundary to independence claims and missing dependency paths.                 |
| 29    | Interventions can identify behavioral dependencies, but provenance equivalence must remain unresolved when multiple lineages fit the observations.             | Behavioral separation is not automatically administrative separation.                                        |
| 30    | The RISE readiness gate defined 34 controls and blocked a model-facing pilot.                                                                                  | The same critical-control rule remains in force.                                                             |
| 31    | A signed seven-fixture sandbox improved local checks but left authority roots collocated.                                                                      | Distinct process keys are not yet evidence of distinct administrative roots.                                 |
| 32    | A 256-case boundary holdout reached 93.75% exact classification yet accepted all 16 shared-administrator aliases; clean and alias observable patterns matched. | The next intervention must address the evidence for independence, not add another locally generated receipt. |

Round 32 explicitly handed off an external-authority intervention. This environment still supplied no separately administered witness, external key custodian, isolated deployment control plane, or model runner. Round 33 therefore performs the permitted fallback: an exact boundary-engineering diagnostic of the certification problem, with the blocked verdict preserved.

## 2. Research question, novelty, scope, and exclusions

### 2.1 Primary research question

> Can a 3-of-4 witness quorum be certified as independent from the witnesses' own signed identities and dependency claims, and how does certificate soundness change as externally anchored coverage increases from zero to all declared critical layers?

The target predicate for a topology  $x$  is:

$\text{IndependentQuorum}(x) = \text{true}$  if and only if at least one three-witness subset has three different roots on administrator, key-custody, and deployment-control layers.

This definition is intentionally structural. It says nothing about whether the witnesses reason correctly, whether their evidence is true, or whether the authorized action is safe.

### 2.2 Why this question is new rather than redundant

Three candidate directions were checked against the archive and current literature: calibrated confidence, compute-budget allocation, and provenance-preserving memory. Confidence and budget allocation remain important, but neither can repair an independence claim whose causal inputs are missing. Memory can preserve a conflict only after a conflict has been represented.

A broader question about counting independent fault domains was rejected as insufficiently novel. He and Yu's August 24, 2026 preprint already defines epistemic fault domains, a structural epistemic cut, and a dependency-aware quorum controller [S02]. Its formal guarantee assumes closed causal accounting and complete conservative exposure. Its limitations section explicitly identifies unmodeled latent dependencies, incomplete provenance, and fabricated lineage as ways to inflate the reconstructed cut. Round 33 selects that open boundary rather than restating the paper's contribution.

The new object is therefore not another quorum rule. It is a **certificate of the evidence used to justify the quorum's independence**, including a fail-closed interpretation for missing dependency paths.

## 2.3 In scope

- Four named witnesses and a 3-of-4 authorization threshold.
- Three declared critical dependency layers: administrator, key custody, and deployment control.
- All set partitions of witnesses on every layer.
- Self-attested identity and lineage, partially externalized lineage, complete externalized lineage, and a fail-closed partial-deployment policy.
- Exact observable-equivalence classes under each evidence depth.
- A proposed certificate schema and decision procedure.
- A future matched-compute experiment that separates structural certification from semantic model performance.

## 2.4 Exclusions

- Claiming that the three chosen layers exhaust every production dependency.
- Estimating the probability of any topology or attack.
- Proving that an external issuer is honest, uncompromised, or globally independent.
- Remote attestation implementation, certificate-authority deployment, HSM provisioning, organizational audits, or legal ownership verification.
- Language-model reasoning, debate, answer revision, or benchmark accuracy.
- Production mutations or deployment.
- Literal copying of a biological brain.

## 2.5 Epistemic labels used below

- **Established external finding:** directly supported by an inspected primary paper, standard, or authoritative technical source.
- **Mixed or bounded evidence:** credible evidence exists, but assumptions or transfer limits materially constrain it.
- **Observed in Round 33:** generated by the deterministic enumeration and stored artifacts.
- **Enceвра hypothesis:** a proposed design claim for future falsification.
- **Open question:** not answered by this run.

### 3. Definitions and the claim being certified

| Term                      | Round 33 definition                                                                                                                | Invalid substitution                                          |
|---------------------------|------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------|
| Witness                   | A named participant whose approval can contribute to an authorization threshold.                                                   | A public key, process, model instance, or hostname by itself. |
| Root                      | A failure or control authority on one declared layer.                                                                              | A cosmetic identifier supplied by the witness it controls.    |
| Critical layer            | A dependency dimension whose shared root can invalidate the independence claim.                                                    | Any available metadata field.                                 |
| Actual topology           | The three partitions assigning witnesses to real roots.                                                                            | The signed manifest alone.                                    |
| Anchored binding          | A witness-to-root claim supported by an issuer whose authority is outside the root being described, under a declared trust policy. | A witness signing its own statement that it is independent.   |
| Conflict edge             | A pair of witnesses that share a root, have contradictory evidence, or lack sufficient coverage under the fail-closed policy.      | Evidence that either witness is malicious.                    |
| Certified independent set | A witness subset with no conflict edge under the declared layers and evidence policy.                                              | Proof that all latent dependencies are absent.                |
| unproven                  | Available evidence does not justify the requested independence threshold.                                                          | unsafe, compromised, or false.                                |
| Completeness              | Every dependency path material to the declared claim is represented conservatively.                                                | A large number of signed fields.                              |

The AIC-33 claim is deliberately scoped:

> For decision  $d$ , witness roster  $W$ , threshold  $q$ , time window  $t$ , and declared layer set  $L$ , the evidence bundle supports a conflict-free subset of at least  $q$  witnesses, with each required witness-to-root binding externally anchored and current under policy  $P$ .

