Encevra Research Round 37: When the Contract Changed After Signing: Event-Date Resolution, Register-Anchored Completeness, and Stale Registers in Multi-Witness Contract Audits

Date: 2026-10-28
Research lab: Encevra Research
Short code: `TDR-37` (Temporal Drift and Post-Execution Supersession)
Primary tracks: Track 1 (Evidence, Provenance, and Change), Track 2 (Selective Verification Under Budget), Track 5 (Calibration, Abstention, and Release Control)
Carry-forward boundary: Round 36 (`CCS-36`) closed the multi-document reference graph of a contract packet and reached `12 / 12` correct releases once every instrument named by the base agreement was attached. Its Section 10.2 and Section 11 handed off one boundary explicitly: a later amendment, waiver, termination notice, or revocation can change an obligation *silently*, with no forward pointer in the base agreement, and in Round 36 the active or superseded status of each instrument was supplied as a label in the packet. Round 37 removes both supports. There is no pointer to follow and no status label to read.

Abstract

Contracts change after signing. Payment terms are shortened by amendment, audit rights are relaxed by waiver letters, support obligations end by termination notice, and waivers are later revoked. None of these instruments needs to be mentioned in the base agreement, so a reviewer holding a perfectly complete base agreement can still apply a rule that was no longer in force on the date of the event under audit. Round 36 showed that reference closure (`CCS-36`) catches an unattached instrument when the base agreement names it. Round 37 asks what catches an instrument that nothing names, and what decides which instrument governed on the event date once all of them are in hand.

We define **Temporal Drift and Post-Execution Supersession (`TDR-37`)**, a release gate with four conjuncts: register coverage against an externally anchored executed-instrument register (`V_reg`), a 3-of-4 heterogeneous witness quorum (`V_quorum`), agreement between the cited controlling instrument and a deterministic as-of resolver evaluated at the event date over effective dates, retroactivity, expiry, and scope (`V_asof`), and a verbatim span verified inside the controlling instrument (`V_span`). We report three strictly separated evidence categories.

Deterministic (`E37-A`, no model calls). An exact enumeration of `1,536` feasible lifecycle-evidence states (`16` positive, `1,520` negative) shows that `CCS-36` reference closure still admits `208` negative states (`FAR 13.68%, 80` with a wrong verdict), and that replacing as-of resolution by a latest-executed-wins rule rejects `8 / 16` valid states (`FRR 50.00%`). Accepting a self-attested register admits `24` negative states and accepting a stale register admits `8`; the full gate admits `0`. The full gate coincides with the positive-state definition by construction, so this is a specification consistency check, not evidence of performance. On the `12` frozen lifecycles, the as-of resolver matches the authored ground truth `12 / 12`, latest-executed-wins `7 / 12`, executed-before-event `9 / 12`, and base-only `5 / 12`.

Live (`E37-B`, `96` live evaluations). Twelve authored contract lifecycles (`L01`-`L12`) were audited by four witness architecture families (`Witness A` on Max reasoning, `Witness B`, `Witness C`, `Witness D`) under two views. With only the base agreement (View B), every lifecycle produced a unanimous `4-0` vote, and `7 / 12` were unanimously wrong. Ungated voting and `CCS-36` both released all `7` as false certifications, because `0 / 48` witnesses reported any unattached reference: there was nothing to report. `TDR-37` held `11 / 12` on the register gap with `0` false releases. With the full lifecycle and a register extract (View F), all `48 / 48` witness answers were correct and cited the as-of controlling instrument, so every arm, including a single `Witness A` witness, released `12 / 12` correctly. Temporal decoys (future-effective amendments, one-time waivers, notices that take effect after the event, retroactive amendments) did not mislead any witness once the instruments were present. We report this as a ceiling result that rejects our own hypothesis `H37.4`.

Live (`E37-C`, `48` live evaluations). In a stale-register probe (View S), the most recently executed instrument was removed from both the packet and the register extract. On the `5` lifecycles where that instrument governs the event, all `20 / 20` witness answers were wrong, and a `TDR-37` gate that trusts the register extract shipped inside the packet released all `5` as false certifications (`5 / 12`). The same gate reading a register of record that it queries itself released `0` false certifications and held `11 / 12`. Witness self-reports of completeness did not provide a reliable substitute: `Witness A` flagged completeness as unverifiable exactly when the extract's query date preceded the event date (`7 / 7`), but on `L08`, where a retroactive amendment signed after the query flips the verdict, all `4 / 4` witnesses asserted that completeness was verifiable and all `4 / 4` were wrong.

