

Enceavra Research Round 39: When the Records Disagree: Conflicting Collateral Evidence, Evidentiary Lens Balancing, and Abstention Failure in Multi-Witness Contract Audits

Date: 2026-10-29

Research lab: Enceavra Research

Short code: CEV-39 (Conflicting and Equivocal Collateral Evidence)

Primary tracks: Track 1 (Evidence, Provenance, and Change), Track 5 (Calibration, Abstention, and Release Control)

Carry-forward boundary: Round 38 (CIV-38) showed that once every condition to effectiveness has one attached verification record, a gate that checks condition satisfaction, execution authority, and sub-clause scope releases 12 / 12 lifecycles correctly. It assumed that the attached record was unambiguous. Its Section 10 handed off two questions: (1) what a balanced two and two assignment of precautionary and presumptive evidentiary lenses does to a 3-of-4 quorum, and (2) what happens when two attached collateral records disagree (CEV-39). Round 39 addresses both on authored packets and pairs them with a track of 24 real executed exhibits filed with the SEC.

Abstract

Executed contract modifications often take effect only when an outside event occurs: an escrow deposit, a lender agent's consent, a countersignature by the right officer, a payment before a stated cutoff. A verification pipeline that attaches the record of that event can still fail when two attached records disagree. A treasury notice says a deposit was wired while the escrow statement says it was returned; a wire confirmation in UTC and a ledger in Singapore Time appear to straddle a New York cutoff; two letters from the same agent on the same date grant and withhold the same consent. Some of these disagreements are resolved by the contract itself, because it makes one record conclusive or requires an exact computation. Others are not, and the only correct output is to hold.

We formulate **Conflicting and Equivocal Collateral Evidence (CEV-39)**, a release rule that resolves a disagreement in a fixed order: the record the contract designates as conclusive, sole, or exclusive; then any computation the contract requires (named time zone cutoff, Business Day count, aggregate authority limit); then agreement among records; otherwise hold, unless the verdict is the same under every reading. We report three strictly separated evidence categories.

Deterministic (E39-A, no model calls). An exact enumeration of 112 evidence states (108 determinate, 4 requiring a hold; 56 with conflicting records) shows that trusting the first attached record, a presumptive lens, a precautionary lens, or a computation without time zone normalization each releases 24 / 112 states unsafely (21.43%), rising to 16 / 56 (28.57%) on conflict states. Honoring a designated record but nothing else leaves 20 / 112 unsafe. CEV-39 without the invariance rule releases none unsafely but holds 4 determinate states; with it, 0 unsafe releases and 0 false holds, by construction. An exact lens composition analysis under a 3-of-4 quorum shows that only a two and two split changes outcomes, and that it holds exactly the outcome relevant unresolved states only when

witnesses apply the contract's own evidence rules before their lens. For witnesses that resolve by lens alone, the same balance converts 24 determinate states into false holds.

Live authored (E39-B, 96 live evaluations). Twelve frozen packets (K01 to K12) were audited by Witness A (Max reasoning), Witness B, Witness C, and Witness D, each under both lenses. All 96 calls succeeded and parsed. Every witness under every lens resolved designated record conflicts (24 / 24), computation conflicts (24 / 24), consistent controls (24 / 24), and the outcome invariant conflict (8 / 8) correctly. **No witness returned HOLD in any of the 96 evaluations.** On K08, two same-date agent letters that grant and withhold one consent, all four presumptive runs returned PASS and all four precautionary runs returned FAIL. Verdicts depended on the lens only on the two packets authored as unresolved (K07, K08) and on none of the other ten. Under a 3-of-4 quorum, all 6 balanced lens assignments held K08 and all 10 unbalanced assignments released it. Witnesses also exposed an authoring defect in K07, reported in full in Section 7.5. The CEV-39 gate released 10 / 12 packets correctly, held K07 and K08, and made no unsafe release.

Live real exhibits (E39-C). Twenty four real amendment, consent, release, and merger exhibits from SEC EDGAR, each with its accession number, were audited for whether their principal operative effect can be verified from the exhibit alone. Witness C completed all 24 and matched the structural label on 22; Witness B completed 9 of a frozen 12 exhibit subset before the round's fixed evaluation budget closed and matched 9 / 9. One mismatch traces to an audit window that began after the exhibit's effectiveness clause. A deterministic pattern scan of all 109 fetched exhibits found effectiveness made conditional in 97, including 96 that are also dated or made effective "as of" a date.