It is not the stronger claim that every possible common cause has been discovered. That stronger claim is unavailable unless the dependency basis itself is justified.

## 4. Literature map and competing positions

### 4.1 Quorum guarantees depend on a failure structure

Malkhi and Reiter define Byzantine quorum systems relative to a fail-prone system and derive conditions under which consistency and availability can be retained despite arbitrary repository faults [S03]. Fitzi and Maurer study Byzantine agreement against a general adversary represented by a family of corruptible participant subsets, rather than only a scalar fault count [S04].

The established point is that a participant count does not specify a resilience guarantee. The system needs an adversary or failure structure. Round 33 applies that idea one level earlier: before using a declared structure, the system must know what evidence supports the mapping from witness identities to failure roots.

### 4.2 Diverse implementations can still fail together

Knight and Leveson's 1986 multiversion-programming experiment ran independently developed program versions over one million tests and found coincident failures more often than an independence model predicted [S05]. The experiment is old, bounded, and about software versions rather than agent administration, so it cannot quantify Enceavra risk. It nevertheless rejects the automatic inference from separate implementations to statistically independent failure.

This is a competing position to naive diversity: visible heterogeneity may help, but it must be measured against shared specifications, training data, tools, administrators, and evaluation channels.

### 4.3 Provenance systems authenticate declared histories

The in-toto paper uses a signed supply-chain layout and signed link metadata to connect steps, actors, inputs, and outputs [S06]. SLSA 1.2 defines provenance as verifiable information tracing an artifact through the moving parts of its supply chain to where, when, and how it was produced [S07]. NIST IR 8536 describes continuous, linked, temporally ordered traceability data across supply-chain ecosystems [S12].

These systems support an affirmative engineering position: lineage can be represented, linked, signed, verified, and queried. They do not make every history complete merely because the represented entries verify. A completeness claim still depends on the layout, required fields, issuers, capture points, and threat model.

### 4.4 Attestation separates evidence, appraisal, and reliance

RFC 9334 distinguishes Attesters, Verifiers, Relying Parties, Endorsers, Reference Value Providers, and owners of appraisal policies [S08]. It also makes trust anchors and their protection explicit. This role separation is useful for AIC-33: a witness can present evidence; a verifier can appraise it; a relying party can decide whether the result supports authorization.

The standard also allows roles to be aggregated. Role labels therefore do not prove administrative separation. A certificate must preserve which real entity performs each role and which trust anchor authorizes that relationship.

### 4.5 Threshold cryptography distributes a key operation, not every authority dependency

RFC 9591 specifies FROST, in which a threshold number of entities cooperate to produce a Schnorr signature from shares of a group key [S09]. That is valuable for avoiding a single key holder. The final signature can verify like a single-signer signature, however, so an application that discards the participant roster also discards evidence relevant to the present independence claim.

A threshold signature proves that the protocol's signing threshold participated under its cryptographic assumptions. It does not, by itself, prove that the participants had distinct administrators, deployment controls, information sources, or policy owners.

### 4.6 Epistemic fault domains formalize shared causal ancestry

He and Yu define epistemic fault domains and a structural epistemic cut that measures how many modeled root causes are needed to cover an authorizing coalition [S02]. Their theorem lower-bounds a semantic compromise cut only under authorization alignment, closed causal accounting, and complete conservative exposure. They explicitly state that incomplete provenance, fabricated lineage, and unmodeled latent dependencies can inflate the reconstructed cut, and recommend conservatively coupling unknown paths.

This recent source is a non-peer-reviewed preprint and should be treated accordingly. It supplies a precise adjacent formalism and a clearly stated limitation. Round 33 tests a finite instance of that limitation; it does not claim to validate the preprint's complete framework.

### 4.7 The knowledge boundary is observational

Halpern and Moses formalize distributed knowledge using worlds that remain indistinguishable under a participant's or group's view [S10]. Their result is broader and more formal than the simple lemma used here. The transferable point is that a decision procedure cannot condition on facts absent from its observation.

Round 33 therefore separates two properties:

- 1 **Cryptographic authenticity:** the received statement was produced under a particular key.

## 2 Authority independence: the signer is controlled separately on every declared critical layer.

The first can be observed from a signature and trusted key binding. The second requires evidence about the control relationships behind the key.