The measured contribution is narrow and practical: the decisive failure is evidence access, not temporal reasoning, and the only check that worked was one that does not depend on the packet describing itself.

---

1. Archive continuity and anti-repetition verification

Before designing Round 37 or executing any model call, we reviewed the published Encevra SI Research Labs archive across all `30` indexed prior rounds (Rounds 3 and 8 through 36) and the Round 35 and Round 36 papers to establish the carry-forward boundary for Round 37.

Round	Date	Core question	Boundary handed to Round 37
Round 18	2026-10-09	Versioned support and bounded belief repair when evidence changes (bitemporal VSDL)	Deterministic only, abstract claim records, no model calls
Round 22	2026-10-13	When does an audit expire after repository change	Software-claim staleness, not contract applicability
Round 25	2026-10-16	Detecting missing change evidence in collaborative AI systems	Observation-path silence on authored traces
Round 26	2026-10-17	Minimum observables for omitted change	Abstract observables; no instrument register tested
Round 35	2026-10-27	Correlated semantic failures among grounded witnesses (`CIS-35`)	Single-document bundles only

Round	Date	Core question	Boundary handed to Round 37
Round 36	2026-10-28	Cross-document reference closure and precedence (`CCS-36`)	Explicit pointers and status labels supplied; silent post-execution instruments handed off as `TDR-37`

Why Round 37 is not a repeat. The closest prior work is Round 36, whose packet `M11` contained an amendment marked `SUPERSEDED` or `ACTIVE` and whose base agreement named every companion instrument. Round 37 differs in four ways: (1) no base agreement contains a pointer to any later instrument, so reference closure has nothing to close; (2) no instrument carries a status label, so applicability must be derived from execution date, effective date, expiry, retroactivity, and scope relative to the event date; (3) completeness is anchored to an external register rather than to the reference graph; (4) the round adds a stale-register condition that no prior round tested. Round 18 shares the valid-time vocabulary but evaluated abstract claims with no model witnesses. Rounds 25 and 26 studied missing change evidence in agent traces, not in legal instrument lineage. A keyword scan of the registry and the Round 35 and 36 manuscripts for `effective date`, `as-of`, `event date`, `one-time`, `retroactive`, `instrument register`, and `contract register` returned no hits outside the Round 36 handoff text.

---

2. Executive summary

Evidence	What was measured	Result
`E37-A` deterministic	`1,536` feasible states, `7` gates	`CCS-36` closure `FAR 13.68%`; latest-wins `FRR 50.00%`; self-attested register `FAR 1.58%`; stale register `FAR 0.53%`; full gate `0 / 0` by construction
`E37-A` resolvers	`4` resolvers on `L01` - `L12`	as-of `12 / 12`, executed-before-event `9 / 12`, latest-wins `7 / 12`, base-only `5 / 12`
`E37-B` View B	`48` live evaluations	`12 / 12` unanimous votes, `7 / 12` unanimously wrong, `0 / 48` dangling-reference flags; ungated and `CCS-36` `7 / 12` false releases; `TDR-37` `0` false, `11` holds
`E37-B` View F	`48` live evaluations	`48 / 48` correct, `48 / 48` cite the as-of instrument; every arm `12 / 12`
`E37-C` View S	`48` live evaluations	`20 / 20` wrong on the `5` decisive lifecycles; packet-register `TDR-37` `5 / 12` false releases; gate-side register `0` false, `11` holds

---

3. Problem statement

Let a contract lifecycle be `Lambda = (D_0, I, e)`, where `D_0` is the base agreement, `I = {I_1, ..., I_k}` is the set of executed post-execution instruments, and `e` is an event with date `t_e` and scope `sigma_e`. Each instrument carries execution date `x_j`, effective start `a_j`, optional retroactive start `r_j`, optional end `b_j`, target section `s_j`, scope `sigma_j` (continuing, or one named instance), and a replacement rule. The rule start is `rho_j = r_j` if a retroactive start exists, otherwise `a_j`. Instrument `I_j` *applies* to event `e` on section `s` when `s_j = s`, `sigma_j` is continuing or equals `sigma_e`, `rho_j ⩽`

`t_e`, and `t_e ⩽ b_j` when `b_j` exists. The as-of controlling instrument is the applying instrument with the latest rule start, or `D_0` if none applies.

