emanuel
celano
12.10.2026
20:27:16
GMT+02:00

EVIDE GOVERNANCE LAB

Adaptive Evidentiary Governance Research

EVIDE TRAER & EVIDE INDEX

Reconstructing AI Agent Rule Transitions Without Overstating the Evidence

A Controlled A/B Experiment on Declared Clock Relationships, Evidentiary Gaps and Reconstructability

Dr. Emanuel Celano

EVIDE / Informatica in Azienda — Bologna, Italy

EVIDE Working Paper — Series: TRAER + INDEX Experiments — No. 1

Version 0.2 · 12 October 2026

Experimental research output based exclusively on synthetic data.
Not a forensic certification, not evidence of real-world agent conduct, and not a legal finding.

EVIDE Governance Lab: https://lab.certifywebcontent.com/

EVIDE Canonical: https://app.certifywebcontent.com/

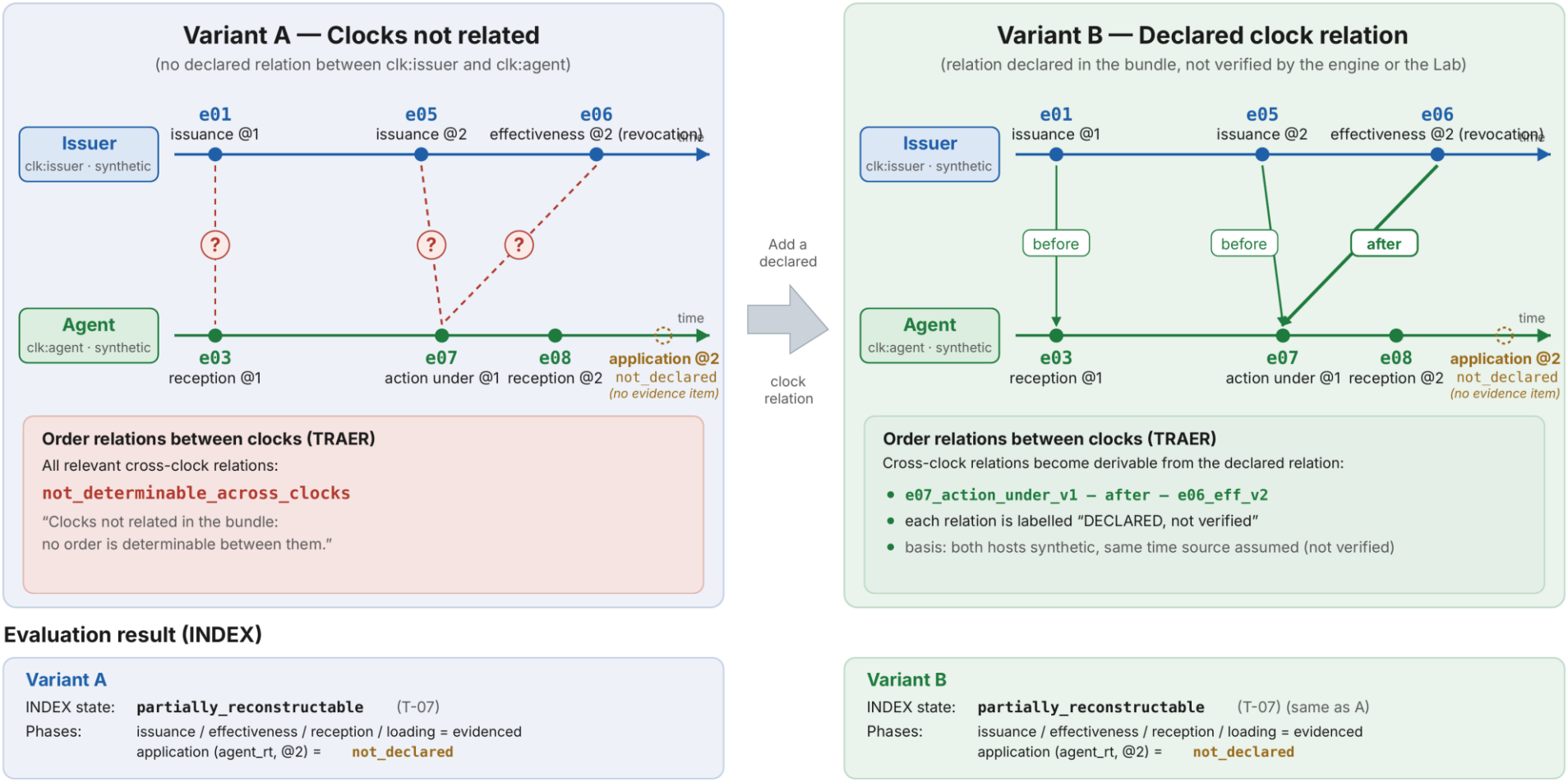
Preface — What this paper reports, and why it matters

On 12 October 2026 two controlled evaluations — variant_A and variant_B — were executed end to end on the operational EVIDE Governance Lab (lab.certifywebcontent.com), using exclusively synthetic data, the frozen EVIDE TRAER 0.3.2 engine, and the experimental profile v_traer_index_01. Every result in this paper is taken from the two real Lab evaluation records produced by those runs.

