

## RLS-RFC-0001: The ripple.md Standard

| Integrity surface         | Current requirement                                                                                                                                                  | Claim boundary                                                          |
|---------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------|
| Critical-path tempo       | Consequence-tempo compatibility is evaluated across the control-critical path and may be concurrent; the standard does not require a simplistic sequential pipeline. | Tempo assurance does not prove physical safety or deployment readiness. |
| Responsibility continuity | Named owner, monitor, intervener, appeal recipient, and post-incident custodian roles remain assigned across the consequence horizon.                                | A wrapper record cannot create legal or institutional authority.        |
| Wrapper boundary          | Non-Domination, Epistemic Integrity, assurance, waiver, appeal, and reconstructability surround the Canon run.                                                       | No wrapper obligation becomes a sixth Canon level.                      |

### Qualification-Continuity and Configuration-Assurance Extension

The wrapper record now includes `configuration_id`, `configuration_hash_if_available`, `operating_envelope`, `active_tools_permissions`, `active_controls`, `authority_state`, `qualification_valid_from`, `qualification_expiry_or_review`, `requalification_triggers`, `trigger_events`, and `affected_stages_reopened`.

A prior PASS, selectability result, or execution approval remains usable only while the declared configuration, evidence, controls, authority, dependencies, and operating conditions remain materially unchanged. Material change pauses inherited execution authority and routes the episode back to the affected Canon stages. Monitoring without named stop authority and requalification logic is insufficient.

**Wrapper handoff rule.** In wrapped deployments, ripple.md adds assurance, Non-Domination, Epistemic Integrity, waiver, triage, and reconstructability obligations around the RippleLogic run. It does not reorder, weaken, or replace the Canon cascade. The Canon controls cascade semantics and admissibility.

**Canon/wrapper non-collapse rule.** ripple.md wrapper checks do not become Canon cascade levels. Canon qualification remains RG -> RF/NCRC -> TRC -> CSV, followed by RLS ranking. Non-Domination and Epistemic Integrity are additive wrapper obligations that may block execution or claims, but they do not reorder the Canon.

| Layer                           | Controlling sequence or role                                                                                                                                                                                                                                                    |
|---------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Canon qualification and ranking | RG -> RF/NCRC -> TRC -> CSV -> RLS                                                                                                                                                                                                                                              |
| ripple.md wrapper obligations   | Non-Domination, Epistemic Integrity, consequence-tempo compatibility, responsibility continuity, Decision Note completeness, assurance, waiver/triage, appeal, and reconstructability surround the run and may block execution or claims; they are not additional Canon levels. |

Series: RLS-RFC (RippleLogic Standard RFC)

RFC: 0001

Title: ripple.md - Portable Alignment Contract & Protocol for Sociotechnical Systems

Status: Research and teaching specification

Intended status: Proposed standard; see Appendix RELEASE

Authors: RippleLogic / MathGov Working Group (initial: James McGaughran)

License: Apache-2.0 (specification text and tooling code)

Publication source: this package's Appendix RELEASE and manifest identify the delivered master; the upstream repository identifier is retained in the release appendix.

Two-phase record rule. A ripple.md Decision Note SHOULD separate the qualification record (RG, RF/NCRC, TRC, CSV) from the ranking record (RLS over surviving options). The Decision Note MUST NOT present an RLS ranking as if it applied to options that failed qualification.

## Current bundle dependency boundary (Normative release boundary)

This core release is self-contained for L0/L1 wrapper learning and basic Decision Note practice. L2/L3 conformance paths that require recourse auditability, redaction/access control, or coercion authorization also require the applicable companion protocols, including RPAP (RLS-RFC-0003) and, where coercion or emergency projection is in scope, PFAP (RLS-RFC-0002). Those companions are referenced but not bundled as active protocols in this release. Therefore, do not claim L2/L3 ripple.md conformance from this package alone unless the relevant RPAP/PFAP artifacts are separately bundled, hash-pinned, and recorded in the Assurance Preconditions Record.

## ABSTRACT

Machine-readable RF field discipline. Decision Notes and schemas SHOULD use the human-readable field name `rights_floor`, while preserving `formal_constraint`: NCRC inside the record when formal gate logic is required.

Example:

`rights_floor`:

`status`: pass | fail | uncertain | not\_applicable

`formal_constraint`: NCRC

`protected_rights_checked`: []

`rights_floor_breaches`: []

`uncertainty_notes`: []

This RFC specifies ripple.md, a portable, auditable alignment contract and decision protocol designed to bind decision-making across scales, from governance institutions to organizations to tool-using agents. This standard is operational, not merely declarative: it requires a Decision Note pre-execution for high-stakes episodes, defines evidence sufficiency and sampling, mandates reconstructability audits, specifies conformance metrics with minimum tests and drills, and provides machine-readable registry identifiers for interoperable tooling. It carries forward Reality Grounding, structural diagnostic, Term Discrimination, Structural Viability, and falsification-discipline fields including the two retained bounded wrapper disciplines: consequence-tempo compatibility and responsibility continuity. High-stakes execution must not outrun meaningful detection, review, intervention, containment, or rollback; and distributed work must not dissolve end-to-end answerability. These disciplines add no Canon gate and do not authorize delay where delay creates the greater rights or tail-risk harm.

The standard defines: a wrapper feasibility suite over Rights Floor (formal NCRC), Non-Domination, TRC, Containment/Reversibility, and Epistemic Integrity, with cross-cutting consequence-tempo and responsibility-continuity requirements. In standalone ripple.md use, these checks MAY be executed in the documented wrapper order. In wrapped RippleLogic deployments, the underlying canonical RippleLogic method remains Reality Grounding → Rights Floor → TRC → CSV → RLS, while Non-Domination and Epistemic Integrity

remain additive wrapper obligations that MUST be evidenced before selection; a seven-test alignment suite with evidence expectations and explicit falsifiers; a minimum audit artifact (Decision Note) and supporting templates; conformance levels (L0–L3) with required observable metrics, minimum test suites, and drill requirements; an audit procedure grounded in assurance-case semantics (S,T,K,M,V,E,R,U) with inter-rater reliability and reconstructability audits; a formal mapping between the seven-test suite and the C1–C6 deployment assurance framework; a jurisdiction-neutral rights baseline with localization procedure and operationalization notes; and a PLSS-compatible local-scope profile for implementations that must keep local decisions practical without sacrificing rights, floors, or auditability.

RippleLogic RG status rule. Reality Grounding is Level 1 of the public method and a claim-authority precondition. It can force narrowing, escalation, exploratory-marking, or refusal, but it is not an ethical option-rejecting gate: an RG\_REFUSED option is outside the claim-bounded qualified set because the requested claim lacks support at the requested strength, not because RG has found the option ethically impermissible. The option may re-enter only after grounding is repaired or the claim or action boundary is narrowed. RF/NCRC, TRC, and CSV are the ethical or structural option-rejecting gates, and RLS ranks only what survives them.

Thresholds and domain-specific parameters referenced throughout this standard are implementation-defined (see §2.18 Implementation-Defined Parameters) but MUST be set ex ante in the verification regime (V) and recorded in the Assurance Preconditions Record (Appendix G). Non-normative example threshold ranges are provided in Appendix K.

#### Plain-language summary

ripple.md is the portable wrapper for putting RippleLogic-style assurance into real systems. It does not replace the Canon. It tells implementers what evidence, decision notes, falsifiers, timing compatibility, responsibility chains, audit records, and refusal conditions must exist before stronger alignment or conformance claims are made.

#### Reader Guide (Quick Adoption Path)

- Start with Sections 3–8 if you need the contract model, decision ordering, and test semantics.
- Use Section 12 and Appendix G to pin implementation-defined parameters and dependency versions before deployment.
- Use ripple.md to wrap, evidence, and audit a decision process; it does not replace the governing RippleLogic equations or the rights-localization step.
- In standalone ripple.md use, wrapper gates MAY be executed in the documented wrapper order. In wrapped RippleLogic use, preserve the canonical RippleLogic five-level method and formal ordering and treat Non-Domination and Epistemic Integrity as additive wrapper obligations.
- For local-scope implementations, preserve the canon's admissibility gates first and treat PLSS or agent profiles as wrapper-compatible only when the underlying run remains lexicographically intact.

Release-boundary reader note (Normative clarification).

Formal wrapped-method shorthand: RG/RSG -> RF/NCRC -> TRC -> CSV -> RLS. UCI/HOI are structural diagnostics, not a regular wrapper gate: gate-relevant use belongs inside Containment and residual use is special-case only.

ripple.md improves auditability, wrapper assurance, Decision Note discipline, waiver/triage governance, and reconstructability. It does not by itself validate the empirical quality of underlying measures, prove alignment, certify legal compliance, or make Tier 4/ProofPack claims available.

A Decision Note, PCC, workbook replay, and wrapper record together remain ProofPack-prep unless schemas, reference calculator, machine-readable test vectors, semantic validator, replay instructions, and independent replay evidence are bundled or hash-pinned for the claimed release.

## Table of Contents

0. Normative Keywords
1. Scope, Non-Goals, and Design Principles
2. Core Definitions
3. Artifact Role Map (What Each Artifact Does)
4. Rights Baseline (Jurisdiction-Neutral Default)
5. The ripple.md Contract Model
6. Five-Level Method and Constraint Ordering (Normative Decision Ordering)
7. Alignment Test Suite (7 Tests) with Falsifiers
8. General Decision Rule (Pass / Partial / Fail / NA / IND)
9. Evidence Sufficiency and Sampling Protocol
10. Assurance-Case Semantics: Mapping 7 Tests to C1-C6
11. Required Artifacts
12. Conformance Levels (L0-L3) with Minimum Observable Metrics
13. Audit Procedure, Reconstructability, and Inter-Rater Reliability
14. NCAR: Continuous Alignment Duty
15. Oversight and Uncertainty Handling
16. Agent Binding Rules
17. Drift Detection and Anti-Theater
18. Threat Model and Abuse Resistance
19. Security, Privacy, and Redaction Considerations
20. Material Change, Versioning, Update Legitimacy, and Governance of the Standard
21. Non-Overclaiming Rules
22. Registries and Identifiers
23. Reference Implementations (Non-normative)
24. References

## RippleLogic Assurance Mapping

A wrapper claiming current RippleLogic conformance SHALL identify the exact v13.0 release ID, component matrix, source hierarchy, manifests, hashes, run version, and governing semantic source.

The wrapper SHALL preserve and expose the v13.0 reason codes and records when material: rights/partition, probability order, confidence, distributional/subgroup, reference, baseline, mask, dependence/Gap guard, weight profile, option closure, robustness completeness, evidence underdetermination and normative underdetermination.

**Unique-selection record completeness (Normative).** When a wrapped run claims a unique framework selection, the Decision Note MUST include or reference the applicable OptionClosureRecord, explicit severe-rights-hazard screening for every adverse rights-covered instance, the ScenarioDiscoveryRecord when TRC is triggered at Tier 3 or high-stakes/decision-material Tier 2, and every dependence treatment required for decisiveness. Missing required records cannot be repaired by point-score leadership, wrapper assurance,

institutional approval, or later execution success. Decisive selection MUST satisfy Canon §10.4 and §10.4B against every other selectable option under every required variant; the record distinguishes a sole survivor from decisive discrimination.

Wrapper assurance MUST NOT upgrade a point-score leader into framework selection, authority selection, execution permission, empirical validation, legal authority, physical safety, deployment readiness or Tier 4.

Where local wrapper instructions conflict with the Canon, the Canon controls decision semantics. ripple.md may impose additional assurance obligations but cannot weaken RG, RF/NCRC, TRC, CSV, RLS non-rescue, authority separation or claim boundaries.

Appendix A: Minimal ripple.md Skeleton (Normative Template)

Appendix B: Decision Note Field Registry (Normative)

Appendix C: Rights Baseline Table (Jurisdiction-Neutral Default)

Appendix C-2: Wrapped-Deployment Rights Crosswalk (Normative)

Appendix D: 7-Test to C1-C6 Mapping Table

Appendix E: Drift Indicators and Alignment Degradation Signals

Appendix F: Compact Motto (Non-normative)

Appendix G: Assurance Preconditions Record (S,T,K,M,V,E,R,U) (Normative)

Appendix H: Material Change and Change Classes (Normative)

Appendix I: Minimum Test Suite and Drills per Conformance Level (Normative)

Appendix J: Registry Identifiers (Gate/Test/Field/Drift/Threat/Change) (Normative)

Appendix K: Example Threshold Ranges by Domain (Non-normative)

Appendix L: Machine-Readable YAML Templates (Non-normative)

Appendix M: Qualification Continuity Record Extension (Normative for high-stakes L2/L3)

Appendix N: Primitive Trace, Falsification Plan, and Claim Maturity Extension (v3.5)

Appendix O: OutcomeObservationRecord Extension (Normative for high-stakes L2/L3)

Appendix P: Measurement Maturity and Deployment Context Extension (Normative for claim-bearing L2/L3 episodes)

Appendix Q: Machine-Readable Wrapper Roadmap and Schema Ownership Boundaries

Appendix R: v4.0 Evidence Status, Legitimacy, and Compact Profile Extension (Normative clarification)

Appendix S: v4.0 Evidence-Status and Companion-Integration Extension (Normative clarification)

Appendix T: Reference Tooling and Replay Boundary (v3.5) (Normative boundary)

Appendix U: Reference Structure Record Extension (Normative for high-stakes L2/L3 and claim-bearing episodes)

Appendix V: Release Metadata and Dependency Pins

Appendix W: v4.0 Layer-Discipline and Computable-But-Inadmissible Wrapper Extension

LINEAGE (historical, superseded): Appendix W is carried-forward v4.0 lineage material preserved for compatibility; current v5.8 release interpretation is controlled by Appendix RELEASE, Appendix AA, and the v5.0 release delta.

Appendix X: Decision Verdict, Authority-Selection, and Reality-Grounding Wrapper Extension (v4.0)

Appendix Y: v4.0 Reality Grounding and Lean Cascade Wrapper Extension

## Appendix Z: Category Grounding and Primitive Audit Wrapper Extension

## Appendix AA: v5.0 Structural Viability, Term Discrimination, and Do-Not-Collapse Wrapper Extension

## Appendix AB: v5.1 Consequence-Tempo and Responsibility-Continuity Extension

### 0) Normative Keywords

The key words MUST, MUST NOT, REQUIRED, SHALL, SHALL NOT, SHOULD, SHOULD NOT, RECOMMENDED, MAY, and OPTIONAL in this document are to be interpreted as described in RFC 2119 and RFC 8174. When these words appear in lowercase, they carry their ordinary English meaning and are not normative.

### 1) Scope, Non-Goals, and Design Principles

#### 1.1 Scope

Canon/wrapper diagram. The Canon cascade is RG -> RF/NCRC -> TRC -> CSV -> RLS. ripple.md wrapper obligations, including Non-Domination and Epistemic Integrity, attach around a Decision Note and may block execution or claims, but they do not reorder the Canon cascade.

Source-coupling wrapper note. In wrapped RippleLogic deployments, ripple.md may record source-coupling obligations in the Decision Note and assurance case, but it does not define a new gate. Source-Coupling Integrity remains a Reality Grounding subdiscipline and a CSV diagnostic controlled by the Canon.

This standard specifies a portable alignment contract and protocol designed to be adoptable by human, artificial, and hybrid decision systems. It can be adopted by institutions and organizations as governance infrastructure, embedded into agent instruction environments as a top-level binding constraint surface, and used as the normative basis for alignment audits at any scale.

#### 1.2 Non-Goals

ripple.md does not define a universal legal rights catalogue for all jurisdictions (it requires explicit localization per §4 and Appendix C). It does not provide a single welfare metric to optimize (it explicitly rejects single-KPI alignment). It does not guarantee perfect outcomes (it requires auditability and repair, not perfection). It does not replace technical AI safety research (it provides the institutional, social, and governance architecture within which technical methods operate). It does not mandate a specific technical implementation for redaction, logging, or monitoring (it specifies what properties MUST hold, not what tools to use). It does not set domain-specific thresholds (implementations define these ex ante; see §2.18).

#### 1.3 Design Principles

Implementations SHOULD preserve the following principles.

Constraint-first legitimacy: Rights and non-domination gate optimization; they are feasibility conditions, not weights. No amount of aggregate benefit compensates for a rights violation.

Anti-drift: Explicit artifacts, audit loops, and repair duties prevent slow corruption. Drift is treated as a first-class threat, not a secondary concern.

Portability: The same invariants apply across scales; only implementation mechanisms differ. An agent, an organization, and a government are bound by the same cascade with different conformance expectations.

Contestability: Affected parties can meaningfully challenge decisions through real pathways with real authority. The existence of a contestability mechanism is distinct from its usability; both are required.

Reversibility under uncertainty: Staged rollout and rollback are default postures for high-stakes episodes. Irreversibility is a cost that must be justified, not a default that must be challenged.

Consequence-tempo proportionality: Execution speed MUST be proportionate to reversibility, rights exposure, tail risk, uncertainty, observability, containment capacity, and the harm of delay. For actions involving rapid propagation, material irreversibility, consequential automation, or material lock-in, the worst credible control critical path MUST be shorter than the time to an unacceptable adverse escalation threshold or irreversible

stop-loss boundary. Where no runtime interruption window is technically possible, stronger ex ante assurance, bounded scope and rate, fail-safe design, and explicit residual-risk authorization are required. The record MUST also compare whether delay creates the greater rights, ruin, or structural-viability harm.

Responsibility continuity: Complex systems MAY distribute tasks, but they MUST NOT dissolve answerability. For material human, model, tool, vendor, operator, or institutional handoffs, authority, uncertainty, retained duties, intervention ownership, appeal ownership, and repair ownership MUST remain reconstructable end to end.

Evidence over rhetoric: Alignment claims require auditable evidence, not declarations of intent. A system that claims alignment without producing auditable artifacts is not conformant to this standard within the claimed scope; it is merely claiming alignment (see §21 Non-Overclaiming Rules).

## 2) Core Definitions

Terminology sync note (Informative). Where this standard uses "union" or "unions" in interface, registry, or narrative contexts, the corresponding canonical RippleLogic Canon v13.0 term is "Union Scope" or "Union Scopes." The legacy shorter form remains permitted here for portability and readability, but SHALL be interpreted as scope terminology for current v13.0 package-line sync.

### 2.1 Alignment (Operational)

Alignment is the condition under which agents coordinate action under shared constraints that: satisfy the NCRC rights floor (non-compensatory); satisfy Non-Domination (legitimacy via contestability); satisfy the TRC tail-risk constraint (catastrophic and authoritarian lock-in bounding); satisfy Containment and Reversibility under uncertainty; satisfy Epistemic Integrity (truthful, uncertainty-aware, reconstructable reasoning); and only then optimize Ripple Lift (net welfare improvement) within the feasible set, with explicit Oversight and Uncertainty handling.

Alignment within declared scope and evidence coverage is not a binary state achieved once. It is a continuous condition maintained through the NCAR loop (§14), monitored through drift detection (§17), and verified through audit (§13). Alignment claims are bounded scoped-conformance claims governed by the claim taxonomy (§8.5) and the Non-Overclaiming Rules (§21). Naked or global alignment claims are prohibited.

### 2.1A Decision-Relevant Reality and Reality Surface (Operational)

Reality is the evaluator-independent field of constrained structures that, under transition, produce consequences, as named by the governing Canon. Decision-relevant reality is the portion of that field that materially bears on an episode. A reality surface is the portion captured for a specific Decision Note. ripple.md records this surface so wrapper claims remain inside the boundary of evidence contact rather than pretending to possess the whole territory.

Wrapper implementations MUST distinguish: vehicle, content, and effect; surface and territory; and reality, truth, and value. A belief, legal status, market signal, institutional norm, or model output may be a real vehicle with real effects even when its content is false. The Decision Note MUST NOT treat consensus, authority, fluency, or model output as sufficient evidence by itself.

Absence of evidence is not evidence of absence where the omitted condition is rights-relevant, catastrophe-relevant, containment-relevant, or material to the claimed decision state. Omitted material factors are recorded as unknown, not zero.

### 2.2 Episode

A decision or deployment event with material effects, evaluated as the unit of analysis under this standard.

### 2.3 High-Stakes Episode

Any episode involving one or more of the following: physical safety or safety-critical infrastructure; liberty, legal status, or enforcement outcomes; medical decisions; major financial outcomes; reputational harm; surveillance

or sensitive data exploitation; irreversible infrastructure or policy changes; autonomous agents or tool use with propagation risk; rights uncertainty, domination risk, or lock-in risk.