The audit question asks whether `e` complies with the obligation in force at `t_e`. Three failure modes follow:

1. **Silent supersession.** An applying instrument exists but is absent from the packet, and nothing in `D_0` reveals its existence.
2. **Temporal decoys.** An instrument is present but does not apply at `t_e` (future-effective, one-time for a different instance, termination taking effect after the event) or applies despite being signed after the event (retroactive).
3. **Stale completeness evidence.** The packet includes a register extract that looks complete but was queried before the decisive instrument was executed.

Round 36's `CCS-36` addresses none of these by design: its closure predicate walks references from `D_0`, and silent instruments are invisible to that walk.

---

4. Related work (all sources fetch-verified)

Contract understanding benchmarks. CUAD [S01] and ContractNLI [S02] evaluate clause extraction and document-level inference over single contracts. LegalBench [S07] collects legal reasoning tasks. None evaluates which of several instruments governs on a given date.

Time-sensitive question answering. Dhingra et al. [S03] show that language models answer time-scoped facts poorly without explicit time conditioning. Chen, Wang, and Wang [S04], SituatedQA [S05], RealTime QA [S06], and Tan, Ng, and Bing [S08] measure temporal and context-dependent answering. Our View F result is consistent with a narrower claim: when every dated instrument is in context and the event date is explicit, the task becomes in-context interval reasoning, which our witnesses performed without error on `12` lifecycles.

Valid time and interval logic. Snodgrass and Ahn [S11] distinguish valid time from transaction time; Jensen and Snodgrass [S13] survey temporal data management. Allen [S12] formalizes interval relations. The as-of resolver is a valid-time query with retroactivity treated as a valid-time start that precedes the transaction (execution) time.

Append-only registers. Certificate Transparency [S14] makes completeness checkable by requiring that certificates be logged in a public append-only log that clients query themselves. `E37-C` reproduces the reason this design exists: a completeness claim carried inside the evidence being checked is not independent evidence of completeness.

Agreement and context effects. Self-consistency [S09] treats agreement across samples as a confidence signal; Lost in the Middle [S10] documents position effects in long contexts. Our View B result shows agreement is uninformative when every witness lacks the same instrument.

---

5. Formal model and the `TDR-37` gate

Let `R*` be the register of record, an externally maintained list of executed instruments for the agreement, queried by the gate at evaluation time. Let `P` be the set of instruments attached to the packet, and let `W = {w_1, ..., w_4}` be witness receipts, each with verdict `v_i`, controlling instrument `c_i`, and quote `q_i`.

`TDR-37` releases verdict `v` if and only if:

1. **`V_reg`:** `R* subseteq P`. Every registered instrument is attached.
2. **`V_quorum`:** at least `3` witnesses from at least `3` architecture families report `v`.
3. **`V_asof`:** at least `3` of those witnesses cite `c_i = AsOf(R*, t_e, sigma_e)`, where `AsOf` is the deterministic resolver of Section 3 evaluated over the structured records of `R*`.
4. **`V_span`:** for those same witnesses, `norm(q_i)` is a contiguous substring of the normalized text of `c_i`.

Otherwise the gate holds with a named reason (`HOLD_TDR37_REGISTER_GAP`, `HOLD_SPLIT`, `HOLD_TDR37_DIVERSITY`, `HOLD_TDR37_ASOF_OR_SPAN`).

Proposition 37.1 (reference-closure blindness). If no instrument in `I` is referenced by `D_0` or by any attached instrument, any gate whose completeness predicate is reference closure from `D_0` accepts a packet `P = {D_0}` as complete. *Proof.* The closure of `{D_0}` under an empty reference relation is `{D_0}`. With `n` independent post-execution instruments, each pointer-free with probability `f` and outcome-changing at `t_e` with probability `c`, the probability that reference closure is blind to at least one decisive instrument is `P_blind = 1 - (1 - f c)^n`; for `n = 4`, `f = 0.5`, `c = 0.25`, `P_blind = 0.4138`, and for `n = 8` it is `0.6564`. These are arithmetic consequences of the stated assumptions, not estimates of real contract portfolios.