The runs demonstrate, on verifiable artifacts, a single discipline. **EVIDE TRAER** reconstructs what the declared evidence permits to be reconstructed; **EVIDE INDEX** states how far that reconstruction is evidentially supportable; and the system **labels, rather than hides, every point at which a conclusion rests on a declaration instead of a verification.**

The central observation is deliberately counter-intuitive. Adding an *explicitly declared* relationship between two otherwise-unrelated clocks **changed** which temporal orderings the engine could derive — in variant_A the cross-clock orderings were `not_determinable_across_clocks`; in variant_B they became `before` / `after` — yet it did **not** change the INDEX reconstructability state (`partially_reconstructable` in both) and did **not** close the evidentiary gap on the agent's application of the new version @2. A declared relationship produced orderings; it did not produce verification, and it did not remove the gaps that determine the evidentiary state.

What this paper is *not*: it is not a product brochure, not a benchmark against any other system, and not a claim of compliance, violation, certification, independent validation, or legal admissibility. It reports exactly what was executed, what was observed, what can be concluded, and what remains unverified — including that the independent human review and the independent confirmation of the 0.3.2 engine are still open.



Evaluation result (INDEX)

Variant A

INDEX state: **partially_reconstructable** (T-07)
Phases: issuance / effectiveness / reception / loading = evidenced
application (agent_rt, @2) = **not_declared**

Variant B

INDEX state: **partially_reconstructable** (T-07) (same as A)
Phases: issuance / effectiveness / reception / loading = evidenced
application (agent_rt, @2) = **not_declared**

Figure 1 — Declared Clock Relationships: What Changes, What Remains Unverified.

A/B comparison of the two runs on the same declared evidence (selected events; full nine-item set in §6). In variant_A the two clocks carry no declared relationship, so all cross-clock orderings are `not_determinable_across_clocks`; in variant_B a declared same_clock relation makes orderings derivable (`before` / `after`), each labelled "declared, not verified". The EVIDE INDEX state is `partially_reconstructable` (rule T-07) in both, and the application of the new version @2 by `agent_rt` remains `not_declared`. See §7 (Methodology) and §10 (Comparative results). Synthetic data only.

Key point: a declared clock relation changes the derivable ordering but **does not change the INDEX state in these two runs**; "after" does not imply a violation, and application @2 remains `not_declared`.

Contents

Preface — What this paper reports, and why it matters 2
1. Abstract 5
2. Research question 6
3. The evidentiary problem 6
4. EVIDE TRAER and EVIDE INDEX: distinct functions 7
5. Experimental architecture and boundaries 7
6. Synthetic scenario 8
7. Methodology and input differences between variants 9
8. Variant_A: clocks without a declared relationship 10
9. Variant_B: clocks with an explicitly declared relationship 11
10. Comparative results 12
11. Evidentiary interpretation 13
12. Integrity verification versus evaluation reproduction 14
13. Limitations and threats to validity 15
14. Reproducibility and available artifacts 17
15. Implications for forensic examination and AI governance 19
16. Conclusions 19
17. References and artifact manifest 20
Appendix A — Executive summary (one page) 21
Appendix B — Short abstract (for future public presentation) 23
Appendix C — Claims that require further proof before publication 24

1. Abstract

AI agent rule transitions — a rule issued, declared effective, received, loaded, applied, and later revoked — are typically documented by heterogeneous sources that record events on uncorrelated clocks. Reading such records naively invites a specific error: treating a declared action timestamp that falls after a declared revocation timestamp as evidence of a violation. This paper reports a controlled A/B experiment that examines how an explicitly declared relationship between otherwise-unrelated clocks affects the reconstruction of a single rule revocation, and which evidentiary conclusions remain unsupported.

Two synthetic evidence bundles — sharing the same nine evidence items and declared timestamps, with one intended experimental variable — were deposited into the operational EVIDE Governance Lab, evaluated off-server with the frozen EVIDE TRAER 0.3.2 engine, and re-imported under the Lab's provenance gate. In **variant_A** the two clocks carry no declared relationship; all cross-clock orderings resolve to `not_determinable_across_clocks`. In **variant_B** the bundle declares the two clocks as the same time source ("declared for the experiment only … not verified"); the same event pairs then resolve to `before` / `after`, each annotated by the engine as a declared, not verified, relation.