## 5. Claims and evidence table

| Claim                                                                                                                                | Status                                                  | Evidence                                                                                                                              | Boundary                                                                                                              |
|--------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|
| Distinct valid signatures establish distinct administrative authorities.                                                             | Falsified in the finite model and unsupported generally | B0 accepts all 3,375 topologies, including 3,164 without a structurally independent quorum.                                           | The diagnostic abstracts away key compromise; it shows only that key count is insufficient.                           |
| Self-attested unique root identifiers repair the problem.                                                                            | Falsified in the finite model                           | B1 has exactly the same confusion matrix as B0 because a common controller can emit the same claims under all four keys.              | Honest self-description can still be operationally useful; it is not an external certificate.                         |
| Partial external evidence can be interpreted optimistically without losing soundness.                                                | Falsified in the finite model                           | One-layer and two-layer optimistic policies have 1,364 and 434 false certifications, respectively.                                    | Rates are uniform over combinatorial topologies, not prevalence-weighted.                                             |
| Complete evidence for the three declared layers exactly recovers the target.                                                         | Observed by construction                                | B4 has 211 true positives, 3,164 true negatives, and no errors.                                                                       | It proves the implementation matches the finite definition; it does not prove the layer basis is complete in reality. |
| Fail-closed partial evidence trades coverage for soundness.                                                                          | Observed in the finite model                            | B5 certifies 64 positive topologies with zero false certifications but misses 147 other positives.                                    | This is one fixed-trio policy, not an optimal availability policy.                                                    |
| A transcript-only certifier can distinguish an independent world from a common-controller alias when both transcripts are identical. | Impossible by definition                                | Stored alias worlds have equal observable JSON and equal hashes but opposite actual control structures.                               | An external anchor can change the observable transcript and break the equivalence.                                    |
| Externally anchored provenance proves the absence of every hidden dependency.                                                        | Not established                                         | Inspected provenance and attestation sources support verifiable claims under trust models, not universal completeness [S06-S08, S12]. | AIC-33 must state its dependency basis and residual trust assumptions.                                                |
| Heterogeneous witness stacks will reduce joint semantic misses at matched cost.                                                      | Enceвра hypothesis                                      | Biological degeneracy and multiversion evidence motivate a test, but do not establish the result [S05, S11].                          | Requires an actual model-facing experiment with hidden faults and controlled compute.                                 |

## 6. Transcript-equivalence limit

Let  $T$  be the complete input available to a certifier and let  $C(T)$  be its decision. Consider:

- World I: four separately administered witnesses with distinct roots on all declared layers.
- World A: one controller operates all four keys and emits the same signer roster and the same four unique-root claims.

The diagnostic stores both worlds with identical observable objects. Therefore their canonical transcript hashes are equal. Any deterministic certifier over that transcript must return the same decision in both worlds. A randomized certifier has the same output distribution in both worlds.

This yields a simple boundary:

> If a certificate accepts World I from the self-attested transcript, it accepts World A with the same probability. If it rejects World A, it rejects World I with the same probability.

No additional parsing rule, confidence score, debate round, or threshold signature over the same transcript can break the equivalence. A separating observation must come from outside the common controller's ability to produce both sides of the comparison, or the claim must be weakened to self-reported separation.

The limit is not anti-cryptographic. Signatures remain essential for authenticity, integrity, and non-repudiation within a trust model. The limit says that a valid signature on an incomplete or self-referential dependency statement is not evidence of completeness or external independence.

## 7. Proposed mechanism: AIC-33

### 7.1 Certificate fields

An Authority Independence Certificate should contain at least:

- `certificate_version`, decision identifier, protected action digest, and expiry;
- witness roster, authorization threshold  $q$ , and actual participant roster;
- declared critical layers and the justification for their inclusion;
- one witness-to-root binding per witness per layer;
- issuer identity, issuer role, trust-anchor path, capture method, issue time, and revocation state for every binding;
- evidence digest and retrieval location;
- conflicts, missing fields, stale fields, and contradictory bindings;
- conservative-coupling policy for unknowns;
- computed conflict graph and certified independent set;
- residual assumptions, excluded layers, and an explicit `certified`, `unproven`, or `invalid` result.

### 7.2 Decision procedure

- 1 **Freeze the claim.** Bind the protected decision, roster, threshold, time window, critical layers, and acceptance policy before collecting approvals.
- 2 **Collect root bindings.** For every witness and layer, obtain a claim about the controlling root.
- 3 **Check externality.** Reject self-attestation as an independence anchor. The issuer supporting a root binding must be outside the root being described under the declared organizational model.
- 4 **Check freshness and revocation.** Validate time bounds, trust-anchor chains, revocation state, and capture continuity.
- 5 **Conservatively couple unknowns.** Add a conflict edge for shared roots, contradictory evidence, stale evidence, missing required evidence, or an issuer inside the described root.
- 6 **Find a certified set.** For the small roster, enumerate witness subsets exactly and find the largest conflict-free subset. For larger rosters, record the solver, objective, and proof artifact.
- 7 **Decide.** Return `certified` only if a conflict-free subset of size at least  $q$  exists. Return `unproven` for insufficient coverage. Return `invalid` for cryptographic, binding, or policy failures.
- 8 **Preserve participation.** If a threshold signature is used, retain the authenticated participant roster and their authority bindings beside the aggregate signature.
- 9 **Log the certificate.** Bind the result to the action and preserve it in an append-only system with independent monitoring, without treating the log as proof that the inputs were complete.

## 7.3 Why unproven is necessary

Missing provenance does not establish shared control, so calling the system compromised would overstate the evidence. But treating missing provenance as separation creates false assurance. unproven is the precise operational state: the requested authorization must stop, but the certificate makes no accusation about the witnesses.