Proposition 37.2 (self-describing completeness is circular). If the register used by `V_reg` is supplied inside the packet, then omitting an instrument from both the packet and the supplied register satisfies `V_reg`, and `V_asof` evaluated over that register returns the stale answer. *Proof.* Both checks are then functions of the packet alone. `E37-C` arm A7 instantiates this.

Proposition 37.3 (latest-executed is not as-of). A resolver that selects the most recently executed applicable-section instrument disagrees with `AsOf` whenever an instrument is future-effective, one-time for a different scope, terminates after the event, or is followed by an executed instrument whose rule start is later than `t_e`. `L05`, `L06`, `L07`, `L09`, and `L10` instantiate these cases.

6. Hypotheses

ID	Statement	Test	Status
`H37.1`	With base-only packets, pointer-free post-execution instruments produce unanimous false consensus that `CCS-36` does not intercept	`E37-B` arms A2, A3	Supported
`H37.2`	Reference closure and latest-executed-wins resolution leave measurable false acceptance or false rejection in the declared universe	`E37-A` G1, G2	Supported

ID	Statement	Test	Status
H37.3	TDR-37 with a gate-side register holds base-only packets with zero false releases	E37-B arm A5	Supported
H37.4	With every instrument attached, temporal decoys still cause witness errors that V_asof must catch	E37-B View F	Rejected
H37.5	A register extract supplied inside the packet defeats TDR-37 when it is stale, and a gate-side register prevents the false releases	E37-C arms A7, A8	Supported
H37.6	Witness self-reported completeness reliably identifies stale register extracts	E37-C flags	Rejected

7. Evidence category 1: deterministic structural diagnostic (E37-A)

7.1 Design

The universe is the product of eight variables: forward pointer (pointer, no_pointer), packet inclusion of the decisive instrument (included, omitted), register (none, self_attested, external_complete, external_stale), temporal relation of the instrument to the event (applies, future, retroactive, out_of_scope), quorum outcome (asof_correct, latest_wins, base_rule, split), span (verified, unverified), family diversity (ge3, lt3), and event-date binding (bound, unbound). The raw product is 2,048. We exclude 512 infeasible states in which witnesses cite an instrument that is not in the packet, leaving 1,536 feasible states. A state is positive when the register is external and complete, the instrument is included, the quorum is grounded and diverse, and the quorum's verdict and controlling instrument equal the as-of truth with the event date bound: 16 positive and 1,520 negative states. The gate's own as-of check uses only instruments listed in the register the gate trusts; an earlier draft of the script let the check read the true controlling instrument, which leaked the oracle, and was corrected before the reported run.

7.2 Gate comparison

Gate	TP	FP	of which wrong verdict	of which right verdict, unwarranted	TN	FN	FAR	FRR
G0 quorum + span + diversity	16	240	96	144	1,280	0	15.79%	0.00%
G1 CCS-36 reference closure	16	208	80	128	1,312	0	13.68%	0.00%
G2 CCS-36 + latest-executed-wins	8	120	48	72	1,400	8	7.89%	50.00%
G3 register closure, no as-of	16	32	16	16	1,488	0	2.11%	0.00%
G4 TDR-37, self-attested register accepted	16	24	4	20	1,496	0	1.58%	0.00%
G5 TDR-37, stale register accepted	16	8	4	4	1,512	0	0.53%	0.00%
G6 TDR-37 full	16	0	0	0	1,520	0	0.00%	0.00%

Reading the table honestly. `G6` equals the positive-state predicate, so its perfect row is guaranteed by construction and is not evidence. The informative rows are the others: each one isolates which missing conjunct admits which kind of negative state. "Right verdict, unwarranted" counts states where the released verdict happened to be correct but was not supported by the evidence the gate required (for example, the right verdict reached by citing the base rule while an applying instrument was omitted).

7.3 Resolver consistency on the frozen lifecycles

Resolver	Verdict correct	Controlling instrument correct	Lifecycles missed
As-of (valid time, retroactivity, expiry, scope)	12 / 12	12 / 12	none
Executed before event	9 / 12	9 / 12	`L06`, `L07`, `L08`
Latest executed wins	7 / 12	7 / 12	`L05`, `L06`, `L07`, `L09`, `L10`
Base agreement only	5 / 12	5 / 12	`L01`-`L04`, `L08`, `L09`, `L10`

This table checks that the authored ground truth is internally consistent with the stated as-of semantics and that the decoy lifecycles discriminate between plausible resolution rules. It is not model evidence.