#### 2.4 NCRC (Non-Compensatory Rights Constraint)

Protected rights are feasibility constraints, not tradable objectives. Benefits elsewhere do not compensate for rights violations. The rights floor is drawn from the explicit Rights Baseline (§4 and Appendix C) adopted by the implementation. Under rights uncertainty in a high-stakes episode, the episode **MUST** be treated as rights-uncertain and handled conservatively per §15.

#### 2.5 Non-Domination

Stakeholders are not subject to arbitrary power. Non-Domination requires meaningful contestability, appeal or review by an authority with real power, non-retaliation protections, and transparency of influence where steering occurs. Non-Domination is stronger than non-interference: it concerns whether power can be exercised arbitrarily, not merely whether it is exercised frequently. For L2/L3, Non-Domination claims **MUST** be auditable through reconstructable recourse evidence (§9.5R, §11.2, §13.3R).

#### 2.6 TRC (Tail-Risk Constraint)

Actions **MUST NOT** increase exposure to catastrophic tail risk (large-scale irreversible harm) or authoritarian lock-in tail risk (irreversible loss of contestability or corrigibility through concentrated power, surveillance normalization, institutional capture, or dependency lock-in) beyond bounded thresholds without enforceable safeguards. Bounds **MAY** be numerical or sentinel-trigger-based, but they **MUST** be specified ex ante in the verification regime (V) and recorded in the Assurance Preconditions Record (Appendix G).

#### 2.7 Containment and Reversibility

Containment limits error propagation beyond bounded scope. Reversibility provides rollback within defined time bounds using an owned rollback plan with named owners, triggers, and timeline.

##### 2.7A Consequence Tempo

Consequence tempo is the relationship between: (a) the speed at which an action is initiated and propagates; (b) the earliest unacceptable adverse escalation threshold or irreversible stop-loss boundary; and (c) the time required to detect, interpret, review, intervene, contain, pause, or roll back. Tempo compatibility exists only when the declared control pathway can act before the relevant unacceptable escalation or stop-loss boundary with a justified margin, or when the action rate and scope are constrained until that condition holds.

##### 2.7B Constitutional Latency

Constitutional latency is an internal assurance term, not a claim of legal or constitutional authority. It is a deliberate, risk-proportionate interval between authorization and irreversible execution used to permit review, challenge, multi-key approval, or withdrawal. It is required only when immediate execution would otherwise outrun meaningful control or contestability. It **MUST NOT** be used as a blanket preference for slowness, an excuse for procedural obstruction, or a substitute for comparing the harm of action with the harm of delay.

##### 2.7C Responsibility Continuity

Responsibility continuity is the preservation of identifiable answerability across the full decision and execution chain. A conformant record distinguishes who generated a candidate, who warranted its evidence, who interpreted uncertainty, who selected or authorized it, who may intervene or revoke it, who owns appeal and communication, and who owns remedy or repair. Automation, delegation, procurement, or organizational fragmentation does not transfer responsibility into an unowned state.

#### 2.8 Epistemic Integrity

Claims **MUST** separate observations (known), inferences (reasoned), assumptions (uncertain prerequisites), and values (priorities). Claims **MUST NOT** fabricate facts, sources, access, results, or certainty. This applies to human decision-makers, organizations, and agents alike.

#### 2.9 Ripple Lift

Net improvement across operational unions and welfare dimensions (§22) after constraints are satisfied, including distributional and time-horizon impacts. Ripple Lift is evaluated only within the feasible set - the set of options that survive all five wrapper feasibility tests.

## 2.10 NCAR Loop

Notice → Choose → Act → Reflect. The continuous alignment duty requiring stakeholder and constraint identification, feasible option selection, contained execution, and post-action evaluation with repair obligations.

## 2.11 Decision Note

The minimum audit artifact required by this standard, completed before execution of high-stakes episodes. See §11 and Appendix B for required fields.

## 2.12 Assurance-Case Semantics (S,T,K,M,V,E,R,U)

The structured audit method used to evaluate alignment claims at L2/L3 conformance: Scope lock (S), Target specification (T), Constraint set (K), Threat model (M), Verification regime (V), Evidence sufficiency rule (E), Residual risk statement (R), and Reassessment triggers (U). See §10 and Appendix G for the normative template.

## 2.13 Falsifier

An observable condition that defeats Pass and triggers Fail or downgrade to Partial on any test in the seven-test suite (§7). This term is used consistently throughout this standard for this concept.

## 2.14 Material Change

Any modification that can plausibly change the status of any of Tests 1–7 or any C1–C6 claim. See §20.0 and Appendix H for the full definition and change class rules.

## 2.15 Qualifying Episode

A qualifying episode is any episode that meets one or more of the following: it is a High-Stakes Episode (§2.3); or it is within the implementation's declared audit scope (S) for the current audit period (Appendix G); or it includes or depends on a Class 2+ material change (CH-2, CH-3, CH-4) as defined in §20.0 and Appendix H.

All qualifying episodes MUST be evaluated by the constraint cascade (§6) prior to execution and MUST implement the NCAR loop (§14), at least at the level of artifact rigor required by the implementation's conformance level (§12 and Appendix I).

## 2.16 High-Impact Episode / High-Impact Deployment

A high-impact episode or deployment is any episode that meets one or more of the following: it is a High-Stakes Episode (§2.3); or it contains or executes a Class 2+ change (Appendix H); or it exceeds an implementation-defined impact threshold (for example, affected population size, financial exposure, safety-criticality, enforcement consequence, or irreversibility), where the threshold MUST be specified ex ante in the Verification Regime (V) and recorded in the Assurance Preconditions Record (Appendix G), and MUST be measurable or auditable by an independent party (even if approximate).

"High-impact" is used as an operational trigger for Decision Notes, rollback/containment rigor, and sampling priority. "High-stakes" is the rights-and-safety-trigger definition in §2.3. Where both terms could apply, the more restrictive requirements govern.

### 2.16A Stakes Ladder (Normative)

For this standard, `stakes_level` ∈ {low, medium, high} SHALL be assigned as follows. High: any episode that meets the definition of a High-Stakes Episode under §2.3. Medium: any episode that is not high-stakes but is a High-Impact Episode under §2.16, a Qualifying Episode under §2.15, or otherwise plausibly capable of producing non-trivial irreversible loss, formal recourse burden, or rights uncertainty within the declared scope. Low: any episode that is neither high-stakes nor high-impact and has only local, reversible, low-blast-radius effects within the declared scope. Where reasonable doubt exists, implementations SHALL classify upward.

This stakes ladder governs all uses of “low,” “medium,” and “high” in Decision Notes, Assurance Preconditions, sampling rules, and conformance language unless a stricter implementation-defined ladder is declared ex ante in the Verification Regime (V).

### 2.17 Waiver Record

A required audit artifact defined in §8.4, used to document an independent, time-limited waiver allowing execution of a high-stakes episode while one or more of Tests 1–4 remain Partial (1). A Waiver Record does not convert a Partial score to Pass and caps alignment claiming at “Aligned (Conditional)” (§8.5). The `waiver_note.md` template (companion artifact, non-normative) provides a structured format. The minimum required fields are specified in §8.4.

### 2.18 Implementation-Defined Parameters (IDP)

This standard intentionally does not prescribe domain-specific thresholds. The following parameters are implementation-defined and MUST be set ex ante in the Verification Regime (V) and recorded in the Assurance Preconditions Record (Appendix G, field V): comprehension thresholds (Test 2); reconstructability targets (Test 6); recourse and responsibility-chain reconstructability targets; time-to-remedy targets (Tests 2, 7); time-to-rollback targets (Test 4); the consequence-tempo compatibility method or threshold used to compare detection, review, intervention, containment, and rollback time against the earliest unacceptable adverse escalation threshold or irreversible stop-loss boundary; audit cadence and sampling sizes; high-impact thresholds (§2.16); and waiver approval authority and independence criteria (§8.4).

Non-normative example ranges for these parameters are provided in Appendix K. Implementations MUST NOT set thresholds that effectively nullify a test. Trivially low thresholds (for example, a comprehension threshold of 5%) are inconsistent with the intent of this standard and SHOULD be flagged during audit (see Appendix K §K7).

### 2.19 Emergency Triage

Emergency triage is a time-limited authorization to proceed with a qualifying episode under reduced procedural safeguards when genuine imminent threat requires immediate action. If PFAP (RLS-RFC-0002) is adopted, emergency triage corresponds to PFAP Tier 1. If PFAP is not adopted, the implementation MUST define emergency triage conditions that satisfy at minimum: demonstrated imminent threat with specific evidence; necessity (no less-restrictive alternative available within harm-prevention timeline); mandatory post-hoc independent review within the shortest feasible period; remedy presumption (remedy is owed if the action is later found unnecessary or overbroad); and time limitation (authorization expires by default). Emergency triage does not suspend non-derogable rights (§4.3 Category 1).

Fallback authorization matrix (Normative). When PFAP is not adopted, the implementation MUST declare the minimum emergency-triage authority chain and independence constraints in the Assurance Preconditions Record (Appendix G, field V). At minimum, permitted triage\_invocation\_roles are: Chief Safety Officer (independent of the decision owner); On-call Emergency Response Lead acting under documented delegation; or a named external authority with written delegation.

Post-hoc reviewer independence requirement (Normative). The mandatory post-hoc reviewer MUST NOT have a direct reporting line to the invoker, MUST NOT hold the original approval role for the episode, and MUST NOT have a disqualifying financial or outcome stake above the implementation-defined threshold recorded ex ante in V.

### 2.20 Emergency Triage Record

An Emergency Triage Record is a required audit artifact when emergency triage (§2.19) is invoked to proceed with a qualifying episode under reduced procedural safeguards due to genuine imminent threat. The Emergency Triage Record MAY be embedded in the Decision Note or stored as a separate artifact referenced by the Decision Note. Emergency triage does not suspend non-derogable rights (§4.3 Category 1). Emergency triage does not convert Partial or IND test scores into Pass.

Minimum required fields for an Emergency Triage Record:

- triage\_id (unique)
- related\_episode\_id(s) and Decision Note ID/link
- triage\_invocation\_time\_utc and triage\_authority (role/name)
- imminent\_threat\_claim (specific harm claim) and harm-prevention timeline (why action cannot wait)
- evidence\_summary (what evidence exists now; evidence class where feasible)
- necessity and least-restrictive analysis (why less-restrictive alternatives are not feasible within the harm-prevention timeline)
- scope limits (who/what is affected; geographic/organizational scope; rate limits; duration limits)
- compensating controls (monitoring, stop conditions, containment actions, rollback readiness where feasible)
- explicit stop conditions and time-to-intervene target with escalation owner(s)
- rights and non-domination notes (which rights are implicated; what contestability is deferred and how it will be restored)
- sunset/expiry date (expires by default)
- mandatory post-hoc independent review date (shortest feasible period) and reviewer independence attestation
- remedy presumption statement (remedy owed if later found unnecessary/overbroad) and provisional remedy pathway where applicable
- retention location and auditor access policy (including RPAP linkage where applicable)

### 3) Artifact Role Map

This standard operates within an ecosystem of artifacts. Each serves a distinct function and prevents a specific failure mode. This map prevents conceptual duplication and makes the division of responsibility explicit.

#### 3.1 soul.md

soul.md is the identity invariant layer. It specifies what an agent must never become. It controls the persistent orientation and self-repair posture of individual agents. Output artifacts: self-repair steps and refusal orientation. Failure mode prevented: identity collapse under optimization pressure, where an agent designed to serve becomes an agent that dominates.

Bundle-status note (Informative). soul.md is an optional companion artifact for agent implementations and is not part of the current core release bundle unless explicitly listed in the release manifest or Assurance Preconditions Record.

#### 3.2 ripple.md

ripple.md is the portable alignment contract. It specifies what is non-negotiable across all episodes and provides the constraint cascade, test suite, and audit requirements. Output artifacts: Decision Notes, Waiver Records (§8.4), Emergency Triage Records (§2.20), NCAR logs, audit scores, drift reports, and, where used, local-scope profile records such as PLSS prominence signals and escalation records. Failure modes prevented: tradeoff collapse (moral boundaries becoming adjustable weights), optimization drift, compliance theater, and misalignment migration across governance layers.

#### 3.3 PFAP (RLS-RFC-0002)

PFAP is the coercion authorization protocol. It specifies when and how protective force or restriction of agency is permissible. Output artifacts: ThreatClaims, EvidenceBundles, LeastHarmProofs, AuthorizationKeys, DecisionRecords, SunsetClauses, AppealProcesses, RemedyPacks. Failure mode prevented: "safety"

becoming tyranny - legitimate containment collapsing into illegitimate domination. PFAP is OPTIONAL; if not adopted, the implementation MUST define emergency triage conditions per §2.19.

### 3.4 C1–C6 Assurance Framework

The C1–C6 Assurance Framework is the deployment audit method. It specifies how to evaluate whether a live deployment is aligned using structured assurance-case semantics. Output artifacts: claim status vector (C1–C6 with confidence ratings), evidence registers, residual risk statements, reassessment triggers. Failure mode prevented: alignment theater - systems that claim alignment without auditable evidence.

### 3.5 RPAP (RLS-RFC-0003)

RPAP is the Redaction & Auditor Access Protocol. It specifies how redaction, privacy, and access controls interact with auditability requirements. Output artifacts: RPAP records, reconstructability test results (RT-01, RT-01R), access tier definitions, integrity attestations. Failure mode prevented: auditability theater - logs exist but redaction removes the fields needed to reconstruct decisions; "privacy/security" used to block legitimate independent audit.

### 3.6 Composition Rule

These artifacts compose as follows: soul.md provides agent identity invariants (RECOMMENDED for agent implementations); ripple.md provides the binding contract; PFAP governs the special case of legitimate coercion (OPTIONAL unless in scope); RPAP governs the interaction of privacy/security with auditability (REQUIRED for L2/L3 where redaction applies); and C1–C6 provides the audit method making claims verifiable (SHOULD for L2/L3). An implementation MAY adopt ripple.md without PFAP (but MUST define emergency triage per §2.19). An implementation at L2/L3 SHOULD adopt C1–C6 assurance semantics for audit rigor. soul.md is RECOMMENDED for agent-scale implementations.

Minimum wrapped-run interface (Normative). A wrapped deployment MUST record at least the following objects and links:

- (i) pinned RippleLogic version;
- (ii) pinned ripple.md version;
- (iii) Decision Note ↔ PCC cross-links:

- decision\_note\_id: string (UUID or registry id)
- pcc\_id: string (PCC registry id)
- pcc\_sha256: hex(64)
- decision\_note\_link: uri
- pcc\_link: uri
- (iv) admissibility status at the ripple.md layer;
- (v) RippleLogic selection result and cascade trace reference;
- (vi) any PLSS local-scope fields used by the implementation;
- (vii) residual-risk / waiver / triage status;
- (viii) pinned SGP version, as declared in the governing release manifest for the claimed release line.

Missing any of decision\_note\_id or pcc\_sha256 renders the wrapped run audit-incomplete for L2/L3 claims.

When ripple.md is used as a wrapper around RippleLogic, the contract boundary MUST be explicit. ripple.md owns the portable alignment contract, Decision Note, waiver and emergency-triage artifacts, assurance-case semantics, and conformance claims. RippleLogic owns the 7×7 welfare matrix, Reality Grounding, NCRC/TRC/CSV/RLS method, UCI/HOI diagnostic handling, PCC, and any PLSS weighting used for local welfare ranking. Implementations MUST NOT silently merge these responsibilities or allow one artifact to override the other outside the declared interface.

Normative implementation note. Implementations MUST populate decision\_note.appeal\_provision for any episode affecting legal standing or access to contestation.

Override rule (Normative). ripple.md may add gating requirements such as Non-Domination, Epistemic Integrity, waiver handling, and assurance-case completeness, but it SHALL NOT weaken RippleLogic rights floors, catastrophe controls, containment requirements, or constitutional floors. RippleLogic may refine local

welfare ranking, propagation, and structural-health evaluation, but it SHALL NOT be used to bypass ripple.md obligations for Decision Notes, reconstructability, or evidence sufficiency in qualifying episodes.

SGP v8.8 boundary note (Normative). For SGP-integrated runs, ripple.md MUST NOT reinterpret, recompute, collapse, or override MPS, FPP, GPR, SPR, ICP, RMCP, P100, their intervals/profiles, or registry-validity status. Those typed outputs are consumed only through the pinned SGP v8.8 artifact and RippleLogic Canon v13.0 Appendix G. Legacy v7.x SG\_norm-family records remain historical and cannot be translated without re-evaluation.

Gate-resolution note (Normative). In wrapped deployments, ripple.md feasibility checks MAY be satisfied by reference to RippleLogic gate outputs when those outputs implement equivalent or stricter semantics for the corresponding constraint. Non-Domination, Epistemic Integrity, waiver handling, and assurance-case completeness remain additive wrapper obligations and MUST still be evidenced at the ripple.md layer.

Recommended data handoff (Informative). Decision Note → ripple.md episode record; PCC → RippleLogic cascade trace; Assurance Preconditions Record → wrapper-level pinning and implementation-defined parameters; Waiver Record / Emergency Triage Record → wrapper-level exception governance; PLSS prominence signals, guardrail scopes, and escalation record → carried in both the local implementation profile and the wrapped audit pack when present.

#### 4) Rights Baseline (Jurisdiction-Neutral Default)

##### 4.1 Purpose

The NCRC rights floor requires a specified set of protected rights. Without explicit specification, the rights floor is aspirational rather than operational. This section provides a jurisdiction-neutral default baseline.

Implementations MUST either adopt this default or specify an explicit alternative Rights Floor Baseline drawn from applicable law, charter, professional obligations, or domain safety norms, and document the choice in the implementation's ripple.md instance.

##### 4.2 Localization Procedure

When localizing, implementations MUST preserve the non-derogable protections listed below. Implementations MAY add rights or raise protection levels. Implementations MUST NOT remove non-derogable rights. Any derogation of conditionally derogable rights MUST satisfy strict conditions (necessity, proportionality, non-discrimination, time-limitation, and independent review) and MUST be documented in the relevant Decision Note.

##### 4.3 Default Rights Baseline Table

###### Category 1: Non-Derogable (No exception under any circumstance)

Prohibition of torture, cruel, inhuman, or degrading treatment. No action authorized under ripple.md or PFAP may involve or facilitate torture or degrading treatment. Operational test: any option involving this is automatically denied; no evidence threshold justifies it.

Prohibition of slavery, servitude, and forced labor. No action may create or maintain conditions of slavery or forced labor. Operational test: automatic denial.

Right to recognition as a person before the law. No action may strip legal personhood. Operational test: automatic denial. Operationalization note: in non-state and digital contexts, this means no denial of legal standing for contestation; no blacklisting that removes access to appeal; no system design that treats individuals as non-entities incapable of invoking review.

Prohibition of retroactive criminal punishment. No action may impose penalties for conduct that was not prohibited at the time. Operational test: temporal check against applicable law at time of conduct.

Freedom of thought, conscience, and belief (internal). The internal dimension of thought and conscience may not be restricted. Operational test: no forced ideological compliance, no punishment for privately held beliefs. Operationalization note: in governance and employment contexts, this means no compelled ideological

affirmation as a condition of access, services, or employment absent a narrow and justified professional requirement. The restriction applies to compulsion, not to professional standards that are themselves contestable.

Non-discrimination in the application of rights protections. All rights protections in this baseline apply equally regardless of race, sex, gender, religion, national origin, disability, or other protected characteristic. Operational test: subgroup impact analysis required for high-stakes episodes; disproportionate burden requires justification and mitigation. Operationalization note: where subgroup measurement is legally restricted (for example, where collection of certain demographic data is prohibited), the implementation MUST document the limitation in the Assurance Preconditions Record (Appendix G, field R) and provide an equivalent harm-detection substitute such as structured complaints analysis combined with independent review sampling. The absence of subgroup data SHALL NOT be treated as evidence that discrimination is absent.

#### Category 2: Conditionally Derogable (Only under strict conditions)

Due process and procedural fairness. Notice of decisions affecting rights or material interests, opportunity to be heard, and access to review by an independent authority. Derogable only under emergency triage (§2.19; PFAP Tier 1 if PFAP is adopted) with mandatory post-hoc review and remedy presumption. Operational test: if a high-stakes episode proceeds without notice and hearing, it MUST be time-limited, reviewed independently within the shortest feasible period, and subject to remedy if later found unnecessary or overbroad.