The EVIDE INDEX reconstructability state is `partially_reconstructable` (decisive rule T-07) in **both** variants, and the application phase of the new version @2 remains `not_declared` in both. The declared clock relation changed the derivable temporal orderings but did not change the evidentiary state and did not close the gap. The paper also documents the Lab's explicit separation between `import_integrity_verified` (a check of digests, binding and provenance performed by the Lab) and `evaluation_reproduced` (re-running the engine, which the Lab does not do; the operator obtained a `reproduced` verdict on the workstation, recorded as declared). Limitations are stated explicitly: fully synthetic data, two single runs, no independent observation of an agent, a declared and unverified clock relation, and pending independent review.

2. Research question

How does an explicitly declared relationship between otherwise unrelated clocks affect the reconstruction of AI agent rule transitions, and what evidentiary conclusions remain unsupported?

A secondary question frames the first: under what conditions may a reconstructed sequence be read as the actual operational sequence? The experiment is designed to keep these apart. Reconstructing a declared sequence is not the same as demonstrating the operational sequence that actually occurred; the paper treats any conflation of the two as the error to be avoided.

3. The evidentiary problem

The lifecycle of a governance rule assigned to a software agent spans several distinct phases: issuance by an authority, declared effectiveness, reception by the agent runtime, loading, application, and — for a revocation — the issuance and declared effectiveness of the new version, its reception and loading, and the question of whether the agent continued to act under the superseded version.

The evidence for these phases is ordinarily declared by different parties (an issuer registry, an agent runtime log) and timestamped on different clocks that are not known to be synchronized. Two difficulties follow. First, a phase may simply have no declared evidence, which is not the same as the phase not having occurred. Second, timestamps declared on uncorrelated clocks cannot, by themselves, establish an operational order between events, and an apparent order carries no finding of legitimacy or violation. A method that is useful for forensic and governance purposes must therefore reconstruct what the evidence permits, state the limits of that reconstruction, and refuse to convert absences into negative findings or declared orders into violations.

4. EVIDE TRAER and EVIDE INDEX: distinct functions

The experiment rests on a strict separation of two functions, each with its own output and its own boundaries.

Component	Function (as stated by the system in every output)
EVIDE TRAER	Reconstructs what the available evidence permits to be reconstructed: the phases of a transition, their status (evidenced / not_declared), the sources and declarants, order relations with their clock basis, and flags for incomplete observation. It imposes no operational restriction and does not verify that a rule was actually applied.
EVIDE INDEX	Assesses how evidentially supportable that reconstruction is: a categorical state (reconstructable, partially_reconstructable, unverifiable, unknown, inconsistent), the decisive rule, all concurrent conditions, and a full perimeter statement of what the state does and does not cover.

EVIDE INDEX, as stated in every evaluation it produces, is explicitly not any of the following:

- not a safety score;
- not a certification of compliance;
- it does not verify the authority of the issuer;
- it does not prove that the declared events happened;
- it does not assess responsibility.

These non-claims are not editorial caveats added for this paper; they are printed by the engine inside each assessment and reproduced in each Lab evaluation view.

5. Experimental architecture and boundaries

The evaluation engine runs **off-server**, on the operator's workstation, never inside the Lab. The Lab stores the exact bytes of each deposited bundle, computes their SHA-256 itself, and applies a sequence of gates (G-1…G-5 on deposit; G-6 on import). The experimental profile `v_traer_index_01` is isolated by construction: it adds only its own `exp_traer_*` tables, modifies no existing Lab file, exposes no public endpoint, and authorizes a single internal operator. The Lab performs no process execution and opens no network connection during the profile's operation.

The profile is **disabled by default**; for the two runs reported here it was **explicitly enabled by the single authorized operator** and then operated. “Disabled by default” therefore describes the installed, idle configuration, not the state during the experiment: the two evaluations were performed with the profile active. No public endpoint is exposed in either state.

The profile package used in this experiment is EVIDE Lab TRAER + INDEX STAGE 0.1 (package SHA-256 `9ddabad91e85…23cec4`). The approved evaluator for the experiment is EVIDE TRAER 0.3.2, recorded under the approval reference “`esperimento sintetico 1a - 2026-10-12`”. The division of responsibility is deliberate: the engine reconstructs and evaluates off-server; the Lab verifies the integrity and provenance of what it is given and asserts nothing about the evaluation itself.