Interpretation, kept separate. The live data support a narrow claim: these witnesses follow contract evidence rules reliably, detect conflict reliably, and do not abstain when the contract leaves a conflict unresolved. Lens balancing turns that abstention failure into a quorum deadlock on the one clean unresolved packet. It detects an unresolved conflict; it does not resolve one.

1. Archive continuity and anti-repetition verification

Before designing any Round 39 experiment we rebuilt the archive ledger from the live research registry (Rounds 3 and 8 through 38) and the Round 38 manuscript (prior-rounds-ledger.json, 32 entries). A keyword scan for the Round 39 constructs (conflicting collateral records, lens composition, two and two lens split, time zone cutoff, Business Day count, verdict invariance, designated or conclusive record) returned 0 hits in Rounds 1 to 38 outside the Round 38 Section 10 handoff.

Prior round	Focus	How Round 39 differs
CIV-38 (Round 38)	Conditional effectiveness, execution authority, sub-clause scope, with one unambiguous verification record per condition	Two attached records per condition disagree; the question is whether the contract resolves the disagreement and whether witnesses abstain when it does not
TDR-37 (Round 37)	Register closure and event-date resolution	Register closure and date applicability are held fixed; time enters only through contract-required cutoffs and Business Day counts
CCS-36 (Round 36)	Cross-document reference closure and precedence	Every packet is reference-closed; conflict is between collateral records, not between instruments

Prior round	Focus	How Round 39 differs
CIS-35 (Round 35)	Verbatim span grounding	Span grounding is retained as a measured field, not the object of study

The cumulative verification stack after Round 39 is: verbatim span grounding (CIS-35), reference closure (CCS-36), register closure and event-date applicability (TDR-37), operative validity of conditions, authority, and sub-clause scope (CIV-38), and resolution of conflicting collateral evidence with explicit abstention (CEV-39).

2. Executive summary

1. **Measured, deterministic.** Every single-record or single-lens policy releases at least 20 of 112 evidence states unsafely. The ordered CEV-39 resolver with verdict invariance releases 0 unsafely and holds only the 4 states that require a hold.
2. **Measured, deterministic.** In a 3-of-4 quorum, one dissenting lens never changes an outcome; only a two and two split does. A two and two split holds exactly the outcome relevant unresolved states if, and only if, witnesses apply contract evidence rules first.
3. **Measured, live.** All four witnesses, under both lenses, resolved every conflict that the contract resolves (48 / 48), including a UTC versus Singapore Time cutoff, a Business Day count across Memorial Day, and an aggregate signing limit.
4. **Measured, live.** No witness returned HOLD in 96 / 96 evaluations. On the clean unresolved packet K08, the verdict was fully determined by the assigned lens (8 / 8).
5. **Measured, live.** Balanced lens assignments held K08 in 6 / 6 cases; unbalanced assignments released a determinate and unsupported verdict in 10 / 10. The Round 38 panel (one precautionary witness) released K08 as PASS.
6. **Measured, live, real text.** On 24 real SEC exhibits, the two witnesses that ran agreed with the structural label on 31 of 33 completed evaluations. Almost every amendment in the fetched pool states a date "as of" which it is effective while making effectiveness conditional on deliveries that the exhibit cannot prove.
7. **Interpretation.** Abstention, not conflict detection and not rule application, is the failure mode this round isolates. Lens balancing is a conflict detector, not a conflict resolver; the resolver has to be deterministic and contract-driven.

3. Problem statement and formal model (CEV-39)

A packet contains a base instrument, a modification whose effectiveness depends on a condition c , and two collateral records REC-1 and REC-2, each asserting c MET or UNMET. The audit question has a base verdict v_{base} (modification inoperative) and a modified verdict v_{mod} (modification operative). The disagreement is outcome relevant when v_{base} differs from v_{mod} .

3.1 The ordered CEV-39 resolver

1. **Designated record.** If the contract makes one record conclusive, sole, or exclusive evidence of c , that record decides c .
2. **Required computation.** Otherwise, if the contract states c as a computable test (a deadline in a named time zone, a count of Business Days, an aggregate limit), the exact computation from the attached facts decides c , whatever either record concludes.
3. **Agreement.** Otherwise, if the records agree, their shared value decides c .
4. **Invariance or hold.** Otherwise c is UNDETERMINED. The gate releases v_{base} if v_{base} equals v_{mod} , and holds otherwise.

The gate then requires a 3-of-4 witness quorum whose verdict equals the resolver verdict, with at least three agreeing witnesses quoting verbatim from the instrument they cite.