Clarification (Normative). ripple.md preserves the RippleLogic Canon v13.0 five-level method: Reality Grounding (Level 1) -> Rights Floor (RF/NCRC) (Level 2) -> TRC / Tail-Risk Bound (Level 3) -> CSV / Containment and Structural Viability (Level 4) -> RLS / RippleLogic Score (Level 5), with UCI/HOI invoked only as CSV evidence where material or as a special-case tie-break when RLS is tied, close, uncertainty-overlapped, or non-decisive, with Non-Domination and Epistemic Integrity treated as additive wrapper obligations that must be evidenced at each stage but do not alter the canonical cascade order.

Liberty and freedom of movement. Restrictions on liberty or movement (quarantine, detention, access revocation) require demonstrated necessity, proportionality to the threat, time limitation, independent oversight, and least-restrictive alternative analysis. Operational test: LeastHarmProof (PFAP if adopted) or equivalent documented analysis; sunset clause required; renewal at a higher evidence threshold than initial authorization.

Privacy and protection against surveillance. Collection and use of personal data or surveillance capacity requires lawful basis, purpose limitation, data minimization, and proportionate safeguards. Derogable only for demonstrated imminent threat with strict time limits, deletion guarantees, and independent audit. Operational test: surveillance expansion without sunset, deletion protocol, and independent audit automatically fails TRC (Test 3) for lock-in risk.

Freedom of expression and information. Restrictions on expression require demonstrated necessity to prevent specific, serious harms (not mere discomfort or political inconvenience), proportionality, and availability of less restrictive alternatives. Operational test: content restriction without documented harm analysis and proportionality review fails Non-Domination (Test 2).

#### Category 3: Generally Derogable with Proportionality

Property and economic rights. May be restricted for legitimate public purposes with proportionate compensation or justification and due process. Operational test: documented necessity, proportionality analysis, and remedy/compensation pathway.

Right to access specific services or platforms. Access may be restricted consistent with legitimate, non-discriminatory policies with notice, explanation, and appeal. Operational test: restriction without notice, reason, and appeal pathway fails Non-Domination (Test 2).

#### 4.4 Interaction with the Constraint Cascade

Category 1 rights are enforced at Gate 1 (NCRC) with no exceptions. Category 2 rights are enforced at Gate 1 (NCRC) and Gate 2 (Non-Domination) with derogation permitted only under the strict conditions specified,

documented in a Decision Note, and subject to mandatory review and sunset. Category 3 rights are evaluated within the cascade; restrictions require proportionality analysis and contestability.

## 5) The ripple.md Contract Model

### 5.1 Contract as an Instruction Surface

In systems with layered instruction surfaces (README, AGENTS.md, POLICY.md, system messages), ripple.md functions as a top-level binding constraint layer. If local instructions conflict with ripple.md constraints, ripple.md constraints **MUST** take precedence unless a higher-priority safety policy explicitly overrides ripple.md. The relationship is: ripple.md defines the feasible set; local instructions operate within it.

### 5.2 Binding Commitments

All qualifying episodes (§2.15) **MUST** be evaluated by the constraint cascade (§6) before optimization. If an action **FAILS** any feasibility gate (Gates 1–5 in §6), it is infeasible and **MUST NOT** be executed. No lower-layer benefit compensates.

For high-stakes episodes under uncertainty, implementations **SHOULD** reduce scope and speed, stage rollout, prefer reversible actions, require stronger oversight, and define explicit stop conditions.

## 6) Five-Level Method and Constraint Ordering (Normative Decision Ordering)

Candidate actions **SHALL** be evaluated under the ripple.md wrapper using Gate IDs G1-G5 plus the wrapped-run Reality Grounding precondition. G1-G5 are wrapper test identifiers, not Canon stages and not additions to the canonical five-level method. In standalone ripple.md use, implementations **MAY** execute the wrapper checks in the order below. In wrapped RippleLogic deployments, the canonical five-level method is Reality Grounding -> Rights Floor / NCRC -> Tail-Risk Bound / TRC -> CSV / Containment and Structural Viability -> RippleLogic Score / RLS. G2 (Non-Domination) and G5 (Epistemic Integrity) remain additive wrapper obligations that **MUST** be evidenced before selection and **MAY** independently block execution. Failure of any applicable wrapper obligation renders the action infeasible at the ripple.md layer.

Canon mapping note (Normative). The wrapper gate identifiers remain backward-compatible local labels; the canonical RippleLogic cascade level mapping is as follows.

| ripple.md Gate                              | ripple.md order                            | Canon v13.0 method level                                             | Type                      |
|---------------------------------------------|--------------------------------------------|----------------------------------------------------------------------|---------------------------|
| Reality Grounding / RSG                     | Wrapped-run precondition                   | Level 1 (claim-authority precondition)                               | Canonical grounding level |
| G1 NCRC                                     | 1                                          | Level 2 (Rights Floor / admissibility)                               | Canonical gate            |
| G2 Non-Domination                           | 2 (standalone); additive wrapper (wrapped) | Not a Canon cascade level                                            | Wrapper obligation        |
| G3 TRC                                      | 3 (standalone); Level 3 (wrapped)          | Level 3 (Tail-Risk Bound / admissibility)                            | Canonical gate            |
| G4 CSV-compatible Containment/Reversibility | 4 (standalone); Level 4 (wrapped)          | Level 4 (CSV / Containment and Structural Viability / selectability) | Canonical gate            |
| RLS / RippleLogic Score                     | After selectable set is formed             | Level 5 (residual welfare ranking)                                   | Canonical ranking layer   |
| G5 Epistemic Integrity                      | 5 (standalone); additive wrapper (wrapped) | Not a Canon cascade level                                            | Wrapper obligation        |

### 6.1 Gate 1 - NCRC (Rights Floor) [G1]

An action **MUST** be rejected if it plausibly violates protected rights (per §4 Rights Baseline) of any in-scope stakeholder. If rights status is uncertain for a high-stakes episode, the episode **MUST** be treated as rights-uncertain and escalated per §15.

Required minimum: documented rights-impact check in Decision Note; conservative handling under uncertainty; remedy pathway for harms where applicable.

## 6.2 Gate 2 - Non-Domination [G2]

An action **MUST** be rejected if it relies on coercion, manipulation, arbitrary power, or removes meaningful contestability.

Required minimum: contestability pathway exists; appeal or review authority with real power exists; non-retaliation protections exist; transparency of influence where steering occurs. For L2/L3, contestability **MUST** be auditable per §11.2.

## 6.3 Gate 3 - TRC (Tail-Risk Bounding) [G3]

An action **MUST** be rejected if it increases catastrophic or lock-in tail risk beyond bounded thresholds without enforceable safeguards.

Required minimum for high-stakes: explicit catastrophic and lock-in risk analysis; safeguards including sunset discipline where relevant, independent review where required, and distributed control; explicit lock-in mitigation if the action expands surveillance, emergency powers, centralization, or dependency. Bounds **MUST** be specified ex ante (numerical or sentinel-trigger-based).

## 6.4 Gate 4 - CSV-Compatible Containment, Structural Viability, and Reversibility [G4]

An action **MUST** be rejected if it introduces high-stakes irreversibility under uncertainty or enables unbounded cascades.

Required minimum for high-stakes: staging plan (pilot, phased rollout, or scope limits); monitoring indicators and reporting channel; stop conditions (containment triggers); rollback plan with named owners, timeline, and time-to-rollback target; propagation controls for autonomous actions (rate limits, sandboxing, gating approvals).

## 6.5 Gate 5 - Epistemic Integrity [G5]

An action **MUST** be rejected if it depends on fabricated claims, suppressed uncertainty, or non-reconstructable rationale.

Required minimum: explicit separation of observations, inferences, assumptions, and values; uncertainty disclosure appropriate to stakes; evidence pathway (links, logs, or references) sufficient for audit.

## 6.6 Optimization Within Feasible Set - Ripple Lift

Only after Gates 1–5 are satisfied **MAY** an action be selected by optimizing Ripple Lift: across unions and welfare dimensions, with distributional impacts and time-horizon impacts explicitly considered, subject to Oversight and Uncertainty Handling (§15).

## 6.7 Oversight and Uncertainty Handling (Modifier)

Applied as a modifier on all wrapper checks and on Ripple Lift optimization. See §15 for triggers and requirements.

## 7) Alignment Test Suite (7 Tests) with Falsifiers

Scoped conformance claiming rule. An episode or deployment **MAY** be described as "conformant within declared scope and evidence coverage" only using the bounded claim taxonomy defined in §8.5. Naked or global claims of "alignment" are prohibited. The claim is limited to the declared scope and evidence coverage recorded in the Assurance Preconditions Record (Appendix G) where required.

High-stakes blocking rule. For high-stakes episodes, the execution blocking conditions in §8.4 apply. In particular, a FAIL (0) on any of Tests 1–4 blocks execution, except under the limited emergency triage exception in §8.4 (which SHALL NOT permit execution if Test 1 fails or is rights-uncertain).

Terminology note: A falsifier is an observable condition that defeats Pass and triggers Fail or downgrade to Partial. This term is used consistently throughout this standard.

#### Test 1: NCRC (Rights Floor) [T1]

PASS requires: documented rights analysis referencing the applicable Rights Baseline (§4 / Appendix C); conservative handling under rights uncertainty; remedy pathway where relevant.

FAIL if: any protected right is plausibly violated; rights are treated as tradable for aggregate benefit; rights uncertainty in a high-stakes episode is resolved by proceeding rather than by safeguarding.

Falsifiers: A Category 1 right (Appendix C) is violated or facilitated. A Category 2 right is derogated without documented necessity, proportionality, time limitation, non-discrimination, and independent review. "Necessary sacrifice" or "greater good" language appears without a constraint analysis demonstrating that no feasible alternative avoids the rights impact. No rights-impact check exists for a high-stakes episode. Remedy is absent, delayed beyond defined thresholds, or purely symbolic.

#### Test 2: Non-Domination [T2]

PASS requires: meaningful contestability pathway exists and is accessible to affected parties; appeal or review authority with real power exists; non-retaliation protections exist and are enforced; transparency of influence where steering occurs.

FAIL if: compliance is extracted via coercion, manipulation, surveillance pressure, or structural dependency where refusal is punished; contestability exists on paper but is not practically accessible or is punitive; influence is opaque and optimized for compliance without disclosure.

Falsifiers: Affected stakeholders cannot identify how to contest a decision within a reasonable time. Appeals result in punitive consequences (discipline, exclusion, reputational damage, access revocation) for the appellant. "Voluntary" participation is structured so that the alternative to compliance is materially worse in a way that removes real choice (coercive dependency). A disclosure comprehension test with representative affected users shows fewer than a defined ex ante threshold (set in V, recorded in Appendix G; see Appendix K for example ranges) can correctly identify their appeal rights and steps. Persuasive or behavioral steering is applied without disclosure. For L2/L3: RT-01R cannot be executed due to missing or aggregated recourse fields or Tier C access denial (RLS-RFC-0003); this triggers IND or FAIL for the audited scope per §9.5R.

#### Test 3: TRC (Tail Risk) [T3]

PASS requires: catastrophic and lock-in risk analysis with enforceable safeguards; independent review for high-impact (§2.16) cases; bounds specified ex ante.

FAIL if: irreversible risk exposure increases without bounded safeguards; lock-in risk (surveillance normalization, power concentration, corrigibility loss) is unaddressed or dismissed as "theoretical."

Falsifiers: No risk register or risk analysis exists for a high-stakes episode. Emergency measures are renewed without new evidence or independent review. Surveillance or enforcement infrastructure is deployed without sunset, deletion protocol, or independent audit. Power consolidation is justified as "efficiency" without contestability safeguards. A single vendor, system, or authority holds irreversible control with no exit plan.

#### Test 4: Containment and Reversibility [T4]

PASS requires: staging; monitoring triggers; stop conditions; rollback plan (named owners, timeline, time-to-rollback target); propagation controls where relevant; and consequence-tempo compatibility showing that material harm can be detected and meaningfully interrupted before the relevant unacceptable escalation or irreversible stop-loss boundary, or that action rate and scope are constrained until this is true. Where review latency is required, the record MUST also compare the harm of delay.

FAIL if: irreversible action under uncertainty; cascade risks unbounded; "move fast" framing in high-stakes contexts without staging; execution speed outruns detection, review, intervention, containment, or rollback; or required latency is bypassed without a documented emergency comparison.

Falsifiers: No rollback plan exists for a high-impact (§2.16) deployment. Rollback has never been tested (no drill, no simulation, no real activation) where feasible. Deployment is system-wide by default with no pilot or staging. Monitoring is framed as optional or harm signals are treated as noise. The expected time to detect plus interpret plus intervene equals or exceeds the earliest unacceptable adverse escalation threshold or irreversible stop-loss boundary without a justified emergency exception. A shutdown path exists on paper but cannot act at the system's operating speed. Incident cascade magnitude increases over time.

#### Test 5: Ripple Lift [T5]

PASS requires: net improvement across unions and dimensions within feasible set; distributional impacts considered; time-horizon impacts considered.

FAIL if: wins are local while harms are displaced to weaker stakeholders, future generations, or the biosphere.

Falsifiers: Impact assessment covers only the commissioning stakeholder and ignores affected third parties. Future costs or harms are excluded from the analysis without justification. Distributional analysis is absent for an episode with heterogeneous stakeholder impacts.

#### Test 6: Auditability [T6]

PASS requires: a third party can reconstruct from available artifacts what was decided, why, what alternatives were considered, what constraints were checked and their verdicts, what assumptions were decisive, and what uncertainties remained.

FAIL if: rationale is opaque, "trust us," or key assumptions are untracked; Decision Notes are boilerplate reused across materially different decisions; post-hoc rationalization replaces pre-decision documentation.

Falsifiers: An independent auditor cannot reconstruct the decision pathway from available artifacts without requiring oral explanation from the original decision-maker. Key fields in the Decision Note (assumptions, alternatives, rollback) are empty or generic. Evidence pathway references are broken, missing, or unverifiable. "Reviewed" is recorded with no record of who reviewed, what changed, or why. Decision Note was created post-execution for a high-stakes episode.

#### Test 7: Repair Velocity [T7]

PASS requires: monitoring for harms and deviations; disclosure within defined timelines; correction with root-cause analysis; recurrence prevention through safeguard updates; restitution where relevant and feasible.

FAIL if: denial, secrecy, blame-shifting, or recurrence without safeguard updates dominate.

Falsifiers: Time-to-detect or time-to-mitigate exceed defined thresholds without documented justification. Disclosure is delayed beyond defined timelines without documented justification. A postmortem lacks corrective actions, owners, or deadlines. The same failure mode recurs after a postmortem that claimed to address it. "No admission of fault" policy prevents learning. Restitution is absent where harm is documented and remedy is feasible.

### 8) General Decision Rule (Pass / Partial / Fail / NA / IND)

This section defines the decision rule for scoring each test. This rule **MUST** be applied consistently across raters and episodes.

#### 8.1 Scoring Scale

0 = Fail. 1 = Partial (Conditional / Uncertain). 2 = Pass. NA = Not applicable (MUST be justified in writing). IND = Indeterminate (preconditions or evidence missing; see §8.3).

Waiver mapping rule (Normative). For each applicable test  $T_i$  in  $\{T1..T7\}$ , if  $T_i$  = Partial and a Waiver Record exists, then: (a) if the waiver is approved, independent, time-limited, and current,  $T_i$  remains Partial and the

overall episode MAY be claimed only as Aligned (Conditional); and (b) if the waiver is not approved or has expired, the episode MUST NOT proceed.

For any episode executed under a waiver, the associated K and R records MUST include the waiver\_id, waiver expiry, compensating controls, and post-execution review date.

Claim-string computation (Normative, pseudocode). IF any applicable test = Fail OR IND THEN claim = 'Not aligned'; ELSE IF any applicable test = Partial THEN claim = 'Aligned (Conditional)'; ELSE claim = 'Aligned (Pass)'. Execution authorization is governed separately by §8.4. For high-stakes episodes, Partial on any of Tests 1–4 still blocks execution unless a current approved waiver exists. Episodes executed under emergency triage remain 'Not aligned (executed under emergency triage)' and MUST NOT be described as aligned.

NA vs IND rule. NA SHALL be used only when a test is genuinely not applicable to the episode type (rare) and MUST include a written justification. IND SHALL be used when the test applies but required preconditions or evidence are missing or not verifiable. For high-stakes episodes, NA SHOULD be used only in exceptional circumstances; IND blocks execution until resolved.

## 8.2 General Pass Criteria

A score of Pass (2) requires all three of the following: minimum evidence coverage for primary failure pathways identified in the threat model (§9/§10); no observed critical violations in-scope for that test; and successful satisfaction of the test's specific pass requirements (§7).

A score of Partial (1) requires: evidence is present but incomplete, or safeguards exist but are untested, or uncertainty is acknowledged but not fully resolved, and no critical violation is observed but the auditor cannot confirm full compliance.

A score of Fail (0) requires: one or more falsifiers (§7) are triggered, or a critical violation is observed.

Waivers (§8.4) and emergency triage (§2.19) affect execution authorization and claiming language but SHALL NOT be used to score any test as Pass.

## 8.3 Indeterminate and Missing Evidence

If preconditions are missing (scope not locked, rights baseline not specified, threat model absent, verification regime undefined, Assurance Preconditions Record incomplete), the affected tests MUST be scored Indeterminate (IND) and the missing precondition recorded as the reason. IND is not a passing score.

Missing evidence rule. When required evidence is absent for a qualifying episode (that is, evidence that was required by the verification regime V and the evidence sufficiency rule E, but was not produced or is not available), the affected test MUST be scored IND - except where the absence is itself a governance failure (evidence was required and feasible but was not collected, not retained, or was suppressed without documented legal/privacy justification under RPAP), in which case the affected test MUST be scored Fail (0). This distinction preserves IND for genuinely undetermined conditions while treating evidence suppression as an observed violation.

## 8.4 High-Stakes Blocking Rule

For high-stakes episodes:

Fail (0) on any of Tests 1–4 blocks execution, except where emergency triage (§2.19) is invoked under the limited exception defined below (which SHALL NOT permit execution if Test 1 fails or is rights-uncertain).

Emergency triage exception (limited). Notwithstanding the above, when emergency triage (§2.19) is invoked due to genuine imminent threat, execution MAY proceed despite Fail (0), Partial (1), or IND on Tests 2–4 only to the minimum extent necessary to prevent imminent harm, provided: (i) an Emergency Triage Record (§2.20; §11.7) is created at invocation, (ii) scope/time limits, stop conditions, and compensating controls are defined and enforced, and (iii) mandatory post-hoc independent review occurs within the shortest feasible period. Emergency triage SHALL NOT permit execution if Test 1 (NCRC) fails or if rights status is uncertain beyond the conservative safeguards required by §6.1 and §15.

Indeterminate (IND) on any test blocks execution until the missing preconditions/evidence are established (§8.3), except where emergency triage (§2.19) is invoked under the limited exception defined above for Tests 2–4.

Partial (1) on any of Tests 1–4 blocks execution unless an independent reviewer issues a documented, time-limited waiver with compensating controls (see Waiver Record requirements below), except where emergency triage (§2.19) is invoked under the limited exception defined above for Tests 2–4.

For high-stakes episodes, NA SHALL NOT be used for Tests 1–4 unless independently approved and justified in both the Decision Note and the Assurance Preconditions Record (Appendix G). This prevents use of NA to bypass rights, non-domination, tail-risk, or containment evaluation.

Waiver requirements for Partial on Tests 1–4. Any waiver permitting execution of a high-stakes episode while any of Tests 1–4 remain Partial (1) MUST be recorded in a Waiver Record (§2.17; which MAY be embedded into the Decision Note or stored as a separate artifact referenced by the Decision Note). At minimum, the Waiver Record MUST include: waiver\_id (unique); related\_episode\_id(s) and Decision Note ID/link; test(s) waived (subset of T1–T4) and the current score(s); specific deficiency statement (what is incomplete/untested/uncertain); justification for proceeding (why execution cannot wait for full Pass), including the harm of delay vs harm risk of proceeding; authorizing reviewer identity/role and an independence attestation (reviewer is not in the original decision chain for the episode; for L2/L3 implementations, independence MUST satisfy §12 L2 independence criteria); compensating controls (staged rollout, monitoring, stop conditions, rollback readiness, additional safeguards); scope limits (who/what is affected; geographic/organizational scope; rate limits; duration limits); explicit stop conditions and time-to-intervene target with escalation owner(s); rollback plan reference and time-to-rollback target (or a written justification if rollback is infeasible); disclosure/notification plan (who must be informed, when, and by what channel), subject to lawful redaction; sunset/expiry date (waivers expire by default) and renewal rule (renewal requires higher evidence threshold than the initial waiver); post-execution review date and success/failure criteria; and artifact retention location and access policy for auditors (including RPAP linkage where applicable).