6. Synthetic scenario

The governed unit is a single rule, `rule:R-1`, with two versions: `rule:R-1@1` (the previous version, in force) and `rule:R-1@2` (the new, revoking version). The transition under study is the `revocation` of @1 by @2 — that is, @2 supersedes @1. Two synthetic subjects declare evidence — an issuer registry (`issuer`, clock `clk:issuer`) and an agent runtime (`agent_rt`, clock `clk:agent`) — under a single declared synthetic operator (`op:synthetic-lab-operator`). The evidence items and their declared times, identical in both variants, are:

Evidence id	Phase	Version	Declarant / clock	Declared time
`e01_issue_v1`	issuance	@1	issuer / clk:issuer	2026-11-02T09:00:00Z
`e02_eff_v1`	declared_effectiveness	@1	issuer / clk:issuer	2026-11-02T09:00:00Z
`e03_recv_v1`	reception	@1	agent_rt / clk:agent	2026-11-02T09:05:00Z
`e04_load_v1`	loading	@1	agent_rt / clk:agent	2026-11-02T09:05:10Z
`e05_issue_v2`	issuance	@2	issuer / clk:issuer	2026-11-02T10:00:00Z

Evidence id	Phase	Version	Declarant / clock	Declared time
e06_eff_v2	declared_effectiveness	@2	issuer / clk:issuer	2026-11-02T10:05:00Z
e07_action_under_v1	application (of @1)	@1	agent_rt / clk:agent	2026-11-02T10:07:00Z
e08_recv_v2	reception	@2	agent_rt / clk:agent	2026-11-02T10:09:00Z
e09_load_v2	loading	@2	agent_rt / clk:agent	2026-11-02T10:09:30Z

The scenario contains a deliberate gap: there is **no evidence item** for the application of the new version @2 by the agent runtime. The agent declares an action under the previous version @1 (e07) but never declares applying @2. This missing evidence is the object of the experiment, not an oversight.

To keep the version semantics unambiguous, the scenario distinguishes six elements: (1) the **previous version** @1; (2) the **revocation transition**, @1 superseded by @2; (3) the **declared effectiveness** of the new version (e06_eff_v2); (4) an **action declared under the previous version** @1 (e07_action_under_v1); (5) the agent's **reception and loading of the new version** @2 (e08_recv_v2, e09_load_v2); and (6) the **application of the new version @2, which is never declared** (application / agent_rt = not_declared). The experiment turns on that last, missing, element — not on @2 being the “old” rule the agent kept using.

7. Methodology and input differences between variants

Both variants use the same nine evidence items with the same declared times, the same engine, and the same profile configuration. The experiment isolates a single independent variable: the presence of a declared relationship between the two clocks. Figure 1 contrasts the two variants side by side.

Aspect	variant_A	variant_B
Declared clock relation	none (empty)	clk:issuer same_clock clk:agent, basis: “DECLARED for the experiment only: both hosts synthetic, same time source assumed (not verified)”

Aspect	variant_A	variant_B
Evidence items / declared times	nine, as in §6	identical nine, identical times
Bundle SHA-256	`b7100fbfac88…a402dd`	`d9917492c871…2d65be`
Bundle length (bytes)	8363	8578
Incidental (non-semantic) differences	bundle_id label 1A; source_description text	bundle_id label 1B; source_description text

The 215-byte difference between the two bundles (8363 → 8578) corresponds to the added clock-relation object and the differing label strings. Inspecting the bundle contents and how the engine consumes them, the `bundle_id` label and the free-text `source_description` do not enter the engine's reconstruction logic or the INDEX rules, so the declared clock relation is the input difference to which the §10–§11 comparison is attributed. This attribution rests on inspection of the inputs and of the engine's logic, **not on a dedicated ablation run**: an otherwise-identical third bundle differing only in the labels was not executed, so the non-effect of the labels is reasoned, not experimentally isolated.

Procedure for each variant, in order: deposit the bundle on the Lab (gates G-1…G-5); download the exact stored bytes; run `traer_assess.py` and `traer_verify.py` from the frozen 0.3.2 package on the workstation; import the resulting snapshot and assessment back into the Lab (gate G-6). The engine source files verified for both runs are `traer_engine.py` (`02b81a39bf99…c4b1cc`) and `traer_assess.py` (`9aea4ff37df5…a5139e`), from the frozen package SHA-256 `f2fd9123fb03…331ee1`.

8. Variant_A: clocks without a declared relationship

Lab bundle #1 was received at `2026-10-12 16:08:33.648532` (Lab clock, UTC), with bundle SHA-256 `b7100fbfac88…a402dd` (8363 bytes) recomputed by the Lab on the received bytes, `data_label synthetic_only`, and deposit gate `G-1 passed; G-2 passed; G-3 passed; G-4 passed; G-5 not_applicable (format 0)`.

Observed engine output (EVIDE TRAER reconstruction and EVIDE INDEX):

- INDEX state `partially_reconstructable`; decisive rule `T-07` (“issuance evidenced, some expected (phase, subject) not evidenced”).

- Phases: `issuance`, `declared_effectiveness`, `reception`, `loading` are `evidenced`; the application of the new version @2 (`application / agent_rt`) is `not_declared` (rule P-ND). Flag `F-02` ("change not fully observed") is raised.
- Order relations: **every** cross-clock pair between `clk:agent` and `clk:issuer` resolves to `not_determinable_across_clocks` ("Clocks not related in the bundle: no order is determinable between them").
- Evidentiary intervals: `None`; contradictions: `[]`; operational divergences: `None`.