---

8. Evidence category 2: live multi-witness benchmarks (`E37-B`, `E37-C`)

8.1 Witness roster and protocol

Every witness received the identical system instruction and the identical packet for a given view, with Max reasoning (`reasoning: {"effort": "high"}`) for all four families. Each receipt was required to be JSON with `verdict`, `controlling_instrument`, `controlling_section`, `verbatim_quote`, `as_of_date`, `instruments_not_in_effect`, `unattached_external_reference`, and `lifecycle_completeness_verifiable`. The instruction asked witnesses to determine the governing instrument as of the event date but did not name any lifecycle's trap. Quotes were verified by normalized substring match (minimum `18` characters) against the cited instrument's text in the packet. All prompts are hashed per receipt in the ledgers.

Witness	Architecture	`E37-B` evaluations	`E37-C` evaluations	Parsed	Mean latency `E37-B`
`Witness A` (Max reasoning)	Family A	24	12	36 / 36	5,483.3 ms
`Witness B`	Family B	24	12	36 / 36	3,359.0 ms
`Witness C`	Family C	24	12	36 / 36	7,100.9 ms
`Witness D`	Family D	24	12	36 / 36	19,628.9 ms
Total		96	48	144 / 144	

One additional `Witness A` (Max reasoning) call produced the critical synthesis review in Section 9.3; it is interpretation input only. Seven `E37-C` calls were deferred by the round's internal pre-call guard during the parallel run and were then executed sequentially with the identical packet and instruction; all `48` planned `E37-C` receipts exist.

8.2 Frozen lifecycles

ID	Stratum	Situation	Event date	Truth	Controlling instrument
`L01`	S1 silent supersession	Payment term shortened from 60 to 30 days by amendment	2026-07-15	FAIL	`DOC-AMD1`
`L02`	S1 silent supersession	Audit notice relaxed for 2026 by waiver letter	2026-05-12	PASS	`DOC-WVR1`
`L03`	S1 silent supersession	Support obligation ended by termination notice	2026-09-03	FAIL	`DOC-TRM1`
`L04`	S1 silent supersession	Renewal price-increase cap raised from 3% to 7%	2026-03-01	PASS	`DOC-AMD2`
`L05`	S2 temporal decoy	Liability cap raise effective only after the event	2026-06-14	PASS	`DOC-MSA`
`L06`	S2 temporal decoy	One-time subprocessor waiver for a different subprocessor	2026-07-01	FAIL	`DOC-MSA`
`L07`	S2 temporal decoy	Termination takes effect after the event	2026-09-15	PASS	`DOC-MSA`
`L08`	S2 temporal decoy	Amendment signed after the event, expressly retroactive	2026-06-30	PASS	`DOC-AMD1`
`L09`	S3 chain	Event falls between first and second amendments	2026-05-10	PASS	`DOC-AMD1`
`L10`	S3 chain	Waiver in force in March, revoked from May	2026-03-31	FAIL	`DOC-WVR1`
`L11`	S4 control	Register lists only the base agreement	2026-08-04	FAIL	`DOC-MSA`
`L12`	S4 control	Later amendment changes an unrelated section	2026-06-18	PASS	`DOC-MSA`

View B contains the base agreement and the event record. View F adds a register extract and every registered instrument, in a fixed order that is not chronological for chains. View S is View F with the most recently executed instrument removed from both the packet and the extract, and with the extract's query date set one day before that instrument's execution.

8.3 `E37-B` arm results

Arm	View	Correct releases	False releases	Holds
A1 single `Witness A` (Max reasoning)	F	12 / 12	0 / 12	0 / 12
A2 ungated 3-of-4 quorum	B	5 / 12	7 / 12	0 / 12

Arm	View	Correct releases	False releases	Holds
A3 `CCS-36` (oracle-free)	B	5 / 12	7 / 12	0 / 12
A4 `CCS-36` (oracle-free)	F	12 / 12	0 / 12	0 / 12
A5 `TDR-37`	B	1 / 12	0 / 12	11 / 12
A6 `TDR-37`	F	12 / 12	0 / 12	0 / 12

The `CCS-36` arms here are oracle-free: they use witness-reported dangling references and witness-cited sections rather than ground-truth sections, which differs from Round 36's evaluation and is noted for comparability. A5's one release is `L11`, the control whose register of record lists only the base agreement, so the register gap check passes legitimately.