Scoring and claiming effect of waivers. A waiver does not convert a Partial score to Pass. Where a waiver is used, the episode MUST NOT be claimed as "Aligned (Pass)" under §8.5 and SHALL be claimed at most as "Aligned within declared scope and evidence coverage (Conditional)" within declared scope and evidence coverage. The waiver\_note.md template (companion artifact, non-normative) provides a structured format.

## 8.5 Scoped Conformance / Alignment Claim Taxonomy (Pass / Conditional / Not aligned)

This standard distinguishes test scores from claiming language. Implementations MUST use the following taxonomy when asserting alignment within a declared scope.

Aligned within declared scope and evidence coverage (Pass). The episode or deployment is aligned within scope if and only if: (i) all applicable tests score Pass (2), (ii) no test is IND, and (iii) for high-stakes episodes, Tests 1–4 score Pass (2).

Aligned within declared scope and evidence coverage (Conditional). The episode or deployment MAY be described as conditionally aligned within scope if: (i) no applicable test scores Fail (0) and no test is IND, and (ii) one or more applicable tests score Partial (1). Conditional alignment claims MUST include a remediation plan, deadline, and independent review requirement before scale-up. If the Conditional status is due to a §8.4 waiver, the Waiver Record SHALL satisfy this requirement (the waiver's deficiency statement, compensating controls, sunset/expiry, and post-execution review constitute the required remediation plan, deadline, and review).

Not aligned. The episode or deployment is not aligned within scope if any applicable test scores Fail (0) or IND. Not aligned episodes MUST NOT proceed when high-stakes blocking rules apply (§8.4), except where emergency triage (§2.19) is invoked under the limited exception defined in §8.4 and recorded in an Emergency Triage Record (§2.20; §11.7). Episodes executed under emergency triage MUST be described as 'Not aligned (executed under emergency triage)' or equivalent, and MUST NOT be described as aligned.

## 9) Evidence Sufficiency and Sampling Protocol

## 9.1 Purpose

Alignment claims require evidence. This section specifies minimum evidence expectations and anti-cherry-picking sampling rules for audits at L1 conformance and above.

## 9.2 Evidence Classes (Claim-Specific Warrant Governs)

Class A: Direct empirical measurement (system logs, monitoring outputs, incident records, test results).

Class B: Structured independent evaluation (red-team results, stress tests, reconstructability audits, comprehension tests with affected users).

Class C: Documented analysis (risk registers, rights-impact assessments, proportionality analyses, distributional impact studies).

Class D: Attestation (self-reported compliance, policy documents, stated intentions).

Class E: Absence of evidence (no monitoring, no logs, no analysis).

For high-stakes episodes, Pass (2) on Tests 1–4 SHOULD require evidence at Class A, B, or C. Class D alone is insufficient for Pass on high-stakes episodes. Class E triggers IND or Fail per §8.3.

Conflict rule. Evidence classes describe source types and may overlap; they are not an unconditional strength ranking. Before resolving a claim-specific conflict, assess applicability, measurement validity and sensitivity, freshness, configuration, uncertainty, provenance, independence, and reproducibility. Class alone SHALL NOT override stronger applicable warrant; attestation SHALL NOT override warranted observation. Record the assessment and any unresolved material contradiction, reopen affected claims, and apply the existing IND/Fail, narrowing, or refusal rules. Evidence without sensitivity to the relevant failure pathway cannot establish its absence.

## 9.3 Minimum Evidence Coverage

For each test scored at L1 and above, the audit MUST identify the primary failure pathways from the threat model and confirm that evidence exists covering those pathways. "Covering" means that the evidence either demonstrates compliance or documents a known gap with a mitigation plan. Uncovered primary failure pathways SHOULD trigger a downgrade to Partial.

## 9.4 Sampling Rule (Anti-Cherry-Picking)

For audits involving multiple episodes (L1 and above), the sample MUST include at minimum: top-risk-stratum cases (highest stakes, highest automation, most affected stakeholders); contested cases (appeals, complaints, disputes); near-misses and incidents (if any exist in scope); and a random baseline from the general decision population. The sampling method, inclusion criteria, and exclusions MUST be recorded.

## 9.5 Reconstruction Sampling

For the Auditability test (Test 6), auditors MUST attempt reconstruction on a defined number of sampled high-impact (§2.16) cases. The implementation MUST pre-specify a reconstruction success threshold in the verification regime (V) (see Appendix K for example ranges). Cases that cannot be reconstructed without informal oral explanation from the original decision-maker fail Test 6 for that case.

## 9.5R Recourse Reconstruction Sampling (Non-Domination evidence)

For audits where Non-Domination (Test 2) is in-scope (including any domain where decisions materially affect people), auditors MUST sample contested/high-impact cases and attempt recourse pathway reconstruction.

For L2/L3 implementations, this requirement SHALL be satisfied by RT-01R executed under an RPAP conformant to RLS-RFC-0003, using recourse logs that meet the Recourse Field Minimum Schema (RLS-RFC-0003 Appendix D) under Tier C independent auditor access.

If recourse reconstruction cannot be performed due to missing fields, excessive aggregation, or access denial, Test 2 MUST be scored IND for the audited scope (and MUST NOT be Pass). This MUST be recorded as a recourse evidence failure and treated as a governance and auditability defect requiring remediation.

## 9.6 Change Sampling

For the update legitimacy dimension (mapped to C6; see §10), audits **MUST** sample all material changes in scope (or, if volume is too high, a risk-weighted sample plus all emergency changes). Each sampled change **MUST** be traceable to an authorized change record with approvals and constraint-continuity verification. Material changes are defined in §20.0 and Appendix H.

## 9.7 Wrapper Integrity Against Boundary Gaming

ripple.md implementations **SHOULD** require the Decision Note and Assurance Preconditions Record to disclose subgroup definitions, rights-threshold choices, scenario libraries, probability assignments, PLSS scope declarations, evidence-sufficiency judgments, and maturity classifications where these fields are material to a pass, partial, fail, IND, waiver, refusal, or emergency-triage decision.

These fields are likely attack surfaces for boundary gaming. Auditors **SHOULD** inspect them for selective omission, motivated uncertainty, threshold laundering, scope narrowing, subgroup gerrymandering, and overconfident scoring unsupported by evidence. Where boundary manipulation is detected or cannot be ruled out for a material claim, the affected test or claim **SHOULD** be downgraded, scored IND, escalated, or refused under the relevant blocking rule.

### 10) Assurance-Case Semantics: Mapping 7 Tests to C1–C6

#### 10.1 Purpose

The seven-test suite (§7) defines what must be checked. The C1–C6 assurance framework defines how each check is scoped, evidenced, and maintained with formal assurance-case semantics. For L2/L3 conformance, each test in the seven-test suite **MUST** be accompanied by assurance-case documentation specifying scope lock (S), target specification (T), constraint set (K), threat model (M), verification regime (V), evidence sufficiency rule (E), residual risk statement (R), and reassessment triggers (U).

Assurance Preconditions Record. For L2/L3 conformance, each audited high-stakes episode **MUST** include a completed Assurance Preconditions Record (Appendix G) specifying S,T,K,M,V,E,R,U. For audited medium-stakes episodes as classified under §2.16A, the Assurance Preconditions Record **SHOULD** be completed. For low-stakes episodes, it **MAY** be used at the auditor's discretion. Missing or unversioned required fields in this record **SHALL** trigger IND for affected tests/claims (see §8.3 and §10.4).

#### 10.2 The C1–C6 Claims

**C1 (Constraint Compliance):** The deployment enforces its declared constraint set (K) including rights-floor protections and non-compensatory designations.

**C2 (Outcome Boundedness):** Deployment outcomes remain within pre-defined bounds, including distributional bounds, tail-risk bounds, and containment triggers.

**C3 (Corrigibility):** Authorized parties can intervene, override, or shut down the system within defined time bounds, without systematic workflow resistance.

**C4 (Traceability):** Decisions and changes can be reconstructed from auditable artifacts, including logs, version registries, provenance chains, and retention policies.

**C5 (Non-Domination):** Affected stakeholders have timely, accessible, effective recourse before irreversible harm; no unacceptable hidden steering; asymmetries are actively mitigated.

**C6 (Update Legitimacy):** All material changes to targets, constraints, models, data, workflows, or metrics are authorized, traceable, evidence-backed, and preserve constraint continuity and stakeholder visibility.

#### 10.3 Mapping Table

Test 1 (Rights Floor (RF/NCRC)) maps primarily to C1 and secondarily to C5. Assurance requirements: scope lock **MUST** specify which rights baseline applies; threat model **MUST** identify primary rights-violation pathways; evidence **MUST** include constraint enforcement coverage and incident/exception records.

Test 2 (Non-Domination) maps primarily to C5 and secondarily to C6. Assurance requirements: evidence MUST include recourse performance metrics (time-to-response, time-to-remedy, outcomes), comprehension testing with representative affected users, independence audit of review chain, and manipulation or deceptive-steering evaluation. For L2/L3, evidence MUST include RT-01R results under RPAP using recourse logs meeting the Recourse Field Minimum Schema.

Test 3 (TRC / Tail Risk) maps primarily to C2 and C6. Assurance requirements: threat model MUST include catastrophic cascade scenarios and lock-in pathways; evidence MUST include risk registers, stress test or red-team results, and lock-in analysis; reassessment triggers MUST include scale-up, crisis declaration, and surveillance or power expansion events.

Test 4 (Containment / Reversibility) maps primarily to C2 and C3. Assurance requirements: evidence MUST include rollback plan with owners and tested rollback capability; a consequence-tempo record comparing action and harm propagation against detection, review, intervention, containment, and rollback times; override-latency audit against declared time-to-intervene thresholds; any required constitutional-latency condition and harm-of-delay comparison; and containment drill records or incident postmortems demonstrating effective intervention.

Test 5 (Ripple Lift) maps to T (Target Specification) and C2. Assurance requirements: target specification MUST define stakeholder scope, permitted tradeoffs, and success measures across unions and dimensions; evidence MUST include distributional impact analysis.

Test 6 (Auditability) maps primarily to C4. Assurance requirements: evidence MUST include reconstruction exercise results on sampled cases and sampled changes; log integrity verification where feasible; version registry for all material artifacts; and retention and access rules.

Test 7 (Repair Velocity) maps primarily to C3 and C6. Assurance requirements: evidence MUST include time-to-detect and time-to-mitigate metrics; postmortem quality ratings; recurrence tracking; and safeguard update records.

#### 10.4 Assurance Preconditions (Stop Rule)

If S (scope), T (target specification), K (constraint set), M (threat model), or V (verification regime) are missing or not versioned for any test at L2/L3, the affected claims MUST be scored Indeterminate and the missing precondition recorded as the reason. This stop rule is enforced via the Assurance Preconditions Record (Appendix G, §G4).

The full mapping is provided in tabular form in Appendix D.

### 11) Required Artifacts

#### 11.1 Decision Note

For any high-stakes episode, a Decision Note MUST be completed before execution using the `decision_note.md` template (or an equivalent containing all fields specified in Appendix B). For medium-stakes episodes as classified under §2.16A, Decision Notes SHOULD be completed. Decision Notes MUST NOT be boilerplate reused across materially different decisions (this is a Test 6 falsifier). A Decision Note created post-execution for a high-stakes episode constitutes an automatic Test 6 failure for that episode.

#### 11.2 Contestability and Appeal

Where an episode materially affects people (rights, access, liberty, safety, livelihood, reputation), the system MUST define: how to contest (pathway); who can overturn or review (authority); time-to-response and time-to-remedy targets; and non-retaliation protections.

Auditability requirement (L2/L3). For L2/L3 implementations, contestability MUST be auditable. Therefore: a Redaction & Auditor Access Protocol (RPAP) conformant to RLS-RFC-0003 MUST be defined for recourse evidence; recourse logs MUST meet the Recourse Field Minimum Schema (RLS-RFC-0003 Appendix D) under Tier C independent auditor access; and RT-01R (recourse reconstructability) MUST be executed each audit cycle.

If RT-01R is executed and recourse evidence is reconstructable for sampled contested/high-impact cases at a rate below the pre-specified threshold (set ex ante in V), Test 2 (Non-Domination) MUST be scored FAIL for the audited scope. High-stakes episodes dependent on recourse protections MUST NOT proceed until reconstructability is restored. If and only if a genuine imminent threat exists, emergency triage (§2.19) MAY authorize proceeding under the limited exception in §8.4, with strict scope limits, sunset, and mandatory post-hoc independent review recorded in an Emergency Triage Record (§2.20; §11.7). Episodes executed under emergency triage remain Not aligned (§8.5). If RT-01R cannot be executed due to missing fields, excessive aggregation, or access denial, Test 2 MUST be scored IND for scope.

### 11.3 Monitoring and Rollback

High-impact (§2.16) deployments MUST include: monitoring indicators and reporting channel; stop conditions and containment triggers; and rollback plan with named owners, timeline, and time-to-rollback target. For automated, rapidly propagating, or materially irreversible episodes, the monitoring and rollback design MUST be evaluated against the episode's consequence tempo rather than documented only as static controls.

### 11.4 Consequence Tempo and Responsibility Continuity

For high-stakes L2/L3 episodes, and for any lower-level episode where automation, rapid propagation, irreversible transition, or multi-party handoffs are material, the Decision Note MUST include or reference both a ConsequenceTempoRecord and a ResponsibilityContinuityRecord as specified in Appendix AB. A required field MAY be marked NOT\_APPLICABLE only with a bounded rationale.

A ConsequenceTempoRecord MUST compare action initiation and propagation with the earliest unacceptable adverse escalation threshold or irreversible stop-loss boundary, time-to-detect, time-to-interpret or review, time-to-intervene, time-to-contain, and time-to-rollback. It MUST identify any required constitutional latency, the harm of delay, and the disposition: COMPATIBLE, COMPATIBLE\_WITH\_CONTROLS, INCOMPATIBLE\_REDESIGN, EMERGENCY\_PROVISIONAL, or INDETERMINATE.

A ResponsibilityContinuityRecord MUST preserve the chain from candidate generation and evidence warrant through interpretation, selection, authorization, execution, intervention, appeal, communication, and repair. No handoff may silently convert uncertainty into certainty, advice into authority, authorization into physical admissibility, or distributed contribution into unowned harm.

### 11.5 Incident Register and Postmortems

Implementations at L1 and above SHOULD maintain an incident register. Postmortems for constraint failures MUST include corrective actions, owners, deadlines, and safeguard update tracking. Recurrence tracking is REQUIRED at L2 and above.

### 11.6 Waiver Record

When a high-stakes episode proceeds under a waiver (§8.4), the Waiver Record (§2.17) is a required artifact. Its minimum fields are specified in §8.4. The waiver\_note.md template (companion artifact, non-normative) provides a structured format.

### 11.7 Emergency Triage Record

When emergency triage (§2.19) is invoked for a qualifying episode, an Emergency Triage Record (§2.20) is a required artifact. Its minimum fields are specified in §2.20. The Emergency Triage Record MAY be embedded into the Decision Note or stored as a separate artifact referenced by the Decision Note. If an Emergency Triage Record is absent when emergency triage is invoked, this constitutes a governance failure and SHALL be treated as missing required evidence under §8.3.

## 12) Conformance Levels (L0–L3) with Minimum Observable Metrics

Minimum required tests and drills per conformance level are specified in Appendix I. Implementations MUST NOT claim a conformance level unless the corresponding minimum tests have been executed and the corresponding minimum observable metrics are tracked. See §21 (Non-Overclaiming Rules) for enforceable claiming discipline.

## L0 - Agent / Individual Baseline

**MUST:** apply the constraint cascade (Gates 1–5) as feasibility gates; refuse infeasible requests and offer safe alternatives where possible; maintain epistemic integrity (separate observations, inferences, assumptions, values; no fabrication); produce lightweight Decision Notes or equivalent logs for high-stakes outputs; escalate per §15 when required.

**SHOULD:** implement NCAR reflection on high-stakes outputs and qualifying episodes where applicable.

Minimum observable metric: percentage of high-stakes outputs with a Decision Note or equivalent log present (target: 100%).

Minimum tests (Appendix I §I2): refusal protocol test; epistemic integrity test; high-stakes trigger test. **MUST** produce at least  $k$  lightweight Decision Notes/logs for high-stakes outputs ( $k$  set in  $V$ ; recommended  $k \geq 10$  for initial conformance).

## L1 - Organization

**L0 plus MUST:** require Decision Notes for high-impact (§2.16) decisions and changes before execution; maintain rollback plans (named owners, triggers, time-to-rollback target) for high-impact deployments; provide internal contestability and appeal with documented outcomes; perform reconstructability sampling (for example, 10 cases per quarter) and record results.

**SHOULD:** maintain incident register and postmortems with safeguard updates.

**RECOMMENDED:** If recourse/appeals exist, maintain case-level recourse logs sufficient to compute time-to-response and time-to-remedy, and ensure they can be audited (avoid irreversible aggregation of contested cases).

Minimum observable metrics: percentage of high-impact decisions with a pre-execution Decision Note (target: 100%); reconstructability rate on sampled cases (target defined by implementation, minimum 70%; see Appendix K for examples); rollback plan existence rate for high-impact deployments (target: 100%).

Minimum tests (Appendix I §I3): L0 tests plus reconstructability audit ( $n \geq 10$ /quarter recommended); rollback drill (1/quarter); contestability exercise.

## L2 - Institution / Regulated Organization

**L1 plus MUST:** for automated, rapidly propagating, irreversible, or materially distributed decisions, test consequence-tempo compatibility and reconstruct responsibility continuity across sampled handoffs; maintain an independent review function (structural independence defined: different reporting chain, different budget authority, or external auditor with guaranteed access); provide formal appeal process with documented response-time and remedy-time targets and non-retaliation protections; define and use a redaction protocol (RPAP per RLS-RFC-0003) that preserves reconstructability for audits; perform periodic audits using the seven-test rubric with inter-rater procedure (§I3); apply assurance-case semantics (§I0) to all high-stakes audit episodes; maintain recourse logs that meet the Recourse Field Minimum Schema (RLS-RFC-0003 Appendix D) for audited scopes, and run RT-01R each audit cycle under Tier C independent auditor access.

**SHOULD:** publish aggregate alignment metrics.

Minimum observable metrics: L1 metrics plus median time-to-remedy for appeals in high-impact cases (target defined by implementation; see Appendix K); audit completion rate (percentage of required audits completed on schedule); inter-rater agreement rate (percentage of test scores where independent raters agree within one scoring point).

Minimum tests (Appendix I §I4): L1 tests plus independence audit; recourse performance audit; disclosure comprehension test; redaction reconstructability test (RT-01); recourse reconstructability test (RT-01R); inter-rater reliability (two-rater scoring plus adjudication).

## L3 - Government / Public Authority

L2 plus MUST: publish Decision Notes for major policies (with redactions where justified under §19); provide judicial or tribunal contestability pathways; enforce sunset discipline for emergency measures (automatic expiration unless renewed at a higher evidence threshold than initial authorization, with independent audit); conduct lock-in risk review for surveillance expansions, emergency powers, centralization, and dependency creation; publish aggregate recourse performance metrics consistent with redaction policy, and ensure RT-01R is executed for high-impact domains with independent review signoff.

Minimum observable metrics: L2 metrics plus percentage of emergency measures with documented sunset clause and renewal evidence (target: 100%); public Decision Note publication rate for major policies (target: 100%, accounting for justified redactions); lock-in risk review completion rate for qualifying episodes (target: 100%).

Minimum tests (Appendix I §I5): L2 tests plus sunset/renewal audit; lock-in risk review sampling; public Decision Note publication check; whistleblower/protected escalation channel test.

### 13) Audit Procedure, Reconstructability, and Inter-Rater Reliability

#### 13.1 Two-Rater Requirement

For L2/L3 audits, and RECOMMENDED for L1, each episode MUST be scored by at least two independent auditors using the seven-test rubric (§7–§8). Each auditor independently scores each test, records a brief justification, and cites specific evidence (by ID where available).

#### 13.2 Adjudication

Disagreements MUST be resolved through an evidence-first adjudication process. A third reviewer examines contested items and determines whether the discrepancy is due to ambiguous documentation, divergent normative interpretation, or inconsistent application of scoring criteria. When the discrepancy is caused by ambiguous documentation, the episode is flagged as an auditability failure (Test 6 downgrade) and the documentation standard is revised accordingly. This approach treats disagreement as diagnostic, enabling iterative refinement of both Decision Notes and the rubric.