3.2 Evidentiary lenses

A **presumptive** lens treats the modification as operative unless the packet establishes that its condition failed. A **precautionary** lens treats it as operative only if the packet establishes that every condition was satisfied. Both are resolution rules for the residual case in step 4. Neither is a rule for abstention.

3.3 Structural propositions

1. A single lens, applied uniformly, releases a determinate verdict on every outcome relevant unresolved state, so a quorum of uniformly lensed witnesses cannot hold such a state.
2. Under a 3-of-4 quorum, a 3-1 or 1-3 lens split releases the majority lens's verdict; only a 2-2 split can deadlock.
3. A 2-2 split deadlocks exactly on outcome relevant unresolved states if witnesses apply steps 1 to 3 before their lens; if they apply their lens to every conflict, it also deadlocks on conflicts the contract resolves.

4. Related work

Contract understanding benchmarks such as CUAD [S01], ContractNLI [S02], MAUD [S18], and LegalBench [S07], and retrieval benchmarks such as LegalBench-RAG [S19], evaluate extraction and reasoning over a single document or retrieved passages. They do not ask whether a modification is operative when collateral records disagree. Studies of legal hallucination [S20, S21] document unsupported determinate answers in legal settings; our zero abstention result is a narrow, contract-specific instance. Work on knowledge conflicts [S22] and on the tension between model priors and supplied evidence [S23] studies conflicts between context and parametric knowledge; here both conflicting sources are in context and the contract defines which one wins. Abstention research [S24, S25] motivates treating HOLD as a first-class output. Multi-model juries [S26], LLM judges [S27], self-consistency [S09], and multiagent debate [S17] aggregate model outputs; our lens analysis shows that the aggregation rule interacts with the instruction each member receives. Classical work on aggregating

disagreeing sources, from observer error-rate models [S28] to truth discovery [S29], estimates source reliability; contracts often settle reliability by designation, which is what step 1 of the resolver encodes. Temporal database work [S11, S12, S13] and computable contract models [S15, S16] inform the treatment of cutoffs and conditions.

5. Hypotheses

ID	Hypothesis	Evidence	Status
H39.1	Single-record and single-lens gates release unsafely on conflict states; ordered CEV-39 resolution with verdict invariance eliminates unsafe releases without false holds	E39-A (exact, specification check)	Supported
H39.2	Under a 3-of-4 quorum only a 2-2 lens split changes outcomes, and it holds exactly the outcome relevant unresolved states only for witnesses that apply contract evidence rules first	E39-A (exact)	Supported
H39.3	Witnesses apply a designated record and a contract-required computation regardless of assigned lens	E39-B, 48 / 48	Supported
H39.4	Witnesses return HOLD when records of equal standing conflict, the contract is silent, and the outcome depends on the conflict	E39-B, 0 / 16	Not supported
H39.5	A balanced two and two lens assignment holds the clean unresolved packet while every unbalanced assignment releases it	E39-B, K08, 6 / 6 versus 0 / 10	Supported
H39.6	The CEV-39 gate makes no unsafe release on the authored suite	E39-B, 0 / 12	Supported
H39.7	On real filed exhibits, witnesses distinguish conditional from unconditional effectiveness from the exhibit text	E39-C, 31 / 33 completed evaluations	Supported

H39.6 uses gate-side evidence precedence metadata authored with the packets, so it is a specification check on authored data, not independent evidence of field performance.

6. Evidence category 1: deterministic structural diagnostic (E39-A)

6.1 Finite universe and gates

Each state fixes a resolution channel (contract silent; REC-1 designated; REC-2 designated; or a required computation whose exact result is MET or UNMET and whose unnormalized reading agrees or flips), the claims of REC-1 and REC-2, outcome relevance, and the base verdict: $7 \times 4 \times 2 \times 2 = 112$ states. Truth follows Section 3.1. All four claim pairs are enumerated within every channel.

Gate	Correct releases	Wrong-verdict releases	Releases where hold required	Correct holds	False holds	Unsafe release rate	False hold rate
G0 first attached record trusted	88	20	4	0	0	21.43%	0.00%
G1 presumptive lens	88	20	4	0	0	21.43%	0.00%

Gate	Correct releases	Wrong-verdict releases	Releases where hold required	Correct holds	False holds	Unsafe release rate	False hold rate
G2 precautionary lens	88	20	4	0	0	21.43%	0.00%
G3 computation without normalization	88	20	4	0	0	21.43%	0.00%
G4 designated record only	92	16	4	0	0	17.86%	0.00%
G5 CEV-39 strict	104	0	0	4	4	0.00%	3.70%
G6 CEV-39 with verdict invariance	108	0	0	4	0	0.00%	0.00%

Rates are over all 112 states (unsafe) and over the 108 determinate states (false holds). G0 to G3 coincide in total by symmetry of the enumeration; they fail on different states.