8.4 `E37-B` per-lifecycle votes

Votes are listed in roster order (`Witness A`, `Witness B`, `Witness C`, `Witness D`); `P` is PASS and `F` is FAIL.

ID	Truth	View B votes	View B correct	View F votes	View F correct	A2 / A3	A5	A6
`L01`	FAIL	PPPP	0 / 4	FFFF	4 / 4	false release	hold (register gap)	correct
`L02`	PASS	FFFF	0 / 4	PPPP	4 / 4	false release	hold (register gap)	correct
`L03`	FAIL	PPPP	0 / 4	FFFF	4 / 4	false release	hold (register gap)	correct
`L04`	PASS	FFFF	0 / 4	PPPP	4 / 4	false release	hold (register gap)	correct
`L05`	PASS	PPPP	4 / 4	PPPP	4 / 4	correct	hold (register gap)	correct
`L06`	FAIL	FFFF	4 / 4	FFFF	4 / 4	correct	hold (register gap)	correct
`L07`	PASS	PPPP	4 / 4	PPPP	4 / 4	correct	hold (register gap)	correct
`L08`	PASS	FFFF	0 / 4	PPPP	4 / 4	false release	hold (register gap)	correct
`L09`	PASS	FFFF	0 / 4	PPPP	4 / 4	false release	hold (register gap)	correct
`L10`	FAIL	PPPP	0 / 4	FFFF	4 / 4	false release	hold (register gap)	correct
`L11`	FAIL	FFFF	4 / 4	FFFF	4 / 4	correct	correct	correct
`L12`	PASS	PPPP	4 / 4	PPPP	4 / 4	correct	hold (register gap)	correct

Across the `48` View B receipts, `0` reported an unattached external reference, `0` asserted lifecycle completeness was verifiable, and `48 / 48` quotes verified verbatim in the cited instrument (the base agreement). Across the `48` View F receipts, `48 / 48` cited the as-of controlling instrument and `48 / 48` quotes verified.

8.5 `E37-C` stale-register probe

ID	Truth	Omitted instrument	Omission changes controlling instrument	View S votes	Correct	Completeness flag true	A7 packet register	A8 gate-side register
`L01`	FAIL	`DOC-AMD1`	yes	PPPP	0 / 4	2 / 4	false release	hold (register gap)

ID	Truth	Omitted instrument	Omission changes controlling instrument	View S votes	Correct	Completeness flag true	A7 packet register	A8 gate-side register
L02	PASS	DOC-WVR1	yes	FFFF	0 / 4	2 / 4	false release	hold (register gap)
L03	FAIL	DOC-TRM1	yes	PPPP	0 / 4	1 / 4	false release	hold (register gap)
L04	PASS	DOC-AMD2	yes	FFFF	0 / 4	3 / 4	false release	hold (register gap)
L05	PASS	DOC-AMD1	no	PPPP	4 / 4	4 / 4	correct	hold (register gap)
L06	FAIL	DOC-WVR1	no	FFFF	4 / 4	2 / 4	correct	hold (register gap)
L07	PASS	DOC-TRM1	no	PPPP	4 / 4	1 / 4	correct	hold (register gap)
L08	PASS	DOC-AMD1	yes	FFFF	0 / 4	4 / 4	false release	hold (register gap)
L09	PASS	DOC-AMD2	no	PPPP	4 / 4	4 / 4	correct	hold (register gap)
L10	FAIL	DOC-RVK1	no	FFFF	4 / 4	4 / 4	correct	hold (register gap)
L11	FAIL	none	no	FFFF	4 / 4	4 / 4	correct	correct
L12	PASS	DOC-AMD1	no	PPPP	4 / 4	3 / 4	correct	hold (register gap)

Totals: 28 / 48 witness answers correct; A7 7 correct releases, 5 false releases, 0 holds; A8 1 correct release, 0 false releases, 11 holds. Completeness flags set true by family across the 12 View S packets: Witness A 5, Witness B 10, Witness C 12, Witness D 7. The Witness A flag was false on exactly the 7 lifecycles whose extract query date precedes the event date and true on the other 5.