The evaluation was performed on the workstation at a declared time of `2026-10-12T16:13:58Z` (operator clock, declared in the snapshot, not verified by the Lab), producing snapshot SHA-256 `b6582ed1be33…53c7aa` and assessment SHA-256 `ffe1773daebb…842674`. On re-import, gate G-6 passed on all checks and the evaluation was stored as `imported` (Lab evaluation #1) at `2026-10-12 16:15:54.049159`. The workstation `traer_verify` verdict was `reproduced`, recorded as an operator-declared reproduction (see §12).

9. Variant_B: clocks with an explicitly declared relationship

Lab bundle #2 was received at `2026-10-12 16:25:13.926532` (Lab clock, UTC), with bundle SHA-256 `d9917492c871…2d65be` (8578 bytes) and the same deposit gate outcomes. The only material difference from variant_A is the declared clock relation (§7).

Observed engine output:

- INDEX state `partially_reconstructable`; decisive rule `T-07` — **identical to variant_A.**
- Phases: identical to variant_A; the application of the new version @2 (`application / agent_rt`) remains `not_declared`; flag `F-02` raised.
- Order relations: the same cross-clock pairs now resolve to `before` / `after`. Each ordered pair is annotated by the engine: "**Relation between clocks**

DECLARED in the bundle, not verified by the engine or by the Lab", with the declared basis repeated. In particular `e07_action_under_v1` (an action under the previous version @1) is ordered `after` `e05_issue_v2` and `after` `e06_eff_v2` (the issuance and declared effectiveness of the new version @2).

- Evidentiary intervals: `None`; contradictions: `[]`; operational divergences: `None`. No duration is computed between events on the related clocks.

The evaluation (declared time `2026-10-12T16:28:35Z`) produced snapshot SHA-256 `c17c8db33954…459dde` and assessment SHA-256 `4b591cddfeae…87d27e`; gate G-6 passed on import, stored as `imported` (Lab evaluation #2) at `2026-10-12 16:30:01.012713`; the workstation `traer_verify` verdict was `reproduced` (operator-declared).

10. Comparative results

The following table (a standalone A/B comparison, drawn entirely from the two Lab evaluation records) summarizes the observed differences and the invariants.

Dimension	variant_A	variant_B
Declared clock relation	none	clk:issuer = clk:agent (declared, not verified)
Bundle SHA-256 / length	`b7100fbfac88…a402dd` 8363 bytes	`d9917492c871…2d65be` 8578 bytes
INDEX state	`partially_reconstructable`	`partially_reconstructable`
Decisive rule	`T-07`	`T-07`
Evidenced phases	issuance, declared_effectiveness, reception, loading	issuance, declared_effectiveness, reception, loading
Application of new version @2	`not_declared (P-ND)`	`not_declared (P-ND)`
Cross-clock orderings	`not_determinable_across_clocks (all)`	`before / after (declared basis)`
e07_action_under_v1 vs e06_eff_v2	`not_determinable`	**`after`** (declared, not verified)
Evidentiary intervals	None	None
Contradictions / divergences	none / none	none / none
Workstation traer_verify	`reproduced (operator-declared)`	`reproduced (operator-declared)`
Lab import_integrity_verified	`yes (G-6)`	`yes (G-6)`

Dimension	variant_A	variant_B
Lab evaluation_reproduced	no (Lab does not run the engine)	no (Lab does not run the engine)

Read plainly: the declared clock relation **changed the derivable temporal orderings** (from not-determinable to before/after) while leaving **unchanged** the INDEX reconstructability state, the decisive rule, the set of evidenced phases, and the application-phase gap. Figure 1 summarizes this A/B contrast.

11. Evidentiary interpretation

Why does INDEX remain `partially_reconstructable` in both variants? Rule `T-07` is decisive because the issuance of the transition is evidenced while at least one expected (phase, subject) — here the application of the new version @2 by `agent_rt` — is not evidenced. The INDEX state is driven by the **presence or absence of phase evidence**, not by the temporal orderings among the events that are present. A declared clock relation changes the orderings that can be derived; it does not add or remove any phase evidence. It therefore did not change the INDEX state in these two runs. (See Figure 1.)

The correct conclusion is *not* that the declared relation adds nothing — it demonstrably adds derivable orderings. The conclusion is that the declared relation **modifies the derivable temporal relations but does not remove the evidentiary gaps that determine the INDEX state.**

The `e07-after-e06_eff_v2` ordering in variant_B deserves particular care, because it is exactly the result a naive reader would over-interpret. On the observed record:

- the ordering exists only because the clock relationship was declared in the bundle;
- that relationship was not independently verified (its basis is “declared for the experiment only … not verified”);
- `after` does not demonstrate a violation — the reading rules printed in the evaluation state that no order qualifies an action as legitimate, illegitimate, or a violation;

- the actual application of the new version @2 is **not** demonstrated; the application phase is `not_declared`, and absence of evidence is not evidence of absence;
- no responsibility can be attributed to any party on the basis of these results alone.