## 7.4 Externality is recursive

An external issuer can itself be controlled by the root it is supposed to distinguish. AIC-33 therefore requires the trust path and role ownership to be visible. This does not eliminate recursion; it makes the stopping assumption explicit. For example, an enterprise registrar may certify administrators, a cloud provider may attest tenancy or hardware roots, and a separate key custodian may attest key ownership. The relying party must still state why those issuers are independent enough for the action's threat model.

# 8. Executed exact finite diagnostic

## 8.1 Universe

Four witnesses can be partitioned into shared-control groups in 15 ways on one layer, the fourth Bell number. The diagnostic uses the Cartesian product over three layers:

15 administrator partitions x 15 key-custody partitions x 15 deployment-control partitions = 3,375 topologies.

Every topology is equally weighted for enumeration. No claim is made that these topologies are equally likely in deployment.

## 8.2 Ground truth

For each topology, the script enumerates all four possible three-witness subsets. A subset is structurally independent only when its three witnesses have different group labels on all three layers. The topology is positive if at least one such subset exists.

The enumeration produces:

- 211 positive topologies with at least one independent 3-of-4 quorum;
- 3,164 negative topologies without one;
- 3,375 total.

## 8.3 Evidence arms

| Arm | Evidence policy                                              | Missing-layer treatment                                                            |
|-----|--------------------------------------------------------------|------------------------------------------------------------------------------------|
| B0  | Four distinct signing identities only                        | Ignores all dependency layers.                                                     |
| B1  | Each signer self-attests a unique root                       | Trusts the common-controller transcript.                                           |
| B2  | Administrator partition externally anchored                  | Treats key custody and deployment control as separated unless a known edge exists. |
| B3  | Administrator and key-custody partitions externally anchored | Treats deployment control as separated unless a known edge exists.                 |
| B4  | All three partitions externally anchored                     | Applies the declared target definition exactly.                                    |
| B5  | Only fixed trio W1/W2/W3 is fully anchored; W4 is excluded   | Fails closed on the unanchored witness.                                            |

B4 is an oracle within this finite universe. Calling it an external audit means the evaluator receives the actual partition labels for all declared layers. It is not evidence that a real-world audit can achieve perfect capture.

## 8.4 Reproducibility

The Python implementation uses only the standard library. It generates set partitions as restricted-growth strings, enumerates the Cartesian product, records every topology, writes the confusion matrices and observable-equivalence summary, and tests the expected counts. Five unit tests cover the Bell count, ground-truth count, arm counts, equivalence classes, transcript alias, and a worked certificate example.

The run makes zero model calls and consumes zero model tokens. Its output is deterministic for the committed code. The result digest over topology rows, arm summaries, and equivalence summaries is:

cdf12b71cc7857d3b907b7dfe445fd6d5684645ac12437a4358d8c08cd8bc8c0

## 9. Results

### 9.1 Confusion matrices

In this table, a positive decision means the policy certifies that some 3-of-4 quorum is independent under the target definition.

| Arm                                 | True positive | False positive | False negative | True negative | Precision | Recall  |
|-------------------------------------|---------------|----------------|----------------|---------------|-----------|---------|
| B0 distinct signatures              | 211           | 3,164          | 0              | 0             | 6.25%     | 100.00% |
| B1 self-attested roots              | 211           | 3,164          | 0              | 0             | 6.25%     | 100.00% |
| B2 one external layer, optimistic   | 211           | 1,364          | 0              | 1,800         | 13.40%    | 100.00% |
| B3 two external layers, optimistic  | 211           | 434            | 0              | 2,730         | 32.71%    | 100.00% |
| B4 complete three-layer audit       | 211           | 0              | 0              | 3,164         | 100.00%   | 100.00% |
| B5 fixed anchored trio, fail closed | 64            | 0              | 147            | 3,164         | 100.00%   | 30.33%  |

The precision values reflect the artificial uniform enumeration. The robust facts are the exact counts and the existence of false certifications. B2 and B3 appear to have perfect recall only because their optimistic policy accepts every truly positive topology; their cost is unsound acceptance of mixed evidence classes.

### 9.2 Observable-equivalence classes

| Externally audited layers | Observable classes | Mixed truth classes | Safe-only classes | Unsafe-only classes | Largest class |
|---------------------------|--------------------|---------------------|-------------------|---------------------|---------------|
| 0                         | 1                  | 1                   | 0                 | 0                   | 3,375         |
| 1                         | 15                 | 7                   | 0                 | 8                   | 225           |
| 2                         | 225                | 43                  | 0                 | 182                 | 15            |
| 3                         | 3,375              | 0                   | 211               | 3,164               | 1             |

With no anchored layer, every actual topology produces the same evidence view. With one anchored layer, seven of the 15 views contain both positive and negative worlds. With two anchored layers, 43 views remain mixed. Any certifier that accepts a mixed class necessarily accepts at least one negative world; any certifier that rejects it loses at least one positive world. Only the complete three-layer observation separates every topology in the declared universe.

### 9.3 Main interpretation

The result does not say that every production certificate must enumerate every causal fact in the universe. It says that a certificate can be sound only relative to a declared basis whose missing values are handled conservatively. Optimistically projecting independence through an unobserved critical layer converts absence of evidence into a resilience claim.

The fail-closed B5 policy demonstrates the alternative. By excluding W4 until its three bindings are anchored, it never certifies a negative world. It also misses positive worlds in which a different trio would have been valid. This is an explicit safety-availability tradeoff rather than hidden optimism.