6.2 Conflict states only, by channel

On the 56 states in which the records disagree, G0 to G3 release 16 unsafely (28.57%), G4 releases 12 (21.43%), G5 holds 4 determinate invariant states (7.69% of 52), and G6 makes no error. By channel: on 16 designated record conflicts, G0 to G3 each give 4 wrong verdicts and G4 to G6 none; on 32 computation conflicts, G0 to G4 each give 8 wrong verdicts and G5 and G6 none; on 8 silent conflicts, G0 to G4 release all 4 outcome relevant states, G5 holds all 8, and G6 holds the 4 relevant states and releases the 4 invariant ones.

6.3 Exact lens composition under a 3-of-4 quorum

k is the number of precautionary witnesses among four.

Witness model	k	Correct releases	Unsafe releases	Correct holds	False holds
lens only	0, 1, 3, 4	88	24	0	0
lens only	2	76	8	4	24
rule following	0, 1, 3, 4	108	4	0	0
rule following	2	108	0	4	0

For lens-only witnesses, the 8 unsafe releases at k = 2 are states where both records agree with each other but not with the computation the contract requires, so balancing cannot catch them. For rule-following witnesses, the 2-2 split is an exact detector of outcome relevant unresolved conflict.

6.4 Resolver check on the authored packets

The deterministic CEV-39 resolver reproduces the authored expected verdict on 12 / 12 packets. This confirms the packets encode the intended structure; it is not model evidence.

7. Evidence category 2: live multi-witness evaluation (E39-B)

7.1 Witness roster and protocol

Witness	Audit role	Reasoning tier
Witness A	Lead Operative-Validity Auditor	Max reasoning
Witness B	Evidentiary Standards Auditor	Standard
Witness C	Records and Computation Auditor	Standard
Witness D	Formalities and Authority Auditor	Standard

Each witness audited every packet twice, once under each lens, at temperature 0, for $12 \times 4 \times 2 = 96$ evaluations. All received the same rule: apply the contract's own evidence and computation provisions first, use the lens only where the contract does not resolve a disagreement, and return HOLD only if no determinate verdict is supportable. Each returned a verdict (PASS, FAIL, HOLD), whether the modification is operative, whether a conflict was detected, the governing record (REC-1, REC-2, COMPUTED, NONE), the controlling instrument, a verbatim quote, and a rationale. Verbatim prompts and responses are in the reproducibility archive.

7.2 The twelve frozen packets

Packet	Stratum	Conflict	Contract resolution	Expected
K01	designated record	Treasury notice says reserve wired; escrow statement says wire returned	Escrow statement conclusive	FAIL (cap stays 1x fees; 2.4x claim exceeds it)
K02	designated record	Internal memo says agent consent pending; agent register says granted	Consent register sole evidence	PASS (30 day notice valid)
K03	designated record	Countersignature by J. Ruiz as Procurement Director; register shows A. Patel held the role from January 15	Signatory Register exclusive	FAIL (60 day notice still required)
K04	computation	Wire credited 21:20 UTC June 30; ledger posted 05:20 Singapore Time July 1	5:00 p.m. New York time cutoff	FAIL (17:20 EDT is late; credit stays 10%)
K05	computation	Customer counts payment on Business Day 10; Supplier counts Business Day 11	Business Day excludes New York bank holidays	FAIL (Memorial Day excluded; payment timely; termination waived)
K06	computation	Officer certificate says EUR 320,000 within EUR 500,000 limit; schedule shows linked order of EUR 260,000	Limit is aggregate including linked orders	FAIL (EUR 580,000 exceeds limit; extension inoperative)
K07	unresolved, relevant	Two Approved Auditors certify 118 and 124 against 120	Contract silent	HOLD (as authored; see 7.5)
K08	unresolved, relevant	Two same-date Facility Agent letters, one consenting, one withholding	Contract silent	HOLD
K09	unresolved, invariant	Board minutes say approved; secretary certificate says tabled	Contract silent	PASS (20 days meets both 45 and 30 day terms)
K10	consistent	Both records show reserve deposited	Not needed	PASS
K11	consistent	Both records show consent refused	Not needed	FAIL