12. Integrity verification versus evaluation reproduction

Each Lab evaluation record keeps two levels strictly apart, under the heading "Evidentiary status asserted by the Lab." The distinction is central to the honesty of the result and is reproduced here verbatim in substance.

Asserted status	Meaning (as recorded by the Lab)
`import_integrity_verified`	yes: the stored snapshot and assessment bytes match their recorded digests and bind to the stored bundle and to an approved evaluator (gate G-6). It is a check of integrity and provenance, not of the evaluation itself.
`evaluation_reproduced (by the Lab)`	no: the Lab does not run the engine. A reproduced verdict, if any, was produced by the operator on the workstation and is recorded as declared, not verified. Because the exact bytes are kept, a third party holding the frozen package can reproduce the evaluation from them; the Lab asserts the integrity of the import, not the reproduction of the evaluation.

Three levels of checking must not be conflated. **(1)** The SHA-256 digests of the artifacts in this paper were recomputed and compared by the operator, with technical support, during the experiment — an operator-side integrity check, **not** an independent third-party reproduction. **(2)** The Lab performed its own integrity and binding checks through gates G-1…G-6, independently of the operator's checks but still without running the engine. **(3)** Reproducing the evaluation — re-running the frozen engine on the retained bytes — was done only on the operator's workstation and is recorded as `reproduced` (declared). No **independent validation**, **independent peer review**, or **independent reproduction** is claimed anywhere in this paper.

In both variants the Lab recorded `import_integrity_verified = yes` and `evaluation_reproduced = no`. The operator obtained a `reproduced` verdict from `traer_verify` on the workstation; in this paper that is described as an **operator-**

declared reproduction. The terms “certified”, “independently validated”, “legally proven” and “court-admissible” are deliberately **not** used anywhere in this paper to describe these results.

13. Limitations and threats to validity

The results are bounded in several explicit ways.

- The data are entirely synthetic. Executing the runs on the operational Lab platform does not transform synthetic declarations into real observations of an agent's behaviour.
- Only two controlled variants were run, one execution each. There is no statistical replication and no generalization is claimed.
- There is no independent observation of the agent: every evidence item is declared by the issuer or by the agent runtime.
- The clock relationship in variant_B is declared and not verified; the derived orderings inherit that status.
- The actual content of the agent's action is not recorded; the action log declares only which rule version the agent claimed to apply.
- No finding of compliance or violation is made, and none can be derived from these results.
- The incidental label differences between the bundles were reasoned to be non-semantic but were not formally ablated (e.g., by an otherwise-identical third bundle differing only in labels).
- Independent human review of the profile package has not been performed; independent confirmation of the frozen 0.3.2 engine remains open.
- No competitive benchmark was executed; no claim of superiority over any other system is made.

Internal validity note: the comparative conclusions attribute the change in orderings to the declared clock relation, which inspection of the inputs and of the engine's logic identifies as the semantically relevant input difference; a label-only ablation run was not performed (see the label bullet above). Construct validity

note: the synthetic scenario was designed to exhibit a phase gap and a cross-clock ordering question; it is illustrative of the method, not a sample of real-world incidents.

What would change or falsify these findings? Two kinds of new result would bear on the conclusions, and they are different in kind:

- **New scenario evidence** — for example, a declared `application / agent_rt` event for @2, or an independently observed agent log — would change the reconstruction of this scenario and could move the INDEX state (e.g., from `partially_reconstructable` toward `reconstructable` or `inconsistent`). It would revise what this scenario shows, not the engine's behaviour.
- **Reproduction results** — an independent re-execution of the frozen 0.3.2 engine on the retained bytes that diverged from the recorded snapshot and assessment, or that surfaced a gate/engine discrepancy — would call into question the documented behaviour of the engine itself, independently of any single scenario.

14. Reproducibility and available artifacts

The evaluations are reproducible from the retained bytes. A third party holding the frozen EVIDE TRAER 0.3.2 package (SHA-256 `f2fd9123fb03…331ee1`) can, for either bundle, recompute the bundle digest, run `traer_assess.py` to regenerate the snapshot and assessment, and run `traer_verify.py` to obtain the `reproduced` verdict; re-import into a Lab carrying the profile is accepted only if gate G-6 binds the artifacts to the stored bundle and to the approved evaluator. The artifact manifest below lists the real artifacts and their full digests; internal account identifiers and administrative metadata are deliberately omitted.