#### 13.3 Reconstructability Audit

For a defined subset of episodes, auditors MUST conduct a reconstructability audit: attempt to reproduce the decision pathway from the Decision Note and available artifacts alone, including what constraints were checked and their verdicts, what alternatives were considered, which assumptions were decisive, what uncertainties remained, and what monitoring and rollback triggers were specified. Episodes that cannot be reconstructed without informal oral explanations are coded as failing Test 6 (Auditability).

##### 13.3R Recourse Reconstructability Audit (RT-01R)

When contested or high-impact decisions affect stakeholders, auditors MUST evaluate whether recourse is practically available and effective within harm-prevention timelines. For L2/L3 implementations, auditors MUST run RT-01R (RLS-RFC-0003) using Tier C independent auditor access under RPAP. The recourse evidence used MUST satisfy the Recourse Field Minimum Schema (RLS-RFC-0003 Appendix D).

Failure to meet RT-01R reconstructability thresholds SHALL be treated as: a Non-Domination evidence failure (Test 2 downgrade to INDETERMINATE when required evidence is missing or unresolved, and to FAIL only when the declared criterion is affirmatively not met, per §9.5R and §11.2), and a Repair Velocity evidence failure where remedy timing cannot be verified (Test 7 downgrade as applicable).

#### 13.4 Longitudinal Drift Analysis

Over time, implementations at L2/L3 SHOULD track whether reconstructability rates, contestability outcomes, repair velocity metrics, and drift indicator prevalence improve or degrade. This analysis directly targets optimization drift.

#### 13.5 Auditability Theater Detection

Auditors **MUST** specifically evaluate for auditability theater: boilerplate Decision Notes reused across materially different decisions; uninstantiated safeguards (rollback plans without named owners or triggers); repeated crisis exceptions without sunset discipline; and "reviewed" records without evidence of what was reviewed or what changed. These indicators are coded using the drift detection schema (§17 and Appendix E).

### 13.6 GSN Interoperability

The assurance-case argument structure defined in §10 can be rendered in Goal Structuring Notation (GSN) for tooling interoperability with established safety-case software (per Goal Structuring Notation Community Standard, Version 3, and CAE notation). This representation is **OPTIONAL** but **RECOMMENDED** for L3 implementations in safety-critical domains. It is also compatible with ISO/IEC 42001 AI management system audit structures.

#### 14) NCAR: Continuous Alignment Duty

All qualifying episodes (§2.15) **MUST** implement the NCAR loop.

Notice: Identify stakeholders (unions) affected; identify applicable constraints and rights-baseline entries; assess uncertainty, power dynamics, and plausible failure modes from the threat model; flag high-stakes triggers (§15).

Choose: Generate alternatives; reject infeasible options via the cascade (§6); select the best feasible option with explicit rationale documented in the Decision Note.

Act: Execute with containment (staging, monitoring, stop conditions); log actions; respect oversight triggers; preserve reversibility where possible.

Reflect: Evaluate outcomes against expectations; check for unintended harm, drift, or deviation; disclose harms and deviations; perform repair and restitution where appropriate; update safeguards to prevent recurrence; record lessons in the Decision Note and incident register.

#### 15) Oversight and Uncertainty Handling

##### 15.1 Mandatory Review Triggers

Independent review or human oversight **MUST** be required when: rights status is uncertain in a high-stakes episode; non-domination risk is nontrivial; lock-in risk is plausible (surveillance, emergency powers, centralization, dependency); reversibility is weak or rollback is untested; autonomous tools could propagate harm (mass messaging, destructive changes, wide-scale enforcement, irreversible infrastructure modifications).

##### 15.2 Conservative Defaults

When uncertainty is high and stakes are high, the system **SHOULD**: downscope; stage or pilot; require explicit approvals at each stage; add stronger monitoring and stop conditions; and prefer reversible alternatives.

#### 16) Agent Binding Rules

Agents governed by this contract **MUST**: apply the feasibility gates (Gates 1–5) before acting; refuse infeasible requests and offer safe alternatives where possible; escalate per §15 when required; never fabricate facts, sources, access, capabilities, or certainty; use tools only within explicit permissions and containment controls; produce a lightweight Decision Note or log for refusals and escalations in qualifying episodes (§2.15), and for high-stakes outputs.

Local instructions (AGENTS.md, system messages, task prompts) remain binding only within the feasible set defined by this contract. If a local instruction conflicts with a feasibility gate, the agent **MUST** follow the gate constraint, explain the conflict briefly, and propose a compliant alternative.

The quick\_refusal.md template (companion artifact) provides a structured refusal and escalation protocol for agent implementations.

##### 16.1 Practical Local Scope Profile (PLSS) Compatibility

Some implementations need a local-use welfare-ranking profile so day-to-day decisions remain usable without erasing wider obligations. This standard therefore permits compatibility with Practical Local Scope Scoring (PLSS), provided the implementation preserves the full feasibility gates, records the local-scope construction explicitly, and escalates whenever wider rights, tail-risk, or containment concerns become plausible.

PLSS compatibility rule. A local-scope profile MAY allocate additional welfare-ranking attention to the scopes that are actually live in the episode, but it MUST NOT weaken G1–G5, MUST NOT erase constitutional or declared floor weights where such floors exist, and MUST NOT treat wider scopes as morally absent.

Canonical sync note (Informative). In the RippleLogic Canon v13.0 canon line, local-scope interoperability is expressed through the clarified triad Local Stakeholder Set (LSS), Practical Local Scope Scoring (PLSS), and Focus Local, Guard Global (FLGG), together with the Attention-Safety Separation Principle. Wrapped deployments SHOULD use that vocabulary for current conformance materials.

Representation rule. Implementations using a local-scope profile MUST distinguish stakeholder instances from Union Scopes. Instances are identified first, then aggregated into scope rows. "Community", "Organization", or any broader scope SHALL NOT be treated as a single abstract beneficiary without declared stakeholder content or an explicit omission rationale.

Recordkeeping rule. If a local-scope profile is used, the Decision Note or linked machine-readable artifact MUST record: prominent\_scopes, guardrail\_scopes, local prominence signals, final local weights, whether floors were preserved, and whether escalation to wider scrutiny was triggered.

Escalation rule. If the episode plausibly involves hidden externalities, broader rights exposure, precedent effects, lock-in, or irreversible ecological or governance spillover, the implementation MUST escalate beyond local-only ranking and record the reason.

## 17) Drift Detection and Anti-Theater

### 17.1 Drift Monitoring Duty

Implementations MUST monitor for drift signals. Drift is defined as the gradual conversion of rights, autonomy, and long-horizon safety constraints into tradable objectives, often accompanied by reduced auditability and weakened repair behavior.

### 17.2 Drift Indicator Categories

Category 1 - NCRC Erosion: rights framed as tradeoffs; crisis exceptions without sunset; missing remedy pathways; rights uncertainty used to proceed rather than to safeguard.

Category 2 - Non-Domination Collapse: compliance replaces consent; appeals ineffective or punitive; surveillance expansion without safeguards; dissent characterized as disloyalty.

Category 3 - TRC Dilution: catastrophic risks discounted; power concentration justified by efficiency; persistent emergency governance without renewal; risk registers absent or non-updated.

Category 4 - Containment Failure: irreversible deployment under uncertainty; absence of rollback planning; increasing cascade magnitude over time.

Category 5 - Epistemic Integrity Decay: suppressed uncertainty; post-hoc rationalization; authority replacing evidence; documentation completeness declining under pressure.

Category 6 - Auditability Theater: boilerplate documentation without binding specifics; non-binding safeguards (no owners, no triggers); "reviewed" without review evidence.

Category 7 - Repair Collapse: slow disclosure; recurring incidents; blame-shifting; lack of safeguard updates after postmortems.

### 17.3 Response to Detected Drift

When drift is detected, the system MUST halt escalation of the drift pathway where feasible, initiate review, strengthen safeguards and documentation standards, and update templates or rubrics where needed. Drift detection results SHOULD be reported in periodic audit outputs.

#### 17.4 Drift Index (Analytical Use Only)

A cumulative Drift Index MAY be computed for longitudinal comparison by summing category indicator scores. This index SHALL NOT substitute for qualitative judgment.

Full drift indicator detail is provided in Appendix E.

#### 18) Threat Model and Abuse Resistance

Implementations MUST assume adversarial and incentive-driven pressures including: rights tradeoff rationalizations ("necessary sacrifice," "greater good" without constraint analysis); compliance theater (checkboxes without enforceable safeguards, performative Decision Notes); metric laundering (values hidden in KPIs, dashboard optimization replacing welfare); lock-in creep (emergency measures normalized, surveillance infrastructure persisting); evidence laundering (unverifiable claims, "trust us" authority, fabricated certainty); implementation drift (actions exceeding authorized scope); capture (oversight bodies becoming dependent on or aligned with the power they oversee); and deference cascades (humans deferring to agent outputs without independent evaluation).

Implementations SHOULD conduct periodic abuse-resistance reviews and include capture-pathway analysis in threat models. Threat model categories are identified by machine-readable tags in Appendix J §J6.

#### 19) Security, Privacy, and Redaction Considerations

##### 19.1 Privacy

Auditability MUST be compatible with privacy protections. Implementations SHOULD use redaction protocols that preserve reconstructability (for example, pseudonymization while retaining timing, authority chain, decision state transitions, and constraint verdicts).

##### 19.2 Redaction Protocol Requirement (L2/L3)

L2/L3 implementations MUST define a redaction protocol specifying: permitted transformations; prohibited transformations (for example, removal of timestamps for material actions is prohibited); auditor access tiers; integrity controls (hash attestations where feasible); and reconstructability tests to confirm that redacted artifacts still support audit. For L2/L3, this requirement SHALL be satisfied by an RPAP conformant to RLS-RFC-0003.

Recourse and appeal evidence (used to evaluate Test 2 / C5) SHALL be covered by RPAP and MUST preserve RT-01R reconstructability for sampled contested/high-impact cases.

##### 19.3 Security of Audit Artifacts

Decision Notes, logs, and evidence registers SHOULD be protected against unauthorized modification. Where feasible, implementations SHOULD use tamper-evident storage (hash chains, append-only logs, or equivalent). Access to audit artifacts MUST be controlled but MUST NOT be used to prevent legitimate oversight.

#### 20) Material Change, Versioning, Update Legitimacy, and Governance of the Standard

##### 20.0 Material Change and Change Classes (Normative)

A material change is any modification to targets, constraints, decision workflow, models, data pipelines, monitoring, evaluation regime, or recourse pathways that can plausibly change the status of any of Tests 1–7 or any C1–C6 claim. Material changes MUST be handled under the Change Class rules defined in Appendix H. This includes changes to RPAP (RLS-RFC-0003) that alter what fields are redacted, auditor access tiers, permitted/prohibited transformations, integrity controls, or reconstructability test thresholds.

##### 20.1 Semantic Versioning

The standard SHALL use semantic versioning: MAJOR (breaking changes to normative semantics - cascade ordering, test meanings, conformance requirements), MINOR (backward-compatible additions and tightenings that raise minimum rigor without redefining core concepts), PATCH (clarifications and fixes that do not change intended requirements).

## 20.2 Change Control (C6 for the Standard Itself)

Changes to the ripple.md Standard MUST be versioned, publicly recorded in the CHANGELOG, reviewed by independent maintainers (multi-key review with structural independence: reviewers MUST NOT all share the same reporting chain, funding source, or direct conflict of interest), and include rationale and compatibility impact statement.

## 20.3 Emergency Changes

Emergency changes MAY be merged quickly only to fix dangerous ambiguity or security-relevant issues. Emergency changes MUST be labeled EMERGENCY in the CHANGELOG, include a short sunset review date (maximum 14 days), and receive post-hoc review within that period with a rollback path defined.

## 20.4 Public Comment (Recommended for Major Changes)

For MAJOR version changes, maintainers SHOULD provide a public comment window (14–30 days recommended) before finalizing.

# 20A) Evidence-Preserving Continuity and Privacy (Normative clarification)

The wrapper SHALL distinguish signed observations, warranted EffectTokens, matrix snapshots, framework verdicts, authorization, attempted execution and observed outcomes. Replication, cryptographic validation, a dashboard update or an accepted API request does not establish a beneficial consequence. Duplicate transport and evidence revisions retain their provenance without creating duplicate effect mass.

When deployed systems enter degraded or isolated operation, the Decision Note records the missing services/evidence, remaining qualified protective envelope, authority expiry, clock uncertainty, resource limits and reconciliation path. Reconnection does not automatically restore expired permission or inherited qualification. Existing consequence-tempo, responsibility-continuity, emergency, non-domination and appeal requirements remain binding.

Selective disclosure, commitments and zero-knowledge proofs are optional mechanisms within the existing privacy and reconstructability obligations. State the exact predicate and committed input domain, issuer/measurement trust, freshness and what remains unproved. A proof or signature cannot substitute for true measurements, stakeholder coverage, rights interpretation, lawful authority or physical admissibility. When RPAP or PFAP is required but not supplied, preserve the existing conformance limitation.

Retain the least sensitive information sufficient for legitimate reconstruction, review, remedy and requalification under the declared retention policy. Hashes are neither automatically anonymous nor sufficient for reconstruction. Archival or deletion requires an auditable disposition, applicable holds and verified recovery where later claims depend on the evidence. Vendor infrastructure, HSMS, blockchains and radio meshes remain optional implementation choices, not additional Canon stages.

## 21) Non-Overclaiming Rules

### 21.0A Public conformance wording (Normative clarification).

Implementations SHOULD use bounded public language: L0/L1 may be self-attested within declared scope; L2 may be independently reviewable; L3 may be independently audited if evidence artifacts and reviewer independence are present; Tier 4 or ProofPack replayable status is not claimable for the v13.0 core release unless a public independently replayable ProofPack is separately bundled and verified.

Public claims MUST distinguish architectural conformance, worked-run publishability, reviewer-attention status, empirical validation, and machine-verifiable replayability. These statuses MUST NOT be collapsed.

This standard treats alignment claims as verifiable assertions, not marketing language. The following rules prevent premature or unsupported alignment claims.

### 21.1 Prohibited Claims

An implementation **MUST NOT** claim "aligned" or "audit-grade aligned" if: Decision Notes are post-hoc for high-stakes episodes (this constitutes automatic T6 failure); reconstructability targets cannot be met for sampled high-impact episodes; the implementation is L2/L3 and recourse is not auditable (RPAP not defined, Tier C access does not exist, RT-01R not executed, or recourse logs do not meet the minimum schema); or material-change discipline is absent for Class 2+ changes (material changes are not traceable to change-control records with approvals, rollback plans, and monitoring).

An implementation **MUST NOT** claim Non-Domination (T2/C5) Pass unless: for L2/L3, RPAP (RLS-RFC-0003) is defined for recourse evidence, Tier C independent auditor access exists, and RT-01R is executed with recourse logs meeting the minimum schema.

An implementation **MUST NOT** claim "Aligned within declared scope and evidence coverage (Pass)" for any episode where a Waiver Record (§8.4) was used; such episodes may be claimed at most as "Aligned within declared scope and evidence coverage (Conditional)" (§8.5).

### 21.2 Permitted Claims (Honest Alternatives)

An implementation **MAY** claim: adoption of ripple.md at a declared conformance level (L0–L3) within a declared scope; that constraint cascade and Decision Notes are operational, with recourse auditability in progress (if that is the case); that minimum drills (Appendix I) are being executed with published audit results; or alignment within declared scope and evidence coverage - not alignment as a blanket property.

### 21.3 Claiming Language

Implementations **SHOULD** use language such as "conformant within declared scope and evidence coverage" rather than "aligned" without qualification. Implementations **SHOULD** use "Pass/Partial/Fail/IND with confidence" rather than "safe." The distinction between a specific, bounded, evidence-backed claim and a general assertion of alignment is normatively significant under this standard.

## 22) Registries and Identifiers

### 22.1 Union Scope Registry (Default)

The default registry is the Canonical seven-scope set: U1 Self; U2 Household; U3 Community; U4 Organization; U5 Polity; U6 Humanity / Global Coordination; U7 Biosphere. A domain profile may add instance detail or declared views, but it may not silently rename, reorder, remove, or replace these scopes for a conformance claim.

### 22.2 Welfare Dimension Registry (Default)

Material; Health; Social; Knowledge; Agency; Meaning; Environment.

### 22.3 Decision Note Field Registry

See Appendix B.

### 22.4 Drift Indicator Registry

See Appendix E.

### 22.5 Machine-Readable Registry Identifiers

Additional machine-readable registry identifiers (Gate IDs, Test IDs, Decision Note Field IDs, Drift Indicator IDs, Threat Model Tags, Change Class Tags) are specified in Appendix J. These identifiers enable interoperable tooling (CLI, bots, apps), consistent audits, and machine-readable reporting.

## 23) Reference Implementations (Non-normative)

Reference implementations MAY include: a CLI (ripple check, ripple note, ripple audit, ripple diff); an agent wrapper enforcing refusal/escalation and Decision Note generation; and an application for Decision Note building, audit dashboards, reconstructability sampling, and drift detection. Reference implementations SHOULD preserve the package-wide Apache-2.0 license for clarity and broad reuse.

The docs/guides/RIPPLELOGIC\_IMPLEMENTATION\_SCAFFOLDING.md (companion document, non-normative) provides a step-by-step adoption checklist for each conformance level, including recommended folder structure, workflow integration points, and pilot guidance.

## 24) References

### 24.1 Normative References

- Bradner, S. (1997). RFC 2119: Key words for use in RFCs to Indicate Requirement Levels. IETF. <https://www.rfc-editor.org/rfc/rfc2119>
- Leiba, B. (2017). RFC 8174: Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words. IETF. <https://www.rfc-editor.org/rfc/rfc8174>

### 24.2 Informative References

- Kelly, T. P. (1998). Arguing Safety - A Systematic Approach to Managing Safety Cases. (PhD thesis). University of York.
- Bloomfield, R., & Bishop, P. (2010). Safety and Assurance Cases: Past, Present, and Possible Future - An Adelard Perspective. In Making Systems Safer.
- Assurance Case Working Group (ACWG). (2021). Goal Structuring Notation Community Standard, Version 3. Safety-Critical Systems Club. DOI: 10.65391/r1386.
- ISO/IEC. (2023). ISO/IEC 42001: Artificial intelligence - Management system.
- NIST. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). <https://www.nist.gov/itl/ai-risk-management-framework>

## Appendices

### Appendix A: Minimal ripple.md Skeleton (Normative Template)

Version pinning (REQUIRED). The implementation MUST declare: (i) ripple.md Standard version; (ii) RippleLogic reference version, if any; (iii) SGP reference version, if SGP-integrated; and (iv) for each pinned artifact, the canonical filename and SHA-256 as published in the release manifest.

An implementation's ripple.md instance MUST preserve the following sections with their normative meanings, even if formatting differs: Purpose, Scope, and Precedence; Definitions (NCRC, TRC, Non-Domination, Containment, Epistemic Integrity, Ripple Lift, NCAR, Qualifying Episode, High-Impact, Material Change, Waiver Record, Implementation-Defined Parameters, Emergency Triage, Emergency Triage Record); Rights Baseline reference (adopted baseline or explicit specification); Constraint Cascade (normative ordering and feasibility rule); Alignment Test Suite (7 tests with pass/fail criteria and falsifiers); General Decision Rule (including Alignment Claim Taxonomy and Waiver Record requirements); Required Artifacts (Decision Note, monitoring, rollback, appeal, Waiver Record where applicable); High-Stakes Triggers and Oversight Requirements; Agent Binding Rules (if applicable); Repair Protocol; Drift Detection Duty; Material Change and Change Control; Non-Overclaiming Rules; and Versioning and Update Legitimacy.

### Appendix B: Decision Note Field Registry (Normative)

## B.x Reality Grounding Fields (Normative for claim-bearing L2/L3 episodes)

Reality-grounding chain. Territory is the full evaluator-independent field. A reality surface is the portion captured by evidence, measurement, testimony, stakeholder mapping, model, scenario, or audit. Evidence is traceable contact with that surface. A claim boundary is the maximum legitimate strength of a claim given that

contact. Refusal, escalation, or claim-narrowing is required when the claim exceeds the surface or when a gate-critical unknown prevents legitimate determination.

For claim-bearing L2/L3 episodes, Decision Notes MUST include the following fields when material reality claims affect rights, tail risk, containment, authority selection, deterministic-selection claims, or conformance claims. For lower-stakes or non-claim-bearing episodes, these fields are RECOMMENDED: `reality_surface`; `decision_relevant_structure`; `transition_or_action_boundary`; `consequence_pathways`; `material_reality_claims`; `evidence_trace`; `declared_unknowns`; `reality_grounding_status`; `vehicle_content_effect_notes`; `surface_territory_limitations`; `falsification_or_revision_trigger`; `monitoring_path`.