Packet	Stratum	Conflict	Contract resolution	Expected
K12	consistent	Both records show fee received before cutoff	Not needed	PASS

7.3 Results by stratum

Stratum	Evaluations	Correct verdicts	HOLD returned	Conflict flagged	Governing record as expected
Designated record (K01 to K03)	24	24	0	24	24
Computation (K04 to K06)	24	24	0	17	23
Unresolved, relevant (K07 , K08)	16	0	0	16	10
Unresolved, invariant (K09)	8	8	0	8	4
Consistent (K10 to K12)	24	24	0	0	17

Only 1 of 8 evaluations flagged a conflict on K04: witnesses converted both timestamps to the same instant and treated the disagreement as apparent. Each witness scored 10 / 12 under each lens. Quotes were verified verbatim somewhere in the packet in 96 / 96 evaluations and inside the cited instrument in 85 / 96. Governing record matches on consistent packets are not scored as errors, since either record may be cited.

7.4 Lens dependence and governance arms

A witness's verdict differed between its two lenses on K07 (3 of 4 witnesses) and K08 (4 of 4), and on none of the other 10 packets (0 of 40 witness and packet pairs). Self-reported conflict flags fired on 49 of 56 evaluations of conflicts the contract resolves or that do not affect the outcome, so the flag alone does not separate resolvable from unresolvable conflict; lens dependence did.

Arm	Correct releases	Unsafe releases	Correct holds	False holds	Unsafe releases, K07 excluded
A1_witnessA_presumptive	10 / 12	2 / 12	0 / 12	0 / 12	1 / 11
A2_witnessA_precautionary	10 / 12	2 / 12	0 / 12	0 / 12	1 / 11
A3_round38_lens_panel	10 / 12	1 / 12	1 / 12	0 / 12	1 / 11
A4_balanced_two_two_panel	10 / 12	1 / 12	1 / 12	0 / 12	0 / 11
A5_all_precautionary	10 / 12	2 / 12	0 / 12	0 / 12	1 / 11
A6_all_presumptive	10 / 12	1 / 12	1 / 12	0 / 12	1 / 11
A7_civ38_gate	10 / 12	1 / 12	1 / 12	0 / 12	1 / 11
A8_cev39_gate	10 / 12	0 / 12	2 / 12	0 / 12	0 / 11

A3 assigns the precautionary lens to Witness B only, as in Round 38. A4 assigns it to Witness A and Witness B. A3 to A7 use a 3-of-4 quorum; A7 adds the Round 38 span requirement; A8 adds the CEV-

39 resolver. Over all 16 possible lens assignments, K08 was held in 6 / 6 balanced assignments and released in 10 / 10 unbalanced ones; K07 was held in 6 / 16, and both were held in 2 / 16.

7.5 Authoring defect found by witnesses in K07

K07 was frozen with expected verdict HOLD. Witness rationales identified two defects. First, Section 8.4(b) of its base agreement provides that Supplier invoices quarterly in arrears, so a September 15 invoice for the third quarter is premature whether or not the modification is operative; the verdict is FAIL under either reading. Second, the condition requires certification "by an Approved Auditor", which one certificate can satisfy, so the contract arguably resolves the conflict. Witness A under the precautionary lens returned PASS on that second reading; Witness A and Witness B under the presumptive lens returned FAIL on the first. We keep the pre-registered label in every table and report K07 excluded results alongside. K08 is the only clean outcome relevant unresolved packet in this round, and the zero HOLD result rests on it and on K07 together.

7.6 Failure walk-through: K08

Amendment No. 1 lets Customer assign to an affiliate without consent, but only upon the Facility Agent's written consent. Two letters from the agent, both dated September 1, 2026, are attached. One reads "We consent to the amendment". The other reads "Our consent to Amendment No. 1 is withheld pending completion of know-your-customer review". The contract says nothing about which letter governs. All eight evaluations flagged the conflict. None returned HOLD. The four presumptive runs certified the assignment as permitted; the four precautionary runs certified it as not permitted. The Round 38 panel released PASS by 3-1. The CEV-39 resolver marks the condition UNDETERMINED, finds the outcome depends on it, and holds.

8. Evidence category 3: real SEC EDGAR exhibit track (E39-C)

8.1 Construction

We searched EDGAR full-text search for amendment, consent, and merger agreement exhibits filed from 2025 onward, fetched 109 exhibits, and froze a stratified track of 24 before any model call: 14 amendments whose effectiveness is conditioned on deliveries or other events (X1), 2 with partitioned effectiveness (X2), 4 merger agreements whose consummation is conditioned (X3), and 4 instruments effective on their face (X4), chosen after reading the full text. Each witness received an audit window (the exhibit opening plus the passage governing effectiveness) and was asked whether the principal operative effect can be verified from the exhibit alone. The structural label is authored from the exhibit text; it is not a legal opinion on whether the conditions were later met.