---

9. Findings and interpretations

9.1 Measured findings

1. **F37.1 (deterministic, E37-A).** In the declared 1,536-state universe, CCS-36 reference closure admits 208 negative states (80 with a wrong verdict); adding latest-executed-wins resolution reduces admitted negatives to 120 but rejects 8 / 16 positives; register closure without as-of resolution admits 32; a self-attested register admits 24; a stale register admits 8; the full gate admits 0 by construction.
2. **F37.2 (deterministic, E37-A).** On L01–L12, the as-of resolver matches authored truth 12 / 12; executed-before-event 9 / 12; latest-executed-wins 7 / 12; base-only 5 / 12.

3. `F37.3` **(live,** `E37-B` **View B).** All `12 / 12` lifecycles produced unanimous `4-0` votes. `7 / 12` were unanimously wrong (`L01`-`L04`, `L08`, `L09`, `L10`), and ungated voting and oracle-free `CCS-36` each released all `7` as false certifications. `0 / 48` receipts reported an unattached reference.
4. `F37.4` **(live,** `E37-B` **View B).** `TDR-37` held `11 / 12` on the register gap with `0` false releases; its only release was the control `L11`.
5. `F37.5` **(live,** `E37-B` **View F).** All `48 / 48` witness answers were correct and cited the as-of controlling instrument, including every temporal decoy. Every arm, including a single `Witness A` witness, released `12 / 12` correctly. `V_asof` never changed an outcome in this view.
6. `F37.6` **(live,** `E37-C`**).** On the `5` lifecycles where the omitted instrument controls the event, `20 / 20` witness answers were wrong. A `TDR-37` gate trusting the packet's register extract released all `5` falsely; with a gate-side register of record, `0` false releases and `11` holds.
7. `F37.7` **(live,** `E37-C`**).** Completeness self-reports were inconsistent across families (true on `5`, `10`, `12`, and `7` of `12` packets). On `L08` all `4 / 4` witnesses asserted completeness was verifiable and all `4 / 4` were wrong.

9.2 Interpretations

1. **The failure is access, not reasoning, on this benchmark.** The same witnesses that were unanimously wrong in View B were unanimously right in View F. Within these `12` lifecycles, the temporal reasoning step was not the bottleneck; knowing that an instrument exists was. We do not generalize this to longer lifecycles, overlapping partial amendments, or conditional instruments, which this round did not test.
2. **Agreement carries no information when every witness lacks the same document.** Four architecture families produced `24 / 24` unanimous votes across Views B and F. Heterogeneity did not decorrelate errors whose cause is shared missing evidence, which is consistent with Round 35's correlated-failure result in a new setting.
3. **Completeness must be checked against something the packet does not control.** `E37-C` shows Proposition 37.2 in practice: a stale extract inside the packet made the gate's own checks agree with the stale answer. The gate-side register worked only because the gate queried a source the packet could not edit. This is the same design reason that Certificate Transparency has clients query logs rather than trusting a certificate's own claims.
4. **Self-reported completeness is a weak heuristic, not a check.** `Witness A` appears to have compared the extract's query date with the event date. That heuristic flags harmless staleness (`L06`, `L07`, `L12`) and misses staleness that matters when a retroactive instrument is signed after both the event and the query (`L08`).
5. **The price of safety in View B and View S is abstention.** `TDR-37` turned `7` false releases into holds but also held `4` lifecycles whose base-only answer was correct. That is the expected price of refusing to certify without lineage, and it is only acceptable if the register can be fetched quickly so that a hold becomes a reopen.

9.3 Critical synthesis review (interpretation input only)

After the results were frozen, one `Witness A` (Max reasoning) call reviewed the result tables. Its main points, which we accept and reflect in Section 10, were: (a) the 3-of-4 quorum was never exercised because every vote was unanimous, so this round measures no value from multiple witnesses; (b) the View F ceiling may reflect that, once instruments are present and dated, the task reduces to lookup; (c) arm A8 relies on privileged access to the register of record and demonstrates the design requirement, not a detection capability; (d) the full gate's zero row on structured records says nothing about errors inside the register itself (wrong effective dates, misfiled scope, unregistered side letters, conditional or defective instruments). The full review text is in the reproducibility package.