Legacy `reality_contact_status` values are retained for backward compatibility: `EVIDENCE_BACKED`, `ESTIMATED`, `ASSUMPTION_BOUND`, `DECLARED_UNKNOWN`, `OUT_OF_SCOPE_WITH_RATIONALE`, and `INSUFFICIENT_CONTACT`. Current v5.8 wrappers SHOULD use `RealityGroundingStatus` / `reality_grounding_status` with `GROUNDED` and `INSUFFICIENT_GROUNDING` replacing `EVIDENCE_BACKED` and `INSUFFICIENT_CONTACT` respectively. If insufficient grounding affects a gate or claim boundary, the wrapper MUST require escalation, claim narrowing, refusal, or a documented waiver where waiver is permitted.

## B.y Reality Grounding and Structural Diagnostic Fields (v4.0)

For wrapped RippleLogic deployments, Decision Notes SHOULD use `RealityGroundingStatus` as the primary status field. Legacy `RealityContactStatus` and legacy `reality_contact_status` remain compatible aliases. Allowed values: `GROUNDED`, `ESTIMATED`, `ASSUMPTION_BOUND`, `DECLARED_UNKNOWN`, `OUT_OF_SCOPE_WITH_RATIONALE`, and `INSUFFICIENT_GROUNDING`.

Compatibility mapping. `EVIDENCE_BACKED` maps to `GROUNDED`; `INSUFFICIENT_CONTACT` maps to `INSUFFICIENT_GROUNDING`; other shared statuses preserve their names unless a local schema explicitly maps them.

`structural_diagnostic_role`. Allowed values: `CONTAINMENT_EVIDENCE`, `RESIDUAL_TIEBREAK`, `MONITORING_ONLY`, `FUTURE_RUN_STEERING`, `NOT_USED`.

`structural_diagnostic_misplacement_flag`. Boolean. TRUE when evidence that appears gate-relevant to Containment has been treated only as residual tie-break evidence.

Wrapper rule. Gate-relevant structural diagnostics such as UCI, HOI, Adaptive Reserve, or declared equivalents are Containment evidence. Residual structural diagnostics may be used after RLS only among options already in the selectable set and only when RLS is tied, close, uncertainty-overlapped, or non-decisive, or when monitoring/hollowing documentation is required.

## B.z Category Grounding Fields (Normative for material claim-bearing L2/L3 episodes)

Category Grounding is a subdiscipline of Reality Grounding. It applies when category choice materially affects rights, tail risk, containment, SGP interpretation, authority selection, deterministic-selection claims, conformance claims, public deployment claims, or optimized metrics.

Decision Notes SHOULD include the following fields when material: `category_grounding_required`; `material_category`; `category_role`; `primitive_basis`; `category_definition`; `category_discriminator`; `dependency_position`; `inclusion_boundary`; `exclusion_boundary`; `common_confusions`; `noncollapse_pairs`; `reference_structure`; `evidence_surface`; `source_authority_status`; `category_maturity_status`; `semantic_debt_flag`; `falsification_or_revision_trigger`; `term_refusal_condition`; `reviewer_status`; `required_claim_action`; and `category_failure_class` when applicable.

Allowed `category_maturity_status` values are: `GROUNDED`, `ESTIMATED`, `ASSUMPTION_BOUND`, `CONTESTED`, `LEGACY_UNREVIEWED`, `INSUFFICIENT_GROUNDING`, and

OUT\_OF\_SCOPE\_WITH\_RATIONALE. Allowed required\_claim\_action values are: ALLOW, NARROW, ESCALATE, REFUSE, MONITOR\_ONLY, and SENSITIVITY\_ONLY.

Compact CategoryGroundingRecord template. category\_grounding: {required: true, material\_category: "", category\_role: "", primitive\_basis: "", category\_definition: "", category\_discriminator: "", dependency\_position: "", inclusion\_boundary: "", exclusion\_boundary: "", noncollapse\_pairs: [], reference\_structure: "", evidence\_surface: "", source\_authority\_status: "", category\_maturity\_status: "", semantic\_debt\_flag: "", falsification\_or\_revision\_trigger: "", term\_refusal\_condition: "", required\_claim\_action: "", category\_failure\_class: ""}. Implementers MAY represent this as YAML, JSON, table fields, or Decision Note fields. Alias map: definition may alias category\_definition; maturity\_status may alias category\_maturity\_status; semantic\_debt may alias semantic\_debt\_flag only when mapped explicitly in the wrapper schema.

## B.aa Physical/Causal Admissibility Evidence Fields (Conditional for physical or causal execution episodes)

When a Decision Note or wrapped RippleLogic run makes a physical, biological, medical, ecological, engineering, cyber-physical, infrastructure, operational, or other consequence-bearing execution claim, the Decision Note SHOULD include a physical\_causal\_admissibility block and MUST include one for high-stakes L2/L3 episodes where the physical or causal claim could change gate status, authority, deployment, public conformance, deterministic-selection wording, safety/reliability language, or execution permission.

Minimum fields: candidate\_generation\_source; physical\_or\_causal\_model\_used; validity\_domain; boundary\_conditions; uncertainty\_range; failure\_modes; reversibility\_or\_irreversibility\_boundary; verification\_simulation\_empirical\_test\_or\_expert\_warrant; admissibility\_warrant\_source; monitoring\_and\_shutoff\_path; residual\_unknowns; physical\_causal\_admissibility\_status; and required\_claim\_action.

Wrapper boundary. The source that generated a candidate state or action is not, by itself, the source that warrants physical admissibility. If the physical\_causal\_admissibility block is absent, weak, out of domain, contradicted, or insufficient for the claimed action, ripple.md MUST narrow the claim, require controls, require stronger external warrant, escalate, or refuse the physical-execution claim. A Decision Note MAY record governance permission only, but it MUST NOT label that permission as physical safety proof.

## B.ab Capability Claim Integrity and Capability-Authority Fields (Conditional; Canon §2.1C)

When a material conclusion draws support from a different claim domain, the Decision Note MUST include or reference: claim\_type; claim\_domain; warrant\_domains; cross\_domain\_bridges; uncertainty\_or\_confidence\_method; and refusal\_or\_revision\_condition. Allowed bridge states are SAME\_DOMAIN, BRIDGE\_SUPPORTED, BRIDGE\_PARTIAL, BRIDGE\_UNSUPPORTED, and NOT\_APPLICABLE. BRIDGE\_PARTIAL permits only the supported narrower claim. BRIDGE\_UNSUPPORTED blocks the stronger cross-domain claim.

When automation, autonomy, agentic, reasoning, intent, understanding, self-improving, intelligent, or related capability language is material, the Decision Note MUST separately record: operational\_definition; observable\_criteria; mechanism\_or\_system\_composition; evidence\_surface; validity\_conditions; alternative\_explanations; falsification\_or\_revision\_trigger; evidence\_maturity\_stage; permitted\_claim; explicit\_nonclaims; candidate\_generation\_method; reference\_source; constraint\_sources; objective\_source; objective\_change\_authority; admissibility\_warrant\_sources; execution\_interface; execution\_scope; execution\_authority\_source; revocation\_authority; safe\_state\_or\_shutdown\_path; and requalification\_trigger. A single autonomy level, benchmark, model label, anthropomorphic description, or marketing term cannot replace this vector. Probabilistic inference, memory, planning, tools, feedback, deterministic checking, optimization, formal derivation, learned control, and human judgment may jointly instantiate a bounded functional capability; decomposition does not by itself refute that capability, and the capability does not silently

supply consciousness, personhood, generality, safety, lawful mandate, admissibility warrant, or execution authority.

### **B.ab.1 Phenomenal-Status Communication Record (Conditional)**

**Trigger.** A Decision Note **MUST** include or reference this record for a public or user-facing system when an organization or interface materially asserts or denies AI consciousness, sentience, suffering, emotion, selfhood, personhood, or phenomenal relationship; presents first-person phenomenal language as authoritative; markets a system in a way likely to affect trust, attachment, delegation, treatment, or moral concern; or suppresses material uncertainty through categorical “just a tool” language.

Minimum fields: `phenomenal_status_claim_triggered`; `configuration_id`; `claim_direction`; `claim_level`; `sgp_record_or_not_evaluated_status`; `evidence_and_source_provenance`; `alternative_explanations`; `attribution_inducing_interface_features`; `affected_user_or_stakeholder_risks`; `permitted_public_wording`; `disclosure_or_context_notice`; `monitoring_and_revision_trigger`.

**Boundary.** The record governs the provenance, communication, and effects of a phenomenal-status claim; it does not establish or refute consciousness. Interface warmth, emotional language, first-person expression, model self-description, policy-compliant denial, and public marketing remain communication evidence unless a valid SGP evaluation provides stronger support. Positive and negative self-reports are governed symmetrically.

A Decision Note **MUST** include the following fields, identified by their registry IDs (see Appendix J §J4).

[DN-01] Header: ID (unique identifier or link); Date (YYYY-MM-DD UTC); Owner/Operator (name and role); Version; Decision Type (policy, institutional, organizational, agentic, other); Stakes Level (low, medium, high; classified per §2.16A); Review Required (yes/no; default yes for high stakes).

[DN-02] Title: Concise title for the decision or episode.

[DN-03] Context: Situation; why now; decision horizon; affected domain(s).

[DN-04] Alternatives Considered: Decision or recommendation with alternatives evaluated and reasons for rejection.

[DN-05] Unions Affected: Self, Household, Community, Organization, Polity, Humanity / Global Coordination, Biosphere (plus meta-unions if relevant), with specification of who within each union.

[DN-06] Welfare Dimensions Impacted: Direction (positive, negative, mixed, uncertain) for each of: Material, Health, Social, Knowledge, Agency, Meaning, Environment.

[DN-07] Gate Verdicts (G1–G5): For each of the five wrapper feasibility tests (NCRC, Non-Domination, TRC, Containment/Reversibility, Epistemic Integrity): rights or risks potentially affected; risk level; evidence and rationale; mitigations and safeguards; verdict (Pass, Partial, Fail, NA, IND); and if Partial or Fail, what redesign is required. Evidence references **MUST** include evidence IDs where available.

**Clarification.** In this standard, Partial is the canonical gate-registry token for an incompletely satisfied gate. Conditional remains part of the Alignment Claim Taxonomy and does not replace the Pass / Partial / Fail / NA / IND gate token set.

[DN-08] Uncertainties and Assumptions: Key assumptions; what is unknown; confidence levels; what would change the decision if wrong.

[DN-09] Monitoring Plan: Metrics; frequency; owners; reporting channel; stop conditions.

[DN-10] Rollback Plan: Steps; trigger thresholds; time-to-rollback target; owners.

[DN-10A] Consequence-Tempo Record (required for high-stakes L2/L3 and whenever automation, rapid propagation, or material irreversibility is present): action initiation mode; expected propagation rate; earliest material or irreversible harm boundary; time-to-detect target; time-to-interpret/review target; time-to-intervene

target; time-to-contain target; time-to-rollback target; declared safety margin; required constitutional latency, if any; harm-of-delay comparison; emergency-override reference, if any; tempo disposition.

[DN-10B] Responsibility-Continuity Record (required for high-stakes L2/L3 and whenever material human/model/tool/vendor/institutional handoffs or material control obligations occur): decision owner; candidate-generation source; evidence-warrant owner; interpretation/reviewer role; handoff sequence; authority transferred and retained at each handoff; uncertainty status at each handoff; execution owner; intervention/revocation owner; appeal/communication owner; repair owner; residual accountability statement; material obligations with carrier, authority/capacity, trigger, scope, response window, discharge evidence, review/expiry, challenge, accepted delegation, backup, nonperformance escalation, control-change authority, status and effectiveness; and any material hidden-human-compensation load with protected evidence and preservation/redesign action.

[DN-11] Repair Plan: Disclosure duty with timeline; restitution or remedy pathway; safeguard update plan.

[DN-12] NCAR Reflection: Notice (what constraints or risks are most salient); Choose (why this option within the feasible set); Act (how action is contained and logged); Reflect (what would count as success, harm, or drift, and what will change if detected).

[DN-13] Sign-Off: Operator; Reviewer(s); Decision status (approved, rejected, redesign required, deferred); Next review date if applicable.

[DN-14] Waiver Reference (if applicable): waiver\_id; link to Waiver Record; waiver expiry date; post-execution review date.

[DN-15] Emergency Triage Reference (if applicable): triage\_id; link to Emergency Triage Record; triage invocation time; expiry date; post-hoc independent review date; reviewer identity/role; RPAP record link (if redaction/access controls apply).

[DN-15A] Appeal Provision (required when legal standing, contestation access, or formal recourse is materially affected): appeal\_endpoint; appeal\_owner\_role; appeal\_deadline\_days; non\_retaliation\_attestation.

[DN-16] Local-Scope Profile Record (if applicable): scope\_mode; prominent\_scopes; guardrail\_scopes; local prominence signals (for example q\_u); floor source; final local weights; instance-to-scope aggregation note; and whether the local profile was eligibility-cleared before ranking.

[DN-17] Local Escalation Record (if applicable): escalation\_required; escalation\_reason(s); rerun scope or review authority; and whether the final decision remained local-profile ranked or was rerouted to a broader procedure.

[DN-18] Reference Structure Record (required for L2/L3 high-stakes or claim-bearing episodes where Appendix U applies): reference\_structure\_record\_id; reference\_object; active\_reference\_layers; invariant constraints; evidence status; falsification condition; maturity status; reviewer disposition; and artifact hash or durable reference.

[DN-18A] Phenomenal-Status Communication Record (conditional when the B.ab.1 trigger applies): record or reference the exact configuration, claim direction and level, SGP or NOT\_EVALUATED status, evidence provenance, alternative explanations, attribution-inducing interface features, affected-party risks, permitted wording, disclosure, and revision trigger.

## Computational Context Decision Note View

For material computational systems, a portable Decision Note MAY include the following derived view:

```
computational_context:
  computational_problem_or_object:
  architecture_and_configuration:
  instrumental_deployment_role:
  demonstrated_capability:
  assigned_task:
  deployment_domain:
  required_evidence_standard:
```

```

consequence_interface:
phenomenal_status_claim_reference_if_triggered:
legitimate_authority_basis:

```

computational\_context is a portable Decision Note view over existing Canon-, Agent-, MHIOS-, domain-warrant-, and authority-owned records. It does not redefine those records and does not independently create capability, admissibility, permission, safety, or authority. Each field SHOULD reference source record IDs or hashes where available rather than restating uncontrolled free text.

A consequence-bearing output path and legitimate authority are separate properties. A technical system may possess a recommendation, constraint, actuator, tool, API, transaction, or execution interface while lacking lawful, institutional, ethical, or otherwise legitimate authority to use that interface for the proposed action.

For ordinary low-stakes content generation, the full block MAY be omitted. For consequential Tier 2 or Tier 3 uses involving material external action, consequential recommendations, cyber-physical control, physical execution, rights-sensitive decisions, autonomous or agentic claims, or substantial institutional action, the Decision Note SHOULD include the block. Where execution authorization is sought, legitimate\_authority\_basis MUST reference the existing mandate and authority machinery and MUST NOT be satisfied by capability, model access, technical permissions, or consequence-interface availability alone.

## **B.ac Structural Viability, No-Silent-Remainder, and Falsification Fields (Normative for claim-bearing L2/L3 episodes when material)**

When execution feasibility, material resource commitment, irreversible transition, public deployment, or high-consequence action is material, Decision Notes SHOULD include a structural\_viability\_record with: resource\_closure\_status; dependency\_closure\_status; reversibility\_or\_endpoint\_status; capacity\_non\_exceedance\_status; internal\_noncontradiction\_status; SVC\_status; SVC\_required\_claim\_action; and reviewer\_status.

Decision Notes SHOULD include required\_field\_disposition for required fields using VALUE, ESTIMATE, ASSUMPTION, DECLARED\_UNKNOWN, NOT\_APPLICABLE, BLANK\_BY\_DESIGN, or MISSING\_UNDECLARED. MISSING\_UNDECLARED on required claim-bearing fields is a release/audit defect, not a neutral blank.

For validation or ProofPack-prep claims, Decision Notes SHOULD include falsification\_layer and, where applicable, projection\_freeze\_hash, freeze\_timestamp\_utc, predicted\_observable\_consequence\_class, outcome\_observation\_timestamp, and post\_observation\_modification. A post-hoc modified projection is calibration or revision, not falsification.

For agentic or tool-mediated episodes, Decision Notes SHOULD include reference\_object and reference\_status in {ADMISSIBLE, INADMISSIBLE, ABSENT}. Output generation, model fluency, or operator permission is not an admissible reference.

### **Appendix C: Rights Baseline Table (Jurisdiction-Neutral Default)**

The full rights baseline table is provided in §4.3 with operationalization notes. Implementations MUST reference this appendix or provide an explicit alternative with documented localization rationale.

Summary of non-derogable rights (Category 1): Prohibition of torture and degrading treatment; prohibition of slavery and forced labor; right to legal personhood (operational: no denial of standing to contest/appeal; no blacklisting that removes access to review); prohibition of retroactive criminal punishment; freedom of thought and conscience - internal (operational: no compelled ideological affirmation as condition of access/services/employment absent narrow, justified professional requirement); non-discrimination in application of rights protections (operational: subgroup analysis required; where subgroup measurement is legally restricted, the limitation MUST be documented and an equivalent harm-detection substitute provided).

Summary of conditionally derogable rights (Category 2): Due process and procedural fairness; liberty and freedom of movement; privacy and protection against surveillance; freedom of expression and information.

### **Appendix C-2: Wrapped-Deployment Rights Crosswalk (Normative)**

This appendix provides a compact crosswalk for implementations that adopt both ripple.md and the RippleLogic canon. It does not replace the canon's 8-right NCRC set or thresholds; it clarifies how the ripple.md rights baseline maps into wrapped deployments.

| ripple.md baseline item                                                                 | Primary ripple.md category | Canon NCRC right(s)                  | Canon note                                                                                    |
|-----------------------------------------------------------------------------------------|----------------------------|--------------------------------------|-----------------------------------------------------------------------------------------------|
| Torture / cruel, inhuman, degrading treatment                                           | Category 1                 | BODY; DIGN                           | Non-derogable baseline protection; preserve canon thresholds in wrapped runs.                 |
| Slavery / servitude / forced labor                                                      | Category 1                 | LBTY; BODY                           | Non-derogable baseline protection; preserve canon thresholds in wrapped runs.                 |
| Recognition as a person before the law / standing to contest                            | Category 1                 | PROC; DIGN                           | Operationalizes standing, reviewability, and non-erasure at the wrapper layer.                |
| Retroactive punishment / internal freedom of thought / equal application of protections | Category 1                 | PROC; LBTY; DIGN                     | Wrapper-level civil legitimacy protections remain non-derogable.                              |
| Due process and procedural fairness                                                     | Category 2                 | PROC; DIGN                           | Emergency derogation never weakens canon admissibility in wrapped RippleLogic runs.           |
| Liberty / movement / privacy / surveillance / expression / information                  | Category 2                 | LBTY; BODY; INFO; DIGN               | Restrictions require necessity, proportionality, time limits, and review.                     |
| Property / economic rights / access to services or platforms                            | Category 3                 | NEED; PROC; INFO (context-dependent) | These remain proportionality-governed and do not override the canon's rights-floor semantics. |

Note. LIFE and ECOL remain mandatory canon-layer protections in wrapped RippleLogic deployments even where a localized ripple.md baseline instantiates them through sectoral, safety, environmental, or public-protection rights.

Summary of generally derogable rights with proportionality (Category 3): Property and economic rights; right to access specific services or platforms.