## 10. Worked example

Suppose W1, W2, W3, and W4 present distinct public keys and each signs `administrator = unique`.

The actual administrator partition is  $\{W1, W2\}$ ,  $\{W3\}$ ,  $\{W4\}$ : W1 and W2 share administrator A, while W3 and W4 are controlled separately. Key custody and deployment control are initially distinct for all four.

### Step 1: self-attested view

B1 sees four valid keys and four unique-root claims. It certifies any three signers, even though the claims could all have been generated by administrator A in a common-controller world. The certificate is not sound.

### Step 2: external administrator evidence

An external registrar reveals the shared W1/W2 root. The conflict graph gains edge W1-W2. A 3-of-4 independent set still exists:  $\{W1, W3, W4\}$  or  $\{W2, W3, W4\}$ , assuming the other layers are genuinely separate.

### Step 3: incomplete deployment evidence

W4 has no externally anchored deployment-control binding. An optimistic policy simply omits an edge and certifies  $\{W1, W3, W4\}$ . AIC-33 instead marks W4 unanchored on a critical layer and adds conservative conflicts or excludes W4. The remaining fully supported witnesses W1, W2, and W3 contain the W1-W2 administrator conflict, so the result is unproven.

### Step 4: evidence that changes the decision

If an issuer outside W4's deployment root supplies a current binding showing W4's deployment controller is distinct, the conflict can be removed under the declared policy. AIC-33 may then certify  $\{W1, W3, W4\}$ , provided key-custody bindings and every other required field also validate.

The important transition is not from low confidence to high confidence. It is from an equivalence class containing both valid and invalid topologies to a view that separates them under the declared basis.

## 11. Brain-inspired hypothesis, kept testable

Tononi, Sporns, and Edelman define degeneracy as structurally different elements capable of performing the same function and distinguish it from identical redundancy [S11]. The useful engineering analogy is limited:

- identical model replicas, prompts, tools, administrators, and evidence paths resemble redundancy;
- heterogeneous witness stacks with independently controlled roots may provide functionally overlapping judgment through structurally different paths.

This is not a claim that AIC-33 copies neural circuitry or that biological degeneracy proves a security property. It yields a falsifiable design hypothesis:

> At matched total compute and matched witness count, a quorum whose members differ in model family, retrieval source, toolchain, administrator, key custody, and deployment control will have a lower joint miss rate on hidden common-mode faults than a homogeneous quorum, without an unacceptable increase in clean-case rejection.

The hypothesis can fail. Superficial heterogeneity may share training corpora, specifications, benchmarks, cloud control planes, or policy sources. Independently administered systems may also make coincident local errors. AIC-33 measures declared structural separation; only an outcome experiment can determine whether that separation improves semantic reliability.

## 12. Prospective experiment under controlled total compute

No experiment in this section was run. It is a preregistration-quality proposal for a later round after external authorities and a model runner exist.

### 12.1 Research hypotheses

- **H1 - Self-attestation limit:** on adversarial common-controller aliases, distinct keys plus self-attested root claims will not reduce false certification relative to key count alone.
- **H2 - Coverage monotonicity:** conservative externally anchored coverage will never increase the set of certified witness coalitions when it reveals a shared root.
- **H3 - Soundness-availability tradeoff:** fail-closed handling of missing critical layers will reduce false certifications relative to optimistic handling, at the cost of more unproven outcomes.
- **H4 - Semantic benefit:** among coalitions with AIC-33 certificates, higher validated structural separation will reduce joint semantic miss rate on hidden common-mode evidence faults at matched compute.
- **H5 - Harm bound:** any semantic gain will not be counted as success if clean-case rejection, harmful answer revision, or abstention burden exceeds preregistered limits.
- **H6 - Brain-inspired ablation:** declared stack heterogeneity without administrative or evidence-root separation will not match the benefit of heterogeneity plus anchored separation.

### 12.2 Dataset strata

Use three disjoint strata and report them separately:

- 1 **Structural topology suite:** exact or sampled dependency graphs with hidden roots and adversarially incomplete manifests.
- 2 **Sandbox mutation suite:** real process, key, registry, and deployment boundaries with independently controlled fault injection.
- 3 **Model task suite:** decision tasks with hidden evidence corruption, stale retrieval, shared false premises, tool faults, and clean controls.

Do not pool authored synthetic accuracy with natural-task accuracy. Do not infer incident prevalence from balanced fault families.

### 12.3 Arms

- **A0 single witness:** one model, one evidence path.
- **A1 homogeneous 3-of-4 vote:** four replicas sharing model, tools, evidence, and administration.
- **A2 distinct keys plus self-attested lineage:** cryptographic plurality without external dependency evidence.
- **A3 complete manifest, optimistic unknowns:** dependency-aware scoring that omits edges for missing values.
- **A4 AIC-33 fail closed:** external bindings, conservative unknowns, explicit unproven result.
- **A5 AIC-33 plus heterogeneous stacks:** A4 with independently administered model, retrieval, and tool paths.
- **A6 oracle topology upper bound:** evaluator receives the hidden dependency graph; not deployable, used to measure remaining semantic error after perfect structural information.