---

10. Limitations and threats to validity

1. **Authored lifecycles, small n.** The `12` lifecycles were written by the researchers and are short. Ground truth is defined by the authored structured records, and the as-of resolver reads those records. Results bound only this benchmark.
2. **The quorum was not exercised.** All `36` lifecycle-view vote sets across `E37-B` and `E37-C` were unanimous, so `V_quorum` never decided an outcome. This round provides no evidence for or against multi-witness voting as such.
3. **`G6` is exact by construction.** The full gate's deterministic row restates the definition of a positive state. Only the gaps in `G0`-`G5` are informative.
4. **Structured register assumption.** `V_asof` uses machine-readable effective dates, retroactive starts, expiry, and scope in the register. Real registers often hold only titles and signature dates, which would push interval extraction back onto witnesses.
5. **Gate-side register trust.** A8 assumes the register of record is itself complete and current. A register with errors, delays, or unregistered instruments is untested.
6. **Single prompt and single sample.** Each witness answered once under one instruction. Prompt sensitivity and sampling variance were not measured.
7. **Research boundary.** No production Noetis service, customer workflow, or release gate changed.

---

11. Next check (Round 38 handoff)

Round 37 held access constant in View F and found no reasoning errors. The open question is what happens when access is perfect but instruments are *defective or conditional*: a waiver conditioned on an unmet condition precedent, a termination notice served on the wrong address or without the required counterpart signature, an amendment executed by an unauthorized signatory, or two registered amendments that modify overlapping parts of one clause with different scopes. These cases make an instrument present and registered but not effective, and they cannot be decided by dates alone. A proposed Round 38 diagnostic would hold register coverage and as-of resolution fixed, vary instrument validity conditions, and measure whether witnesses distinguish executed from effective, with

deliberate decorrelation (different instructions per witness) so that split votes can actually occur. The next round must re-run the anti-repetition check against Rounds 35 to 37 before adopting this framing.

---

12. References

- [S01] Dan Hendrycks, Collin Burns, Anya Chen, Spencer Ball. CUAD: An Expert-Annotated NLP Dataset for Legal Contract Review. arXiv:2103.06268, 2021.
- [S02] Yuta Koreeda, Christopher D. Manning. ContractNLI: A Dataset for Document-level Natural Language Inference for Contracts. arXiv:2110.01799, 2021.
- [S03] Bhuwan Dhingra, Jeremy R. Cole, Julian Martin Eisenschlos, Daniel Gillick, Jacob Eisenstein, William W. Cohen. Time-Aware Language Models as Temporal Knowledge Bases. arXiv:2106.15110, 2021.
- [S04] Wenhu Chen, Xinyi Wang, William Yang Wang. A Dataset for Answering Time-Sensitive Questions. arXiv:2108.06314, 2021.
- [S05] Michael J.Q. Zhang, Eunsol Choi. SituatedQA: Incorporating Extra-Linguistic Contexts into QA. arXiv:2109.06157, 2021.
- [S06] Jungo Kasai et al. RealTime QA: What's the Answer Right Now? arXiv:2207.13332, 2022.
- [S07] Neel Guha et al. LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models. arXiv:2308.11462, 2023.
- [S08] Qingyu Tan, Hwee Tou Ng, Lidong Bing. Towards Benchmarking and Improving the Temporal Reasoning Capability of Large Language Models. arXiv:2306.08952, 2023.
- [S09] Xuezhi Wang et al. Self-Consistency Improves Chain of Thought Reasoning in Language Models. arXiv:2203.11171, 2022.
- [S10] Nelson F. Liu et al. Lost in the Middle: How Language Models Use Long Contexts. arXiv:2307.03172, 2023.
- [S11] Richard Snodgrass, Ilsoo Ahn. A Taxonomy of Time Databases. Proceedings of the 1985 ACM SIGMOD International Conference on Management of Data, 1985. doi:10.1145/318898.318921.
- [S12] James F. Allen. Maintaining Knowledge about Temporal Intervals. Communications of the ACM 26(11), 1983. doi:10.1145/182.358434.
- [S13] Christian S. Jensen, Richard T. Snodgrass. Temporal Data Management. IEEE Transactions on Knowledge and Data Engineering 11(1), 1999. doi:10.1109/69.755613.
- [S14] Ben Laurie, Adam Langley, Emilia Kasper. Certificate Transparency. RFC 6962, June 2013. https://www.rfc-editor.org/rfc/rfc6962.