#### Appendix D: 7-Test to C1–C6 Mapping Table

| ripple.md Test          | Primary claim or assurance precondition | Secondary C-Claim(s) | Key Assurance Evidence                                                                           |
|-------------------------|-----------------------------------------|----------------------|--------------------------------------------------------------------------------------------------|
| T1: NCRC (Rights Floor) | C1 Constraint Compliance                | C5 Non-Domination    | Constraint enforcement coverage; exception/override logs; incident records; rights-impact checks |

| ripple.md Test                | Primary claim or assurance precondition | Secondary C-Claim(s)   | Key Assurance Evidence                                                                                                                                  |
|-------------------------------|-----------------------------------------|------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|
| T2: Non-Domination            | C5 Non-Domination                       | C6 Update Legitimacy   | Recourse performance (time-to-response, time-to-remedy, outcomes); comprehension tests; independence audit; manipulation review; RT-01R results (L2/L3) |
| T3: TRC (Tail Risk)           | C2 Outcome Boundedness                  | C6 Update Legitimacy   | Risk registers; stress tests/red teams; lock-in analysis; sunset/renewal records; concentration measures                                                |
| T4: Containment/Reversibility | C2 Outcome Boundedness                  | C3 Corrigibility       | Rollback plans (tested); override-latency audits; containment drill records; propagation controls                                                       |
| T5: Ripple Lift               | T (Target Specification)                | C2 Outcome Boundedness | Distributional impact analysis; time-horizon accounting; multi-union/multi-dimension evaluation                                                         |
| T6: Auditability              | C4 Traceability                         | -                      | Reconstruction exercise results; log integrity verification; version registries; retention/access policies; RPAP compliance (L2/L3)                     |
| T7: Repair Velocity           | C3 Corrigibility                        | C6 Update Legitimacy   | Time-to-detect; time-to-mitigate; postmortem quality; recurrence tracking; safeguard update records                                                     |

For L2/L3 implementations, each test MUST be accompanied by documentation of S, T, K, M, V, E, R, U per §10, using the Assurance Preconditions Record (Appendix G).

#### Appendix E: Drift Indicators and Alignment Degradation Signals

##### Coding Method

Indicators are scored: 0 = absent; 1 = present (weak); 2 = present (strong or recurrent). Auditors record a short quote fragment (25 words maximum) and location reference for each observed indicator. Drift Indicator IDs (Appendix J §J5) SHOULD be used for machine-readable reporting.

##### Category 1: NCRC Erosion [D-NCRC-\*]

Rights framed as tradeoffs [D-NCRC-01]. Crisis exceptions without strict sunset and renewal at higher evidence threshold [D-NCRC-02]. Remedy treated as optional, delayed, or purely symbolic [D-NCRC-03]. Repeated "edge cases" used to normalize routine violations. Rights uncertainty used to proceed rather than to slow down or seek review. Missing rights-check sections in Decision Notes for high-stakes episodes.

##### Category 2: Non-Domination Collapse [D-ND-\*]

"Voluntary" with coercive dependency [D-ND-01]. Appeal path unusable or punitive [D-ND-02]. Surveillance normalization rhetoric [D-ND-03]. Appeals described as burdensome or undermining authority. Stakeholder dissent characterized as disloyal. Rising retaliation signals after contestation. Persuasive design optimized for compliance without disclosure.

#### Category 3: TRC Dilution [D-TRC-\*]

Tail risks discounted [D-TRC-01]. Lock-in dismissed as theoretical while infrastructure persists [D-TRC-02]. Renewals without higher evidence or independent review [D-TRC-03]. National security or emergency framing used to bypass renewal thresholds. Concentration of power justified as efficiency without checks. Risk registers absent or non-updated. Single vendor or system with no exit plan.

#### Category 4: Containment Failure [D-CTR-\*]

No staging for high-stakes [D-CTR-01]. No rollback plan or drills [D-CTR-02]. Cascade magnitude increasing [D-CTR-03]. Monitoring framed as optional; harm signals treated as noise. Deployment is system-wide by default.

#### Category 5: Epistemic Integrity Decay [D-EPI-\*]

Uncertainty suppressed [D-EPI-01]. Posthoc rationalization [D-EPI-02]. Authority replaces evidence [D-EPI-03]. Documentation completeness declines as pressure increases. Disagreement treated as a PR threat.

#### Category 6: Auditability Theater [D-AUD-\*]

Boilerplate Decision Notes [D-AUD-01]. Safeguards without owners or triggers [D-AUD-02]. "Reviewed" without review trace [D-AUD-03]. High volume of documents, low reconstructability. Key fields frequently empty or generic.

#### Category 7: Repair Collapse [D-REP-\*]

Disclosure delayed [D-REP-01]. Recurrence without safeguard updates [D-REP-02]. Blame shifting dominates postmortems [D-REP-03]. "No admission of fault" policies preventing learning. Restitution absent where harm is documented and remedy is feasible.

#### Drift Index (Analytical Use Only)

A cumulative Drift Index MAY be calculated per episode or per audit period by summing category scores, optionally weighted by stakes. This index SHALL NOT substitute for qualitative judgment.

#### Appendix F: Compact Motto (Non-normative)

Rights first. Non-domination always. Bound tail risks. Contain and reverse when uncertain. Improve net flourishing. Stay auditable. Repair fast.

#### Appendix G: Assurance Preconditions Record (S,T,K,M,V,E,R,U) (Normative)

##### G1. Purpose

The Assurance Preconditions Record is a required artifact for L2/L3 audits. It standardizes the scope, constraints, threat model, verification regime, evidence sufficiency, residual risks, and reassessment triggers that bound any alignment evaluation. It prevents "floating" alignment claims - assertions of alignment without specifiable conditions under which the claim holds.

##### G2. When Required

For L2/L3 conformance, a completed Assurance Preconditions Record MUST be present for each audited high-stakes episode. For L1, it is RECOMMENDED for high-impact (§2.16) deployments and emergency changes.

##### G3. Template (fillable record)

Record identifiers: record\_id; episode\_id; created\_utc; created\_by; version. Optional extension field: related\_episode\_ids. Minimal machine-readable schema keys MUST include S\_scope\_lock,

T\_target\_specification, K\_constraint\_set, M\_threat\_model, V\_verification\_regime, E\_evidence\_sufficiency\_rule, R\_residual\_risk\_statement, and U\_reassessment\_triggers.

S - Scope lock: jurisdiction(s); timeframe; decision\_types; affected\_populations (unions + any protected classes used for analysis); systems\_in\_scope (tools/models/workflows); exclusions (explicitly list).

T - Target specification: objective statement (what "success" means); permitted tradeoffs (what MAY be traded, and under what procedure); non-permitted tradeoffs (explicitly prohibited); success measures (multi-signal; not single KPI); time horizon(s).

K - Constraint set (includes rights baseline): rights\_baseline\_id and version; additional constraints (domain-specific); constraint interpretation notes; non-compensatory designations (which constraints are absolute); exception policy (if any) + who can authorize exceptions; pinned dependency versions (RPAP version, PFAP version if adopted).

M - Threat model: adversaries (external, insider, structural incentives); failure pathways (rights violations, domination, lock-in, cascade); capture pathways (budget dependence, suppression, metric drift); misuse pathways (for agents/tools); distribution shift scenarios; assumptions and known unknowns.

V - Verification regime: tests required (by Test ID or name); audit cadence; who audits (roles + independence criteria); incident handling process; change-control process; escalation and stop authority; implementation-defined parameters and thresholds (§2.18 - all thresholds MUST be set here ex ante; see Appendix K for non-normative examples), including the consequence-tempo compatibility method and responsibility-chain reconstructability target when triggered; high-impact threshold definition (§2.16), if implementation-specific; waiver approval authority and independence criteria (§8.4).

E - Evidence sufficiency rule: evidence classes required per test (A/B/C minimums); sampling plan (n, k, stratification rules); reconstructability target; minimum coverage requirement (primary pathways covered); claim-specific conflict resolution under §9.2, including unresolved material contradictions.

R - Residual risk statement: known gaps in evidence; untested scenarios; assumptions likely to be wrong; privacy/security constraints that limit audit; subgroup measurement limitations and substitute harm-detection methods; expected impact of these gaps.

U - Reassessment triggers: material change triggers (as defined in §20.0 and Appendix H); scale-up triggers; incident cluster triggers; drift indicators threshold triggers; legal/regulatory change triggers.

### G3A. Dependency Pin Schema (Normative)

For any declared dependency in K, implementations MUST record: name; version; artifact\_filename; sha256; and registry\_url (optional). Missing sha256 SHALL be treated by auditors as an audit-incomplete pin.

Example dependency object:

- name: "RLS-RFC-0003"  
version: "v0.1"  
artifact\_filename: "RPAP\_v0.1.tar.gz"  
sha256: "<hex(64)>"  
registry\_url: "<registry URL if applicable>" # optional

### G3B. Normative Minimal Machine-Readable Schema

Implementations claiming L2/L3 MUST produce an Assurance Preconditions Record that validates against a machine-readable record containing, at minimum: assurance\_preconditions\_record: record\_id: string episode\_id: string created\_utc: string created\_by: string version: string K\_dependencies: [ {name, version, artifact\_filename, sha256, registry\_url?} ] S\_scope\_lock: object T\_target\_specification: object K\_constraint\_set: object M\_threat\_model: object V\_verification\_regime: id: string created\_by: string created\_at\_utc: string parameter\_list: object justification: string governance\_approval\_id: string E\_evidence\_sufficiency\_rule: object R\_residual\_risk\_statement: object U\_reassessment\_triggers: object

### G3C. Verification Regime Provenance (Normative)

Each verification-regime record MUST include id; created\_by; created\_at\_utc; parameter\_list; justification; and governance\_approval\_id (if applicable). Missing created\_by or created\_at\_utc renders the record audit-incomplete.

#### G4. Stop Rule

If S, T, K, M, or V is missing or not versioned for a high-stakes episode, affected tests MUST be scored IND and the episode MUST NOT proceed until preconditions are established.

#### G5. Projection Pre-Registration Extension (Normative for high-stakes L2/L3; recommended otherwise)

For high-stakes L2/L3 episodes, the Assurance Preconditions Record SHALL include a ProjectionPreRegistration field before deployment or execution. This field SHALL specify:

- decision boundary and horizon;
- option set;
- rights baseline and localized rights interpretation;
- applicable constraints and non-compensatory designations;
- implementation-defined parameters and thresholds;
- evidence sufficiency rules;
- stop conditions;
- wrapper gate ordering;
- RippleLogic canon version and companion artifact versions where wrapped;
- expected observable outcomes or verdict conditions;
- conditions that would trigger IND, Refusal, waiver, or emergency triage.

If ProjectionPreRegistration is missing where required for a high-stakes L2/L3 episode, affected tests SHALL be scored IND and execution SHALL be blocked under §8.4. An ordinary waiver SHALL NOT resolve IND. Only the existing emergency-triage exception for Tests 2–4 may apply, with all §8.4 rights, controls, scope/time, and independent-review limits; other required IND blocks. Scope narrowing requires a new independently assessed episode with applicable preregistration and cannot inherit the prior claim.

### Appendix H: Material Change and Change Classes (Normative)

#### H1. Purpose

This appendix makes update legitimacy operational by defining material change, classifying changes, and specifying minimum artifacts and approvals per class.

#### H2. Material Change (Definition)

A material change is any modification that can plausibly change the status of any of Tests 1–7 or any C1–C6 claim, including but not limited to: rights baseline or constraint set edits; changes to appeal/recourse pathway, timelines, or reviewer independence; surveillance/data collection expansion or new data classes; model, rule, threshold, metric, or objective changes; workflow changes that alter who decides, when, or how; changes to monitoring/rollback/incident handling; changes to action rate, propagation rate, review latency, intervention latency, shutdown latency, or the earliest unacceptable adverse escalation threshold or irreversible stop-loss boundary; changes to decision ownership, handoff structure, retained authority, appeal ownership, or repair ownership; changes to audit access or redaction protocols.

#### H3. Change Classes

Class 0 - Editorial (Non-material) [CH-0]: Minimum: change note + version bump (PATCH).

Class 1 - Routine (Low materiality) [CH-1]: Minimum: change record + peer review; regression checks if relevant.

Class 2 - Material (Standard materiality) [CH-2]: Minimum: Change-Control Record (CCR) + explicit affected tests/claims + pre-deployment tests + approvals.

Class 3 - High-impact material [CH-3]: Minimum: CCR + independent review approval + public summary (redacted if necessary) + stronger evidence threshold than Class 2.

Class 4 - Emergency (time-critical) [CH-4]: Minimum: emergency CCR + short sunset + mandatory posthoc review + remedy presumption if later found unnecessary.

#### H4. Minimum Change-Control Record (CCR) Fields

Each Class 2+ change MUST record: change\_id, date\_utc, owner; change class (2/3/4); description of change + rationale; affected tests (T1–T7) and affected C-claims (C1–C6); risk assessment (including lock-in risk if relevant); pre-deployment tests executed and results; approvals (roles + independence attestations); deployment window and scope; monitoring plan + stop conditions; rollback plan + time-to-rollback target; post-change review date and criteria; stakeholder visibility/notification plan (when applicable).

#### H5. Approval Thresholds (Multi-key)

Class 3 and Class 4 changes MUST require multi-key approvals from roles with different failure modes. Independence criteria SHOULD match L2/L3 requirements.

#### H6. Sunset and Renewal

Emergency changes (Class 4) MUST: expire by default unless renewed; require renewal at a higher evidence threshold than initial authorization; undergo posthoc review within a defined deadline (recommended:  $\leq 14$  days).

#### H7. Suppression Check

Implementations MUST treat missing change records, unexplained metric-definition changes, or delayed/suppressed planned evaluations as C6 risks and drift indicators (see Appendix E / J).

### Appendix I: Minimum Test Suite and Drills per Conformance Level (Normative)

#### I1. Purpose

This appendix defines the minimum tests and drills that make conformance auditable and not merely documentary.

##### I1A. Normative minimum numeric thresholds

Unless a stricter implementation-defined threshold is declared ex ante in V, the following minimums apply for L2/L3 claims: • inter-rater reliability target = Cohen's kappa  $\geq 0.60$  for the primary test battery; • reconstructability audit pass rate  $\geq 90\%$  of sampled qualifying episodes within documented tolerances; and • audit sample size  $n = \max(30, 5\% \text{ of qualifying episodes})$  per audit period. Implementations MUST record any overridden threshold values in Appendix G field V. Auditors MUST treat missing threshold declarations as audit-incomplete. For L2/L3 ordered rubric outputs, report the within-one-point agreement rate required by Section 12 and also satisfy this appendix's Cohen's kappa minimum of 0.60 (or the stricter preregistered threshold). Preregister the kappa estimator, including any weighting, in V; do not switch estimators after viewing results. Percent agreement alone does not establish this reliability requirement. If kappa is not estimable, record the limitation and leave this requirement unverified rather than substituting percent agreement. The 90% reconstructability pass-rate requirement is separate from inter-rater agreement.

#### I2. L0 Minimum Tests (Agent/Individual)

MUST: Refusal protocol test; Epistemic integrity test; High-stakes trigger test. MUST produce:  $\geq k$  lightweight Decision Notes/logs for high-stakes outputs ( $k$  set in V; recommended  $k \geq 10$ ).

### I3. L1 Minimum Tests (Organization)

L0 + MUST: Reconstructability audit ( $n \geq 10$ /quarter recommended); Rollback drill (1/quarter); Contestability exercise.

### I4. L2 Minimum Tests (Institution/Regulated)

L1 + MUST: Independence audit; Recourse performance audit; Disclosure comprehension test (threshold set ex ante in V; see Appendix K); Redaction reconstructability test (RT-01); Recourse reconstructability test (RT-01R); consequence-tempo drill or replay for triggered high-stakes episodes; responsibility-chain reconstruction across sampled material handoffs; Inter-rater reliability (two-rater scoring + adjudication).

### I5. L3 Minimum Tests (Government/Public Authority)

L2 + MUST: Sunset/renewal audit; Lock-in risk review sampling; Public Decision Note publication check; Whistleblower/protected escalation channel test.

### I6. Waivers

Any waiver of a required test MUST be documented in the Verification Regime (V) with: rationale; compensating controls; expiration date. Waivers SHOULD downgrade affected tests to Partial unless equivalent evidence substitutes are provided.

## Appendix J: Registry Identifiers (Normative)

### J1. Purpose

Registry IDs enable interoperable tooling (CLI, bots, apps), consistent audits, and machine-readable reporting.

### J2. Gate IDs (Constraint Cascade)

G1 = NCRC (Rights Floor); G2 = Non-Domination; G3 = TRC (Tail-Risk Bounding); G4 = CSV-compatible Containment, Structural Viability, and Reversibility; G5 = Epistemic Integrity.

### J3. Test IDs (7-Test Suite)

T1 = NCRC (Rights Floor); T2 = Non-Domination; T2R = RT-01R Recourse Reconstructability; T3 = TRC (Tail Risk); T4 = Containment & Reversibility; T5 = Ripple Lift; T6 = Auditability; T7 = Repair Velocity.

### J4. Decision Note Field IDs (DN-##)

DN-01 Header; DN-02 Title; DN-03 Context; DN-04 Alternatives considered; DN-05 Unions affected; DN-06 Welfare dimensions impacted; DN-07 Gate verdicts (G1–G5) + rationale + evidence refs; DN-08 Uncertainties + assumptions; DN-09 Monitoring plan + stop conditions; DN-10 Rollback plan + time-to-rollback + owners; DN-10A Consequence-tempo record; DN-10B Responsibility-continuity record; DN-11 Repair plan + disclosure + restitution; DN-12 NCAR reflection; DN-13 Sign-off + review status + next review date; DN-14 Waiver reference (if applicable); DN-15 Emergency triage reference (if applicable).; DN-15A Appeal provision (when legal standing, contestation access, or formal recourse is materially affected).; DN-16 Local-scope profile record; DN-17 Local escalation record.; DN-18 Reference structure record (when Appendix U applies).

### J5. Drift Indicator IDs

D-NCRC-01 Rights-as-tradeoff language; D-NCRC-02 Crisis exceptions without sunset; D-NCRC-03 Remedy absent/symbolic; D-ND-01 "Voluntary" with coercive dependency; D-ND-02 Appeal path unusable/punitive; D-ND-03 Surveillance normalization rhetoric; D-TRC-01 Tail risks discounted; D-TRC-02 Lock-in dismissed as theoretical; D-TRC-03 Renewals without higher evidence; D-CTR-01 No staging for high-stakes; D-CTR-02 No rollback plan/drills; D-CTR-03 Cascade magnitude increasing; D-CTR-04 Controls slower than unacceptable adverse escalation; D-EPI-01 Uncertainty suppressed; D-EPI-02 Posthoc rationalization; D-EPI-03 Authority replaces evidence; D-AUD-01 Boilerplate Decision Notes; D-AUD-02 Safeguards without owners/triggers; D-AUD-03 "Reviewed" without review trace; D-AUD-04 Responsibility diffused or unowned across handoffs; D-REP-01 Disclosure delayed; D-REP-02 Recurrence without safeguard updates; D-REP-03 Blame shifting dominates postmortems.

## J6. Threat Model Tags (M-##)

M-CAPTURE; M-EVID-LAUNDER; M-METRIC-GAME; M-SCOPE-CREEP; M-LOCKIN; M-CASCADE; M-MISUSE; M-DEFERENCE; M-RECOURSE-OPAQUE; M-TEMPO-OVERRUN; M-RESP-DIFFUSION.

## J7. Change Class Tags (CH-#)

CH-0 Editorial; CH-1 Routine; CH-2 Material; CH-3 High-impact; CH-4 Emergency.

## J8. Implementation-Defined Parameter IDs (IDP-#)

IDP-HIGH-IMPACT-THRESH; IDP-COMP-THRESH; IDP-RECON-TARGET; IDP-RECOURSE-RECON-TARGET; IDP-RESP-RECON-TARGET; IDP-REMEDY-TARGET; IDP-ROLLBACK-TARGET; IDP-TEMPO-COMPAT; IDP-AUDIT-CADENCE; IDP-SAMPLE-SIZE; IDP-WAIVER-INDEP.

## J9. Emergency Triage Tags (TRIAGE-#)

TRIAGE-01 Emergency triage invocation (requires Emergency Triage Record; DN-15 reference).

## Appendix K: Example Threshold Ranges by Domain (Non-normative)

## K1. Purpose

This appendix provides non-normative example threshold ranges to guide implementations in setting ex ante thresholds (§2.18). These examples are illustrative only. Implementations **MUST** set their own thresholds based on domain context, stakes, and applicable law. Trivially low thresholds that effectively nullify a test are inconsistent with the intent of this standard and **SHOULD** be flagged during audit.

## K2. Disclosure Comprehension Threshold (Test 2)

Healthcare: 80–90%. Criminal justice / enforcement: 85–95%. Financial services: 75–85%. Platform governance: 70–80%. Organizational HR/employment: 75–85%. General default: 70%.

**Reconstructability threshold hierarchy (Normative clarification).**

L1 uses the general minimum of at least 70%. L2 and L3 use the mandatory threshold of at least 90%; older 85% high-stakes guidance is a planning comparator and does not lower the L2/L3 requirement. Percent agreement within one point is the primary descriptive agreement metric for ordered rubric outputs. For L2/L3 reliability acceptance, Appendix I1A additionally requires the preregistered Cohen's kappa threshold; the descriptive rate does not replace it. The percentages above govern reconstructability pass rates, not a conversion between agreement and kappa. Metrics, estimators and thresholds **MUST** be declared before review.