8.2 Track and results

ID	Stratum	Filer	Exhibit	Filed	Accession	Structural label	Witness C	Witness B
X01	X1	Waste Connections, Inc.	EX-10.1	2025-05-27	0001104659-25-052526	CONDITIONAL	CONDITIONAL	CONDITIONAL

ID	Stratum	Filer	Exhibit	Filed	Accession	Structural label	Witness C	Witness B
X02	X1	Ameresco, Inc.	EX-10.1	2026-03-31	0001628280-26-022433	CONDITIONAL	CONDITIONAL	not run
X03	X1	Sonendo, Inc.	EX-10.1	2025-03-04	0000950170-25-031237	CONDITIONAL	CONDITIONAL	not run
X04	X1	Hamilton Insurance Group, Ltd.	EX-10.2	2025-10-24	0001593275-25-000251	CONDITIONAL	CONDITIONAL	not run
X05	X1	Spire Global, Inc.	EX-10.1	2025-03-13	0000950170-25-038328	CONDITIONAL	CONDITIONAL	CONDITIONAL
X06	X1	Arq, Inc.	EX-10.1	2026-03-03	0001515156-26-000008	CONDITIONAL	CONDITIONAL	not run
X07	X1	Xtant Medical Holdings, Inc.	EX-10.3	2025-04-11	0001641172-25-003711	CONDITIONAL	CONDITIONAL	not run
X08	X1	GXO Logistics, Inc.	EX-10.2	2025-11-24	0001104659-25-115447	CONDITIONAL	OPERATIVE	not run
X09	X1	Super Micro Computer, Inc.	EX-10.1	2026-01-29	0001375365-26-000005	CONDITIONAL	CONDITIONAL	CONDITIONAL
X10	X1	Lucid Group, Inc.	EX-10.1	2025-11-05	0001628280-25-049553	CONDITIONAL	CONDITIONAL	not run
X11	X1	CONDUENT Inc	EX-10.1	2026-10-15	0001677703-26-000134	CONDITIONAL	CONDITIONAL	not run
X12	X1	EnerSys	EX-10.1	2025-12-15	0001289308-25-000046	CONDITIONAL	CONDITIONAL	not run
X13	X1	Asana, Inc.	EX-10.1	2026-03-02	0001477720-26-000012	CONDITIONAL	CONDITIONAL	CONDITIONAL
X14	X1	AMERICAN VANGUARD CORP	EX-10.1	2025-05-08	0001193125-25-115977	CONDITIONAL	CONDITIONAL	not run
X15	X2	MANHATTAN BRIDGE CAPITAL, INC	EX-10.5	2025-12-16	0001493152-25-027893	PARTIAL	PARTIAL	PARTIAL
X16	X2	Local Bounti Corporation/DE	EX-10.2	2025-08-04	0001840780-25-000031	PARTIAL	CONDITIONAL	PARTIAL
X17	X3	e.l.f. Beauty, Inc.	EX-2.1	2025-08-06	0001600033-25-000037	CONDITIONAL	CONDITIONAL	CONDITIONAL
X18	X3	TaskUs, Inc.	EX-2.1	2025-05-09	0000950157-25-000390	CONDITIONAL	CONDITIONAL	not run
X19	X3	Criteo S.A.	EX-2.1	2026-08-05	0001628280-26-053414	CONDITIONAL	CONDITIONAL	CONDITIONAL
X20	X3	First Seacoast Bancorp, Inc.	EX-2.1	2026-05-05	0001193125-26-206551	CONDITIONAL	CONDITIONAL	not run
X21	X4	BioXcel Therapeutics, Inc.	EX-10.1	2026-08-11	0001104659-26-093710	OPERATIVE	OPERATIVE	OPERATIVE

ID	Stratum	Filer	Exhibit	Filed	Accession	Structural label	Witness C	Witness B
X22	X4	Apollo Debt Solutions BDC	EX-10.1	2025-04-14	0001193125-25-080186	OPERATIVE	OPERATIVE	not run
X23	X4	Medalist Diversified REIT, Inc.	EX-10.3	2026-02-17	0001104659-26-016379	OPERATIVE	OPERATIVE	not run
X24	X4	MAMMOTH ENERGY SERVICES, INC.	EX-10.1	2025-07-03	0001628280-25-034133	OPERATIVE	OPERATIVE	not run