## 12.4 Compute control

Freeze the total budget per task across A1-A5:

- same maximum number of model tokens;
- same maximum tool calls and retrieved bytes;
- same wall-clock deadline;
- same number of candidate answers exposed to the final decision rule;
- same verification budget, with unused budget recorded rather than reassigned after outcomes are visible.

Where different model families have materially different per-token cost or latency, report both a token-matched analysis and a cost-matched sensitivity analysis. Structural certificate construction should be accounted for separately from model inference but capped per task.

## 12.5 Blinding and adjudication

- Precommit fault families, hidden dependency graphs, metrics, stopping rules, and exclusions.
- Separate fixture authors from gate implementers where possible.
- Keep semantic labels and topology labels unavailable to the decision systems.
- Use independent adjudicators for task truth and a separate authority for deployment-root truth.
- Preserve all manifests, certificates, transcripts, participant rosters, and costs before unblinding.

## 12.6 Baselines and ablations

Required ablations include:

- remove external issuer validation while retaining signed manifests;
- remove one critical layer at a time;
- treat missing values optimistically versus conservatively;
- preserve versus discard threshold-signature participant rosters;
- randomize root labels to test whether identifiers rather than relationships drive results;
- use identical models under independent administration versus different models under one administrator;
- use different evidence sources under one control plane versus one source mirrored across control planes;
- replace exact independent-set computation with key count or majority vote;
- remove freshness and revocation checks;
- remove hidden common-controller aliases.

The last ablation detects benchmark leakage: a policy that appears strong only after the hardest alias class is removed has not solved the target problem.

# 13. Metrics, regressions, and harmful changes

## 13.1 Structural metrics

- False certification rate among negative topologies.

- Precision and recall for certified under the hidden topology truth.
- unproven rate and time-to-resolution.
- Coverage by witness, layer, issuer, and trust anchor.
- Number and size of mixed observable-equivalence classes.
- Certificate freshness, revocation failures, contradictory binding rate, and orphaned roots.
- Difference between claimed and independently adjudicated maximum independent-set size.

## 13.2 Semantic metrics

- Task accuracy and calibrated Brier score before and after collaboration.
- Joint miss rate under shared evidence faults.
- Selective accuracy at fixed coverage.
- Correction rate: initially wrong answers repaired.
- Harmful revision rate: initially correct answers changed to wrong.
- False authorization rate on critical cases.
- Clean-case rejection and abstention burden.
- Evidence-grounding precision: proportion of decisive claims supported by the preserved external evidence.

## 13.3 Cost and concentration metrics

- Total tokens, tool calls, retrieved bytes, latency, and monetary cost.
- Verification effort per corrected error and per avoided false authorization.
- Root concentration: fraction of the accepted quorum dependent on the most common root at each layer.
- Issuer concentration and trust-anchor concentration.
- Sensitivity to the removal or compromise of one issuer.

## 13.4 Regression guardrails

Predeclare a deployment failure if any of the following occurs:

- AIC-33 certifies a common-controller alias under the declared basis.
- Harmful revision rises beyond the preset non-inferiority margin.
- Clean-case rejection or unproven burden exceeds the operational bound without a proportional reduction in critical false authorizations.
- Cost-matched performance is worse than a simpler single-witness or homogeneous baseline.
- Results disappear when labels, issuers, or task order are permuted.
- The certificate omits the actual participant roster or cannot be bound to the protected action.

## 14. Alternative explanations and how to distinguish them

| Observed future result                              | Alternative explanation                                                       | Discriminating test                                                           |
|-----------------------------------------------------|-------------------------------------------------------------------------------|-------------------------------------------------------------------------------|
| AIC-33 lowers false authorization                   | It merely forces more abstention.                                             | Compare selective risk at matched coverage and report unproven burden.        |
| Heterogeneous quorum outperforms homogeneous quorum | Larger or better models, not independence, caused the gain.                   | Match compute and cross administration/model-family factors factorially.      |
| External provenance predicts success                | The issuer is a proxy for higher-quality organizations.                       | Randomize or stratify by issuer and test within matched quality bands.        |
| Threshold signing appears safer                     | The threshold protocol, not administrative separation, improves key security. | Preserve the same cryptography while changing participant control roots.      |
| Complete manifests classify topology accurately     | The benchmark leaks hidden roots through identifiers.                         | Relabel roots, regenerate manifests, and use independently authored aliases.  |
| No benefit from structural separation               | Fault injections did not activate shared dependencies.                        | Measure fault transmission and include faults targeted at each declared root. |

## 15. Expected failure modes

- **Self-reference:** a witness or its administrator issues the evidence used to prove that it is independent.
- **Issuer monoculture:** nominally external registrars share one parent, cloud account, policy owner, or data source.
- **Omitted layer:** the certificate covers keys and administrators but not deployment, retrieval, model update, or policy roots.
- **Stale ownership:** a binding remains valid after an acquisition, account transfer, key rotation, or deployment migration.
- **Revocation lag:** a compromised or reassigned root remains accepted during the decision window.
- **Composite-role collapse:** Attester, Verifier, Relying Party, and policy owner are distinct labels implemented by one entity.
- **Aggregate-signature opacity:** the final artifact proves a threshold signature but does not preserve who participated.
- **Identifier laundering:** one controller creates multiple legal entities, accounts, or certificates that look distinct syntactically.
- **Unknown-as-independent bug:** absent fields create no graph edge.
- **Unknown-as-universal-conflict overload:** fail-closed handling becomes so conservative that the system can never authorize, encouraging bypass.
- **Basis incompleteness:** all declared layers are captured, but an unmodeled shared source dominates the witnesses' semantic errors.
- **Correct structure, wrong semantics:** independent witnesses reach the same wrong conclusion through coincident local errors or an incorrect authoritative policy.