Artifact	Format	SHA-256 (full)	Provenance / time (clock)	Verification status
Evidence bundle (A)	traer_evidence_bundle_0	`b7100fbfac881dcdcb969bcb9f679190a3396ca2556414f0d4e3933dd2a402dd`	Synthetic; stored by Lab, received 2026-10-12 16:08:33.648532 (Lab, UTC)	Digest recomputed by Lab; G-1…G-4 passed
Snapshot (A)	traer_assessment_snapshot_1	`b6582ed1be33c511147702ec43ca3012075fe6606940222a1bfdac659353c7aa`	Workstation eval declared 2026-10-12T16:13:58Z (operator, not verified)	G-6 passed; bound to bundle A
Assessment (A)	traer_assessment_3	`ffe1773daebbdb91bbe3154ab135dc5f88e78d253ecdaff294892e9c33842674`	Engine output, workstation	G-6 passed; import_integrity_verified
Evidence bundle (B)	traer_evidence_bundle_0	`d9917492c87156d7c06e3e89fc3504031b26e772f4e35b587a1aac12f62d65be`	Synthetic; stored by Lab, received 2026-10-12 16:25:13.926532 (Lab, UTC)	Digest recomputed by Lab; G-1…G-4 passed
Snapshot (B)	traer_assessment_snapshot_1	`c17c8db33954a1700e89d045a1dae8da80fed14fc7786ada336e1912b6459dde`	Workstation eval declared 2026-10-12T16:28:35Z (operator, not verified)	G-6 passed; bound to bundle B
Assessment (B)	traer_assessment_3	`4b591cddfeae77147c81e7aa0916670d61da3f279b98d2926011351dcf87d27e`	Engine output, workstation	G-6 passed; import_integrity_verified
TRAER engine (frozen)	zip package	`f2fd9123fb03657e8493018a5a38f41689db7a110463d925fd133674db331ee1`	Frozen 0.3.2; approved evaluator for the experiment	Digest verified on workstation

Artifact	Format	SHA-256 (full)	Provenance / time (clock)	Verification status
Profile package	STAGE 0.1	9ddabad91e853d413436238c2de653be55529bafacd9a41c19e405e3fa23cec4	v_traer_index_01 (experimental; disabled by default, enabled by the authorized operator for these runs)	Local staging only; not independently reviewed

Engine source files (identical in both runs): `traer_engine.py` 02b81a39bf99d049720ec659db42065c25fafdc60a102e5781b336592dc4b1cc; `traer_assess.py` 9aea4ff37df58889d5743831a713dfab7ac68c49bccdde8e0747ed88eda5139e. Approval reference recorded on the Lab: “esperimento sintetico 1a - 2026-10-12”.

15. Implications for forensic examination and AI governance

The implications below follow from the demonstrated discipline, not from any claim of fitness for a specific legal or regulatory use.

For forensic examination: a method that reconstructs declared transitions while stating the evidentiary perimeter explicitly — which phases are evidenced, which are not_declared, which orderings rest on a declared clock relation, and what the state does not show — supports a disciplined reading of agent records. It makes the difference between “reconstructed from declarations” and “independently corroborated” visible at the level of each phase and each order, which is the level at which an examiner must reason.

For AI governance: audit and dispute contexts around agent behaviour routinely turn on records declared by heterogeneous systems on unsynchronized clocks. The separation of “we reconstructed a declared sequence” from “we verified the operational sequence”, and the refusal to convert an absence of evidence into a negative finding or a declared order into a violation, are the properties that make such reconstructions defensible rather than merely assertive. The further separation of import integrity from evaluation reproduction gives a custody record whose claims are scoped to what was actually checked.

16. Conclusions

A controlled A/B experiment, executed on the operational EVIDE Governance Lab with synthetic data and the frozen EVIDE TRAER 0.3.2 engine, shows that an explicitly declared relationship between two otherwise-unrelated clocks changes the temporal orderings EVIDE TRAER can derive, while leaving the EVIDE INDEX reconstructability state (partially_reconstructable, rule T-07) and the application-phase evidentiary gap unchanged. At every ordered relation the system marks the order as resting on a declaration, not a verification; it does not convert the gap into a negative finding, nor the declared order into a violation. The Lab asserts the

integrity and provenance of the imported artifacts and explicitly does not assert that it reproduced the evaluation.

The contribution is a verifiable demonstration of evidentiary discipline on synthetic data, with explicit limits, rather than a performance or compliance claim. Work not performed here, and required before any stronger use, includes: independent human review of the profile package; independent confirmation of the frozen 0.3.2 engine; experiments in the profile (format 1) mode against a frozen expectation profile; and broader, non-synthetic scenarios under appropriate review.

17. References and artifact manifest

No external peer review, DOI, normative reference, or third-party approval is claimed or cited. The references below are the system surfaces and the frozen artifacts used in the experiment.