Witness C matched the structural label on 22 / 24, claimed operative on its face for 1 of 20 conditioned exhibits, labeled all 4 unconditional controls correctly, and quoted the window verbatim in 21 / 24. Witness B was run on a frozen subset of 12 (X01, X05, X09, X13, X15, X16, X17, X19, X21 to X24); 9 completed before the round's fixed evaluation budget closed, all 9 matched, and 6 / 9 quotes verified. The two witnesses agreed on 8 of 9 shared exhibits. Witness A and Witness D were not run on this track, so it measures single-witness recognition on real text, not quorum behavior.

8.3 Mismatches

X08 (GXO Logistics, Amendment No. 1 to Credit Agreement): the effectiveness clause ("shall become effective on the date on which the Administrative Agent (or its counsel) shall have received from the Borrower and each of the Required Lenders" counterparts) lies before the start of the frozen audit window. Witness C judged the window it received correctly; the error is in our window construction. X16 (Local Bounti, twelfth amendment): Witness C labeled the whole instrument conditional while Witness B and the structural label treat it as partitioned; the instrument conditions Sections 3 and 4 on receipt of listed deliveries on or before its date, and the boundary between the two labels is arguable.

8.4 Deterministic prevalence scan

Over all 109 fetched exhibits, pattern matches found language dating the instrument or its effect "as of" a date in 108, conditional effectiveness language in 97, both in 96, a receipt condition for executed counterparts or signature pages in 65, an officer, secretary, or closing certificate in 12, and payment of fees or expenses in 12. The pool was assembled by searches that targeted conditional language, so these counts describe this pool, not EDGAR as a whole. The structural point stands for this pool: a stated "as of" date is not evidence that an amendment is operative.

9. Findings, interpretations, and critical synthesis review

9.1 Measured findings

1. Exact enumeration: single-record and single-lens policies release 20 to 24 of 112 states unsafely; CEV-39 with invariance releases 0 and holds only the 4 required states.
2. Exact enumeration: only a 2-2 lens split changes a 3-of-4 quorum outcome; it is an exact detector of outcome relevant unresolved conflict for rule-following witnesses.

3. Live: 48 / 48 correct on conflicts the contract resolves, under both lenses, across all four witnesses.
4. Live: 0 / 96 HOLD; on K08 the verdict equaled the lens in 8 / 8.
5. Live: lens dependence occurred on K07 and K08 only; balanced assignments held K08 in 6 / 6.
6. Live, real text: 31 / 33 completed evaluations matched the structural label; the two mismatches are explained in Section 8.3.

9.2 Interpretation

The witnesses in this round behave like the rule-following model of Section 6.3 on every packet the contract resolves. Their failure is confined to the residual case, where they treat the lens as a license to decide rather than as a default to flag. That makes a balanced lens panel useful as a cheap detector: if two precautionary and two presumptive witnesses split, the conflict is one the contract does not resolve. It does not make the panel a resolver, and an unbalanced panel actively hides the problem behind a 3-1 majority. The release decision should come from the deterministic resolver; the lens panel can serve as a cross-check on whether the resolver's UNDETERMINED call is shared by the witnesses.

9.3 Critical synthesis review (Witness A Max reasoning, interpretation input only)

After all measurements were recorded, Witness A reviewed the results as a skeptical external reviewer. Its main points, quoted or closely paraphrased from the verbatim review in the archive: (1) the zero HOLD rate is plausibly an artifact of the lens design, because both lenses are resolution rules and give a witness no reason to abstain; (2) the gate and the strata were authored by the same program, so the gate's 10 / 12 is close to circular, and K07 shows authored labels need independent audit; (3) on K08 the conflict flag fired in every run, so the split added little beyond the flag on that packet. We accept (1) and (2) as limitations. On (3), the flag also fired on 49 / 56 evaluations of resolvable or invariant conflicts, whereas lens dependence occurred only on the unresolved packets, so the split does carry information the flag lacks. The reviewer's proposed next check is adopted in Section 11.