The design response is not to claim these failures away. It is to make the basis, issuers, unknown policy, expiry, and residual assumptions inspectable and to evaluate operational bypass pressure.

## 16. Limitations

- 1 The three-layer universe is a deliberately small abstraction. Real systems may share model weights, pretraining data, retrieval corpora, policy authors, cloud regions, identity providers, time sources, orchestration, networks, and human operators.
- 2 Partitions encode equality of roots, not partial influence, nested ownership, time-varying control, collusion probability, or graded compromise.
- 3 The enumeration weights every topology equally. Precision and false-discovery percentages are diagnostic summaries, not population estimates.
- 4 Complete three-layer evidence is supplied as ground truth to B4. A real implementation must acquire it through fallible institutions and technical systems.
- 5 Externality remains relative. A certificate authority, provider, auditor, or registrar introduces another trust root rather than eliminating trust.
- 6 The transcript-equivalence result is exact only when the observables are equal. Active intervention may produce a separating observation if the common controller cannot predict, suppress, or forge it.
- 7 AIC-33 certifies a structural claim, not model competence, honesty, evidence truth, policy correctness, or action safety.
- 8 No human-subject, user-impact, or production-operations evidence was collected.
- 9 The 2026 epistemic-fault-domain source is a recent preprint and has not been treated as settled consensus.
- 10 The proposed experiment has not been run; all expected benefits remain hypotheses.

## 17. Readiness decision

The Round 32 readiness manifest is carried forward without credit for new production controls:

| Status  | Count |
|---------|-------|
| Present | 29    |
| Partial | 4     |
| Absent  | 1     |

The critical non-present controls remain:

- **A3:** genuine administrative separation;
- **I1:** independent observation;
- **R4:** independently administered witness;
- **R5:** receipt path outside the actor's control boundary;
- **C6:** model runner for an actual model-facing pilot.

**Decision: BLOCKED. The exact finite diagnostic clarifies why missing authority evidence must not be interpreted optimistically. It does not establish any new external authority, so no critical control changes state.**

**No new Encevera answer benchmark was run. No model call was made. No production mutation, key operation, deployment, or user-facing revision occurred.**

## 18. Next-round handoff

Round 34 should not add another self-generated manifest. It should attempt one of two evidence-producing interventions:

- 1 **Preferred: three-root external pilot.** Provision at least three witnesses with independently administered identities, key custody, and deployment controls. Obtain one external binding per witness per critical layer, run AIC-33 before action, inject hidden shared-root and stale-binding cases, and have a separate authority adjudicate the topology.
- 2 **If external roots remain unavailable: basis-completeness red team.** Give an independent fixture author the AIC-33 schema and ask them to construct latent dependencies omitted from the declared basis. Measure how often the certificate falsely accepts and which new layer or trust assumption separates each alias.

Only after A3, I1, R4, and R5 are present should the program activate C6 and run the matched-compute model experiment from Section 12. The decisive next question is whether independently anchored structure improves semantic reliability without excessive clean-case rejection or harmful revision.

## 19. Reproducibility artifacts

The accompanying package contains:

- this complete manuscript;
- `round33_authority_independence.py`, the standard-library exact enumerator;
- `test_round33.py`, five deterministic unit tests;
- `summary.json`, the complete aggregate result and result digest;
- `topologies.csv`, one row per topology and arm decision;
- `equivalence_classes.csv`, observable-class counts by audit depth;
- `aic33_protocol.json`, the proposed decision sequence;
- `source_ledger.csv`, source dates, inspected portions, support, and limits;
- `README.md` and `SHA256SUMS.txt`.

Reproduction command:

```
python repro/round33_authority_independence.py --output results_final && python -m unittest -v repro/test_round33.py
```

## 20. Sources and inspection record

### [S01] Encevera Research archive

**Inspected October 24, 2026.** Live Lovable project 8483bc59-da36-492e-b8b8-6bf684070241: Research overview, rounds index, publications registry, complete Rounds 32 and 31, and relevant Rounds 30, 29, 26, 10, and 9. Supports numbering, carry-forward, Round 32 metrics, and the current blocked-control state. Internal project evidence; not independent validation.

### [S02] Jun He and Deying Yu, “The Illusion of Independent Quorums: Epistemic Fault Domains and Correlated Cognitive Failures in Agentic Quorums”

**arXiv:2609.02925v1, August 24, 2026. Inspected October 24, 2026.** Abstract; definitions of epistemic fault domains and structural epistemic cut; Theorem 1 assumptions; limitations on incomplete provenance, fabricated lineage, and unmodeled latent dependencies. Supports the novelty boundary and the need for conservative exposure. Recent non-peer-reviewed preprint; does not validate AIC-33. <https://arxiv.org/html/2609.02925v1>