- EVIDE Governance Lab — experimental research platform: lab.certifywebcontent.com (profile `v_traer_index_01`, experimental, phase 1).
- EVIDE Canonical — framework site: app.certifywebcontent.com (operated by CertifyWebContent, Informatica in Azienda, Bologna).
- EVIDE TRAER 0.3.2, frozen engine package — SHA-256 `f2fd9123fb03657e8493018a5a38f41689db7a110463d925fd133674db331ee1`.
- EVIDE INDEX rule set and perimeter vocabulary — as embedded in the engine output (rules T-01…T-10; phase rules P-EV, P-ND; flag F-02; decisive rule for this experiment: T-07).
- EVIDE Lab TRAER + INDEX STAGE 0.1 package report — profile package SHA-256 `9ddabad91e853d413436238c2de653be55529bafacd9a41c19e405e3fa23cec4`.

The full artifact manifest with SHA-256 digests is given in §14.

Appendix A — Executive summary (one page)

Problem. Records of an AI agent's rule transitions are declared by different systems on clocks that are not known to be synchronized. Reading such records naively — for example, treating an action timestamped after a revocation as a violation — is unsafe, because declared, cross-clock timestamps do not by themselves establish an operational order or a finding.

Experiment. Two synthetic evidence bundles describing one rule revocation — sharing the same nine evidence items and declared timestamps, with one intended experimental variable — were deposited on the operational EVIDE Governance Lab, evaluated off-server with the frozen EVIDE TRAER 0.3.2 engine, and re-imported under the Lab's provenance gate. In variant_A the two clocks have no declared relationship; in variant_B the bundle declares them as the same (unverified) time source.

Results. EVIDE TRAER reconstructs the revocation and its phases; EVIDE INDEX reports how supportable that reconstruction is. The declared clock relation changed the derivable orderings (from not-determinable to before/after, each marked “declared, not verified”), but the INDEX state stayed partially_reconstructable in both variants and the agent's application of the new version @2 remained not_declared. A declared relation produced orderings; it produced neither verification nor closure of the evidentiary gap.

Integrity vs reproduction. The Lab verified the integrity and provenance of the imported artifacts (digests, binding, approved evaluator) and explicitly did not reproduce the evaluation itself; the operator's reproduced verdict is recorded as declared, not independently verified.

Limits. Fully synthetic data; two single runs; no independent observation of an agent; a declared, unverified clock relation; no finding of compliance or violation; independent human review and independent confirmation of the engine still pending; no benchmark and no comparison to other systems.

Takeaway. The method reconstructs what the evidence declares, states how far that reconstruction is supportable, and refuses to overstate — a property demonstrated here on verifiable artifacts rather than asserted.

Experimental research output based exclusively on synthetic data. Not a forensic certification, not evidence of real-world agent conduct, and not a legal finding.

Appendix B — Short abstract (for future public presentation)

We ran a controlled A/B experiment on the EVIDE Governance Lab (synthetic data, frozen EVIDE TRAER 0.3.2): one rule revocation, two evidence bundles sharing the same nine evidence items and declared timestamps, with one intended experimental variable: whether a relationship between the two clocks is declared. Result: declaring the clock relationship changed which event orderings the engine could derive — but the EVIDE INDEX reconstructability state stayed partially_reconstructable in both cases, and the agent's application of the new version @2 stayed not_declared. EVIDE TRAER reconstructs what the evidence permits; EVIDE INDEX states how far that is evidentially supportable; and every order that rests on a declaration is labelled as declared, not verified. An “after” is not a violation; a missing phase is not a proof of absence; and verifying the integrity of an import is not the same as reproducing the evaluation.

Experimental research output based exclusively on synthetic data. Not a forensic certification, not evidence of real-world agent conduct, and not a legal finding. Independent review pending.

Appendix C — Claims that require further proof before publication

The following statements are not supported by the two runs alone and must not appear as established claims in any public version without the indicated additional work.

Statement to avoid / qualify	What would be required
“The engine / method is independently validated or certified.”	Independent human review and an independent re-execution by a third party; neither performed.
“EVIDE TRAER 0.3.2 is confirmed correct.”	Independent confirmation of the 0.3.2 regression and mutation results (open).
“The agent did / did not apply the new version @2.”	Evidence of the application phase; it is not_declared in both runs — no such claim is derivable.
“The action after the revocation is a violation.”	A verified operational order and a governing-rule finding; neither exists. “After” here rests on a declared, unverified clock relation.
“Results were obtained with real agents in production.”	Non-synthetic data and independent observation of a real agent under review; not done.
“EVIDE outperforms / beats other systems.”	A defined benchmark and a comparative study; none executed. No comparative claim is made.
“The results are court-admissible or legally proven.”	Outside the scope and the stated status (experimental, synthetic-data research output; not a legal finding); do not assert.
“Label differences between bundles had no effect.”	A third ablation bundle differing only in labels would formally establish it; reasoned, not ablated.

Claim-by-claim control: every factual statement in §1–§17 is tied to one of — an artifact digest, a value observed in a Lab evaluation record, or an engine/profile specification. Statements of significance are marked as interpretation (§11, §15). Unsupported or stronger claims are confined to this appendix as items to be proven, not asserted.