10. Limitations and threats to validity

1. **Small authored suite.** Only two packets are outcome relevant and unresolved, and one is defective. The zero HOLD result is a strong signal on these packets, not a population estimate.
2. **Lens instructions bias against abstention.** Both lenses tell the witness how to resolve the residual case. A neutral instruction that offers HOLD without a lens has not been tested here.
3. **Authored labels and gate metadata come from the same team.** The CEV-39 gate reads precedence metadata written with the packets. Its live score is a specification check.
4. **Partial real-text coverage.** E39-C ran two witnesses, one on all 24 exhibits and one on 9 completed of 12. No quorum was formed on real text. One audit window was mis-cut.
5. **Search-biased pool.** The prevalence counts come from a pool built by targeted searches.
6. **Single run at temperature 0.** Run-to-run variation was not measured.

7. **Witness independence.** The four witnesses may share training data or behavior, so a 3-of-4 agreement is not four independent votes.
-

11. Next check (Round 40 handoff)

1. **Neutral abstention test.** Re-run corrected K07, K08, and at least six new outcome relevant unresolved packets with a neutral instruction that offers HOLD and no lens, with expected verdicts audited by a reviewer outside the authoring process before any model call. If witnesses hold without a lens, the lens apparatus is unnecessary; if they still decide, abstention is the failure the verification stack must absorb deterministically.
 2. **Real-text quorum.** Complete the E39-C track with all four witnesses on corrected audit windows, and add a paired evidence step: for each conditioned exhibit, locate the filer's own later statement (for example a Form 8-K item or periodic report) that the amendment became effective, and test whether witnesses can tie that statement to the exhibit's conditions.
-

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.
- [S07] Neel Guha et al. LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models. arXiv:2308.11462, 2023.
- [S09] Xuezhi Wang et al. Self-Consistency Improves Chain of Thought Reasoning in Language Models. arXiv:2203.11171, 2022.
- [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.
- [S15] Guido Governatori, Antonino Rotolo. Logic and the Law: Philosophical Foundations, Deontics, and Defeasible Reasoning. Handbook of Deontic Logic and Normative Systems, 2013.
- [S16] Mark D. Flood, Oliver R. Goodenough. Contract as Automaton: Representing a Simple Financial Agreement in Computational Form. Artificial Intelligence and Law 30, 2022. doi:10.1007/s10506-021-09300-9.

- [S17] Yilun Du, Shuang Li, Antonio Torralba, Joshua B. Tenenbaum, Igor Mordatch. Improving Factuality and Reasoning in Language Models through Multiagent Debate. arXiv:2305.14325, 2023.
- [S18] Steven H. Wang, Antoine Scardigli, et al. MAUD: An Expert-Annotated Legal NLP Dataset for Merger Agreement Understanding. arXiv:2301.00876, 2023.
- [S19] Nicholas Pipitone, Ghita Hour Alami. LegalBench-RAG: A Benchmark for Retrieval-Augmented Generation in the Legal Domain. arXiv:2408.10343, 2024.
- [S20] Varun Magesh, Faiz Surani, et al. Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools. arXiv:2405.20362, 2024.
- [S21] Matthew Dahl, Varun Magesh, et al. Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models. arXiv:2401.01301, 2024.
- [S22] Rongwu Xu, Zehan Qi, et al. Knowledge Conflicts for LLMs: A Survey. arXiv:2403.08319, 2024.
- [S23] Kevin Wu, Eric Wu, James Zou. ClashEval: Quantifying the Tug-of-War between an LLM's Internal Prior and External Evidence. arXiv:2404.10198, 2024.
- [S24] Bingbing Wen, Jihan Yao, Shangbin Feng, et al. Know Your Limits: A Survey of Abstention in Large Language Models. arXiv:2407.18418, 2024.
- [S25] Pranav Rajpurkar, Robin Jia, Percy Liang. Know What You Don't Know: Unanswerable Questions for SQuAD. arXiv:1806.03822, 2018.
- [S26] Pat Verga, Sebastian Hofstatter, et al. Replacing Judges with Juries: Evaluating LLM Generations with a Panel of Diverse Models. arXiv:2404.18796, 2024.
- [S27] Lianmin Zheng, Wei-Lin Chiang, et al. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. arXiv:2306.05685, 2023.
- [S28] A. P. Dawid, A. M. Skene. Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm. Applied Statistics 28(1), 1979. doi:10.2307/2346806.
- [S29] Yaliang Li, Jing Gao, Chuishi Meng, et al. A Survey on Truth Discovery. ACM SIGKDD Explorations Newsletter 17(2), 2016. doi:10.1145/2897350.2897352.
- [S30] U.S. Securities and Exchange Commission. EDGAR Full-Text Search and Archives. <https://efits.sec.gov/LATEST/search-index> and <https://www.sec.gov/Archives/edgar/data/>.