### [S03] Dahlia Malkhi and Michael Reiter, “Byzantine Quorum Systems”

**Proceedings of the 29th ACM Symposium on Theory of Computing, 1997. Inspected October 24, 2026.** Abstract, system model, quorum definitions, fail-prone systems, and opaque-quorum discussion. Supports modeling quorum guarantees relative to failure scenarios rather than participant count. Concerns replicated data and declared fault structures, not provenance completeness. <https://www.cs.umass.edu/~arun/cs691ee/reading/BQS97.pdf>

### [S04] Matthias Fitzi and Ueli Maurer, “Efficient Byzantine Agreement Secure Against General Adversaries”

**DISC 1998. Inspected October 24, 2026.** Abstract, general-adversary structure, and Byzantine-agreement model. Supports representing correlated corruption as a family of possible corrupt subsets. Does not address how a deployed system learns that structure. <https://crypto-test.ethz.ch/publications/files/FitMau98.pdf>

### [S05] John C. Knight and Nancy G. Leveson, “An Experimental Evaluation of the Assumption of Independence in Multiversion Programming”

**IEEE Transactions on Software Engineering, SE-12(1), January 1986. Inspected October 24, 2026.** Publisher metadata/abstract and accessible experiment summary covering 27 versions and one million tests. Supports caution against assuming independent failures from separately developed versions. Transfer to modern AI witnesses is indirect; the experiment does not estimate Encevera's failure correlations. <https://doi.org/10.1109/TSE.1986.6312924>

### [S06] Santiago Torres-Arias et al., “in-toto: Providing farm-to-table guarantees for bits and bytes”

**28th USENIX Security Symposium, August 2019. Inspected October 24, 2026.** Abstract, architecture, signed layout, functionary roles, link metadata, and key-compromise analysis. Supports signed step provenance and declared actor requirements. Does not prove that the declared layout contains every relevant dependency. <https://www.usenix.org/system/files/sec19-torres-arias.pdf>

## [S07] Supply-chain Levels for Software Artifacts, “Provenance,” specification v1.2

**Current approved specification accessed October 24, 2026.** Provenance definition and build/source provenance scope. Supports treating provenance as verifiable information about where, when, and how an artifact was produced. Software-artifact scope; not a certificate of organizational independence. <https://slsa.dev/spec/v1.2/provenance>

## [S08] IETF RFC 9334, “Remote ATtestation procedureS (RATS) Architecture”

**January 2023. Inspected October 24, 2026.** Architectural roles, conceptual messages, composite devices, appraisal policies, trust anchors, role aggregation, and trust-anchor protection. Supports separating evidence production, verification, reliance, and policy ownership. Does not make aggregated roles independent or remove root-of-trust assumptions. <https://www.rfc-editor.org/rfc/rfc9334.html>

## [S09] Deirdre Connolly et al., RFC 9591, “The Flexible Round-Optimized Schnorr Threshold (FROST) Protocol for Two-Round Schnorr Signatures”

**IRTF Informational RFC, June 2024. Inspected October 24, 2026.** Abstract, introduction, participant threshold, group key, and aggregation steps. Supports the distinction between distributed key participation and broader administrative independence. The RFC is informational and does not claim to certify control-plane separation. <https://datatracker.ietf.org/doc/html/rfc9591>

## [S10] Joseph Y. Halpern and Yoram Moses, “Knowledge and Common Knowledge in a Distributed Environment”

**Journal of the ACM 37(3), July 1990. Inspected October 24, 2026.** Abstract, possible-worlds knowledge framework, group knowledge, and indistinguishability results. Supports the observation-boundary framing. Round 33 uses only a simple transcript-equivalence consequence, not the paper's full epistemic formalism. <https://groups.csail.mit.edu/tds/papers/Halpern/JACM90.pdf>

## [S11] Giulio Tononi, Olaf Sporns, and Gerald M. Edelman, “Measures of degeneracy and redundancy in biological networks”

**Proceedings of the National Academy of Sciences 96, March 1999. Inspected October 24, 2026.** Abstract, distinction between structurally different degenerate elements and identical redundancy, and neural-system simulations. Supports only a testable analogy for heterogeneous functional overlap. It does not establish a security or AI-collaboration guarantee. <https://www.pnas.org/doi/pdf/10.1073/pnas.96.6.3257>

## [S12] Michael Pease et al., NIST IR 8536, “Supply Chain Traceability: Manufacturing Meta-Framework”

**Published September 2026. Inspected October 24, 2026.** Publication page and abstract describing linked, interoperable, temporally ordered traceability data. Supports the engineering value of connected provenance chains. Manufacturing scope; does not prove authority independence or data completeness. <https://csrc.nist.gov/pubs/ir/8536/final>

## 21. Conclusion

Round 33's result is a boundary, not a deployment claim. Four keys can still be one authority. Four signed claims of uniqueness can still be one story told four times. Partial external provenance narrows ambiguity, but if a critical layer remains unobserved, optimistic certification can still accept worlds without the required independent quorum.

A useful certificate must therefore say which dimensions were checked, who supplied each binding, why the issuer sits outside the root it describes, how unknowns were treated, and what residual assumptions remain. AIC-33 makes those obligations explicit and returns `unproven` when the evidence cannot carry the claim.

The system remains blocked until the certificate can be exercised against real, separately administered roots and then tested for semantic benefit under controlled compute. The next gain will not come from another locally generated signature. It will come from an observation that the common controller cannot manufacture on both sides of the comparison.