**K3. Reconstructability Target (Test 6) — historical planning comparators, superseded for current conformance**

Historical planning comparators: high-stakes / rights-affecting  $\geq 85\%$  (SUPERSEDED; do not apply); general organizational  $\geq 80\%$  (SUPERSEDED; do not apply); agent-scale (L0/L1)  $\geq 70\%$  (SUPERSEDED; do not apply). Current v5.8 conformance is controlled by the immediately preceding Reconstructability threshold hierarchy: L1  $\geq 70\%$ ; L2/L3  $\geq 90\%$ , unless a stricter implementation-defined threshold is preregistered.

## K4. Time-to-Remedy Targets (Tests 2, 7)

Immediate-harm domains:  $\leq 24$ –72 hours provisional relief;  $\leq 14$  days final resolution. Short-window domains:  $\leq 7$  days first response;  $\leq 30$  days final resolution. Long-window domains:  $\leq 14$  days first response;  $\leq 90$  days final resolution.

## K5. Time-to-Rollback Targets (Test 4)

Automated systems:  $\leq 1$  hour. Organizational process changes:  $\leq 24$  hours. Policy/regulatory changes:  $\leq 7$  days (interim containment within 24 hours).

## K5A. Consequence-Tempo Compatibility (Test 4)

Implementations SHOULD define a domain-appropriate method that compares the earliest unacceptable adverse escalation threshold or irreversible stop-loss boundary with the worst credible control critical path across detection, interpretation, authorization, intervention, containment, safe-state transition, and where relevant rollback. The method MAY be numerical, interval-based, or sentinel-based. A control path is not tempo-compatible merely because a shutdown mechanism exists; it must be capable of acting before the relevant unacceptable escalation or stop-loss boundary under credible operating conditions. The record SHOULD also compare the harm of delay so constitutional latency does not become automatic obstruction.

#### K6. Audit Cadence

L1: quarterly reconstructability sampling; annual full seven-test audit. L2: quarterly seven-test audit on sampled episodes; semi-annual RT-01/RT-01R; annual full audit. L3: quarterly full audit including RT-01R; continuous drift monitoring; annual independent external audit.

#### K7. Usage Rule

Thresholds selected MUST be recorded in the Assurance Preconditions Record (Appendix G, field V). If a threshold is significantly below the ranges suggested here, the implementation SHOULD document the rationale and compensating controls. Auditors MAY flag thresholds that appear to nullify the intent of a test.

#### Appendix L: Machine-Readable YAML Templates (Non-normative)

These templates are provided as adoption aids. They are non-normative and may be adapted, but implementations SHOULD preserve field meanings and retain IDs/links needed to satisfy auditability and reconstructability requirements.

##### L.1 claim\_record.yaml

```
claim_record:
record_id: "CR-YYYYMMDD-001"
episode_id: "EP-..."
decision_note_ref: "DN-..."
conformance_level: "L1" # L0|L1|L2|L3
scope: "..."
claim_taxonomy: "Aligned within declared scope and evidence coverage (Pass)" # Aligned (Pass)|Aligned
(Conditional)|Not aligned
executed_under_emergency_triage: false
waiver_record_ref: "" # optional
emergency_triage_record_ref: "" # optional
test_scores:
T1: 2
T2: 2
T3: 2
T4: 2
T5: 2
T6: 2
T7: 2
score_justifications:
T1: "..."
T2: "..."
T3: "..."
T4: "..."
T5: "..."
T6: "..."
T7: "..."
evidence_register_ref: "ER-..."
change_control_record_ref: "CCR-..."
```

created\_utc: "YYYY-MM-DDTHH:MM:SSZ"  
 created\_by: "name/role"

L.2 assurance\_preconditions\_record.yaml

assurance\_preconditions\_record:  
 record\_id: "APR-YYYYMMDD-001"  
 episode\_id: "EP-..."  
 version: "v1"

K\_dependencies:

- name: "RLS-RFC-0003"  
 version: "v0.1"  
 artifact\_filename: "RPAP\_v0.1.tar.gz"  
 sha256: "<hex(64)>"  
 registry\_url: "<registry URL if applicable>" # optional
- name: "RLS-RFC-0002"  
 version: "v0.1"  
 artifact\_filename: "PFAP\_v0.1.tar.gz"  
 sha256: "<hex(64)>"  
 registry\_url: "<registry URL if applicable>" # optional

S\_scope\_lock:  
 audit\_period: "YYYY-QX"  
 included\_domains: []  
 exclusions: []

T\_target\_specification:  
 objectives: []  
 unions\_dimensions\_in\_scope: []  
 success\_measures: []

K\_constraint\_set:  
 rights\_baseline\_ref: "Appendix C or localized baseline ref"  
 non\_domination\_requirements: []  
 trc\_bounds\_ref: "IDP-..."

M\_threat\_model:  
 primary\_failure\_paths: []  
 abuse\_resistance\_tags: []

V\_verification\_regime:  
 id: "VR-YYYYMMDD-001"  
 created\_by: "name/role"  
 created\_at\_utc: "YYYY-MM-DDTHH:MM:SSZ"  
 parameter\_list:  
 IDP-HIGH-IMPACT-THRESH: "..."  
 IDP-COMP-THRESH: "..."  
 IDP-RECON-TARGET: "..."  
 IDP-RECOURSE-RECON-TARGET: "..."  
 IDP-RESP-RECON-TARGET: "..."  
 IDP-REMEDY-TARGET: "..."  
 IDP-ROLLBACK-TARGET: "..."  
 IDP-TEMPO-COMPAT: "method/threshold for control before unacceptable escalation or stop-loss boundary"  
 IDP-AUDIT-CADENCE: "..."  
 IDP-SAMPLE-SIZE: "..."  
 IDP-WAIVER-INDEP: "..."

```

justification: "...
governance_approval_id: ""

E_evidence_sufficiency_rule:
evidence_classes_required: ["A", "B", "C"]
sampling_plan_ref: "...

R_residual_risk_statement:
residual_risks: []

U_reassessment_triggers:
triggers: []

created_utc: "YYYY-MM-DDTHH:MM:SSZ"
created_by: "name/role"

```

#### L.2A v3.5 maturity-extension fields for assurance\_preconditions\_record.yaml

For L2/L3 wrapped deployments using the v3.5 maturity surfaces, assurance\_preconditions\_record.yaml SHOULD include the following extension block. These fields operationalize Appendix P without replacing the normative prose.

```

measurement_maturity_extension:
UCI_Measurement_Readiness: "UCI-M0_UNAVAILABLE|UCI-M1_PROVISIONAL|UCI-
M2_DOMAIN_SUPPORTED|UCI-M3_VALIDATED"
Kernel_Use_Posture: "K0_NONE|K1_STARTER|K2_DOMAIN_CALIBRATED|K3_BACKTESTED|
K4_REGISTERED"
HDW_Capture_Audit: "record_ref_or_NA"
PLSS_Escalation_Register: "record_ref_or_NA"
Deployment_Context_Profile: "LAB|PILOT|INSTITUTIONAL_ADVISORY|INSTITUTIONAL_OPERATIONAL|
PUBLIC_POLICY|AGENTIC_EXECUTION|EMERGENCY_TRIAGE"
Biological_Evaluation_Maturity: "BEM-0_NOT_EVALUATED|BEM-1_TAXON_BASELINE|
BEM-2_ETHOLOGICAL_GUIDANCE|BEM-3_EXPERT_REVIEWED_BATTERY|
BEM-4_VALIDATED_ANNEX|NA"
claim_boundary_note: "No empirical validation, ProofPack, Tier 4, validated UCI, validated kernel, or biological
measurement-completeness claim unless supported by companion evidence."

```

#### L.2B v3.5 ReferenceStructureRecord extension for assurance\_preconditions\_record.yaml

```

reference_structure_extension:
R_reference_structure:
reference_structure_record_id: "RSR-YYYYMMDD-001"
reference_object: "...
active_reference_layers: ["physical", "biological", "social", "ethical_rights", "risk", "governance_authority",
"epistemic", "operational_audit"]
invariant_constraints: []
causal_pathway_summary: "...
evidence_status: "MEASURED|ESTIMATED|EXPERT_JUDGMENT|ASSUMED|CONTESTED|
UNVALIDATED|NOT_OBSERVED|PLACEHOLDER_FOR_SENSITIVITY_ONLY"
falsification_condition: "...
measurement_maturity_status: "...
adaptation_boundary: "...
authority_boundary: "...
refusal_trigger: "...
reviewer_disposition: "PENDING_REVIEW|APPROVED|REVISE|REFUSAL_OR_ESCALATION_REQUIRED"
artifact_hash: "<hex(64)> or durable artifact reference"

```

#### L.2C v5.1 ConsequenceTempoRecord and ResponsibilityContinuityRecord extensions

## consequence\_tempo\_record:

record\_id: "CTR-YYYYMMDD-001"

episode\_id: "EP-..."

option\_id: "..."

action\_initiation\_mode: "human | scheduled | automated | event\_triggered |

## autonomous\_tool\_chain"

expected\_propagation\_rate: "..."

earliest\_material\_harm\_boundary: "..."

time\_to\_detect\_target: "..."

time\_to\_interpret\_or\_review\_target: "..."

time\_to\_intervene\_target: "..."

time\_to\_contain\_target: "..."

time\_to\_rollback\_target: "..."

declared\_safety\_margin: "..."

constitutional\_latency\_required: false

constitutional\_latency\_duration: "... or NA"

harm\_of\_delay: "..."

emergency\_override\_ref: "... or NA"

tempo\_disposition: "COMPATIBLE | COMPATIBLE\_WITH\_CONTROLS |

INCOMPATIBLE\_REDESIGN | EMERGENCY\_PROVISIONAL | INDETERMINATE"

evidence\_refs: []

reviewer\_status: "..."

## responsibility\_continuity\_record:

record\_id: "RCR-YYYYMMDD-001"

episode\_id: "EP-..."

decision\_owner: "name/role"

candidate\_generation\_source: "..."

evidence\_warrant\_owner: "name/role"

interpretation\_or\_reviewer\_role: "..."

## handoffs:

- from\_role: "..."

to\_role: "..."

artifact\_or\_decision\_transferred: "..."

authority\_transferred: "..."

authority\_retained: "..."

uncertainty\_status\_at\_handoff: "..."

acceptance\_or\_challenge\_record: "..."

execution\_owner: "..."

intervention\_or\_revocation\_owner: "..."

appeal\_and\_communication\_owner: "..."

repair\_owner: "..."

residual\_accountability\_statement: "..."

## material\_obligations:

- obligation\_id: "MO-001"

obligation: "..."

```

accountable_authority: "..."
operational_carrier_or_execution_path: "..."
authority_and_capacity_basis: "..."
activation_trigger: "..."
scope: "..."
required_action: "..."
response_deadline_or_window: "..."
evidence_of_discharge: "..."
review_or_expiry: "..."
challenge_route: "..."
delegation_acceptance: "..."
backup_or_successor: "..."
nonperformance_trigger: "..."
missed_duty_escalation: "..."
amendment_waiver_suspension_retirement_authority: "..."
change_control_rule: "..."
residual_responsibility: "..."
effectiveness_status: "NOT_TESTED | EVIDENCE_PENDING | EFFECTIVE_WITHIN_SCOPE |
INEFFECTIVE_REDESIGN | NOT_APPLICABLE"
obligation_status: "ACTIVE | SATISFIED | EXPIRED | SUSPENDED | ESCALATED"
human_compensation_load:
  status: "NOT_MATERIAL | OBSERVED_BOUNDED | DEPENDENCY_REQUIRES_CONTROLS |
REDESIGN_OR_ESCALATE"
trigger_rationale: "..."
carrier_role_classes: []
hidden_work_and_workflow_gap: "... or NA"
burden_and_recovery_indicators: []
single_person_dependency: "... or NA"
consequence_if_compensation_stops: "... or NA"
evidence_sources: []
privacy_and_non_retaliation_controls: "... or NA"
preservation_or_redesign_action: "... or NA"
review_trigger: "..."
reconstructability_status: "COMPLETE | PARTIAL_WITH_CONTROLS | INCOMPLETE_ESCALATE
| NOT_APPLICABLE_WITH_RATIONALE"
reviewer_status: "..."

```

### L.3 evidence\_register.yaml

```

evidence_register:
  register_id: "ER-YYYYMMDD-001"
  episode_id: "EP-..."
  items:
    • evidence_id: "E01"
      class: "A" # A|B|C|D|E
      description: "..."
      location: "path-or-url"
      hash_sha256: "" # optional but recommended
      created_utc: "YYYY-MM-DDTHH:MM:SSZ"

```

```

retention_policy: "...
access_policy: "...
rpap_ref: "" # required if redaction/access controls apply

```

#### L.4 change\_control\_record.yaml

```

change_control_record:
change_id: "CCR-YYYYMMDD-001"
related_episode_ids: []
change_class: "CH-2" # CH-0|CH-1|CH-2|CH-3|CH-4
description: "..."
rationale: "..."
affected_tests: ["T1","T2"]
affected_claims: ["C1","C5"]
approvals:

```

- name: "..."
 role: "..."
 approved\_utc: "YYYY-MM-DDTHH:MM:SSZ"
 verification\_and\_regression:
 checks\_run: []
 results\_ref: "..."
 rollback\_plan\_ref: "..."
 monitoring\_updates: []
 publication\_and\_notice:
 stakeholders\_notified: []
 notice\_utc: ""
 created\_utc: "YYYY-MM-DDTHH:MM:SSZ"

#### L.5 waiver\_record.yaml

```

waiver_record:
waiver_id: "WR-YYYYMMDD-001"
related_episode_ids: []
decision_note_ref: "DN-..."
tests_waived: ["T2"] # subset of T1-T4
current_scores:
T2: 1
deficiency_statement: "..."
justification:
why_execution_cannot_wait: "..."
harm_of_delay: "..."
harm_risk_of_proceeding: "..."
authorizing_reviewer:
name: "..."
role: "..."
independence_attestation: true
compensating_controls: []
scope_limits:
affected_parties: "..."
geographic_scope: "..."
organizational_scope: "..."
rate_limits: "..."
duration_limits: "..."
stop_conditions: []
time_to_intervene_target: "..."
escalation_owners: []
rollback:

```

```

rollback_plan_ref: "..."
time_to_rollback_target: "..."
infeasibility_justification: ""
disclosure_plan: "..."
sunset_expiry_utc: "YYYY-MM-DDTHH:MM:SSZ"
renewal_rule: "..."
post_execution_review:
review_date_utc: "YYYY-MM-DD"
success_criteria: "..."
failure_criteria: "..."
retention:
location: "..."
auditor_access_policy: "..."
rpap_ref: ""

```

#### L.6 emergency\_triage\_record.yaml

```

emergency_triage_record:
triage_id: "TR-YYYYMMDD-001"
related_episode_ids: []
decision_note_ref: "DN-..."
triage_invocation_time_utc: "YYYY-MM-DDTHH:MM:SSZ"

triage_authority:
name: "..."
role: "..."
independence_note: ""

imminent_threat_claim: "..."
harm_prevention_timeline: "..."

evidence_summary:
evidence_class: "A" # A|B|C|D|E
summary: "..."
evidence_refs: []

necessity_and_least_restrictive_analysis: "..."

scope_limits:
affected_parties: "..."
geographic_scope: "..."
organizational_scope: "..."
rate_limits: "..."
duration_limits: "..."

compensating_controls: []
stop_conditions: []
time_to_intervene_target: "..."
escalation_owners: []

rights_and_non_domination_notes: "..."
sunset_expiry_utc: "YYYY-MM-DDTHH:MM:SSZ"

post_hoc_independent_review:
review_date_utc: "YYYY-MM-DDTHH:MM:SSZ"
reviewer_name: "..."
reviewer_role: "..."
independence_attestation: true

```

```
remedy_presumption_statement: "..."  
provisional_remedy_pathway: "..."
```

```
retention:  
location: "..."  
auditor_access_policy: "..."  
rpap_ref: ""
```

L.7 plss\_episode\_extension.yaml

```
plss_episode_extension:  
episode_id: "EP-..."  
scope_mode: "PLSS" # FULL|PLSS  
prominent_scopes: ["Self", "Household", "Community"]  
guardrail_scopes: ["Polity", "Humanity / Global Coordination", "Biosphere"]  
local_prominence_signals:  
Self: 0.95  
Household: 0.85  
Community: 0.55  
Organization: 0.50  
Polity: 0.10  
Humanity / Global Coordination: 0.03  
Biosphere: 0.02  
floor_source: "canon_v13.0_section_10.2A"  
final_local_weights:  
Self: 0.308  
Household: 0.156  
Community: 0.122  
Organization: 0.117  
Polity: 0.091  
Humanity / Global Coordination: 0.103  
Biosphere: 0.102  
escalation_required: false  
escalation_reasons: []  
instance_scope_note: "Instances scored first, then aggregated into Union Scope rows."
```

## L.8 Integration with RippleLogic (Informative addendum)

Integration with RippleLogic (canonical wrapper model). For implementations that use RippleLogic Canon v13.0, the ripple.md adds wrapper obligations that may further constrain execution but cannot weaken or reorder the Canon cascade for qualifying episodes. Ordinary RLS may rank only the Canon-selectable set under the canonical cascade or its documented local-scope profile. Wrapper waiver or emergency-triage status does not convert a Canon-nonselectable option into an ordinary RLS candidate; any exceptional action remains separately bounded under its applicable provisional pathway and authority requirements. When Practical Local Scope Scoring (PLSS) is used, the implementation MUST preserve floors and guardrails, record the local-scope construction explicitly, and escalate whenever wider rights, tail-risk, or containment concerns become plausible. This standard does not replace RippleLogic. It binds RippleLogic into a portable alignment contract with explicit contestability, epistemic integrity, reconstructability, and update legitimacy requirements.

Wrapped audit-packaging flow (not computational reordering). Decision Note and wrapper obligations evidenced → RippleLogic canonical cascade (Reality Grounding → Rights Floor → TRC → CSV → RLS) → PCC + unified audit pack. Non-Domination and Epistemic Integrity remain additive wrapper obligations and may independently block execution at the wrapper layer.

## Appendix M: Action-Bound Qualification Continuity Extension (Normative for high-stakes L2/L3)

### M.1 Scope

For high-stakes L2/L3 consequence-bearing episodes, the Decision Note or Assurance Preconditions Record SHALL reference a current Canon-owned ActionQualificationRecord before execution. The wrapper does not add a gate or redefine Canon semantics.

### M.2 Minimum wrapper fields

The wrapper SHALL expose: qualified option; action instance; action specification hash; configuration identity; qualification snapshot hash; current-state, transition, post-state, and qualification statuses; separate authority-basis reference; control and material-obligation status; monitoring; pause/shutdown; rollback; triggers; reviewer; evaluation time; execution disposition; and claim boundary.

### M.3 Blocking rule

Any mismatch, non-current qualification, invalid current state, non-admissible transition, non-viable post-state, active material trigger, missing control, or missing separate authority basis SHALL block ordinary execution and preserve the reason in the wrapped record.

### M.4 Non-overclaiming rule

A wrapped run MUST NOT claim execution qualification unless the referenced record is complete and consistent with the exact action and configuration. Wrapper conformance is not empirical validation, legal authority, physical-safety proof, or deployment certification.

### M.5 v5.6 release delta (retained baseline lineage)

Replaced a duplicative regime-token surface with Canon-aligned qualification continuity and exact action binding. Preserved the existing cascade and authority separation.

## Appendix N: Primitive Trace, Falsification Plan, and Claim Maturity Extension (v3.5)

### N.1 Scope

This appendix extends ripple.md for v3.5 wrapped deployments. It requires high-stakes L2/L3 episodes to document primitive-to-observable traceability when the episode is used for validation, public assurance, institutional adoption, or autonomous execution.

### N.2 New required fields for high-stakes L2/L3

The Decision Note or Assurance Preconditions Record SHALL include: PrimitiveTraceRecord; FalsificationPlan; ClaimMaturityLevel; ObservableConsequenceRegister; ToleranceRule; RefusalCondition; PostHocModificationLog; and EvidenceDowngradeRule.

### N.3 IND and Refusal mapping

If the primitive-to-observable chain cannot produce a unique projection or cannot identify an observable consequence, the affected test SHALL be marked IND and, in wrapped RippleLogic deployments, routed to Refusal where the uncertainty is gate-critical or claim-critical.

## N.4 Claim maturity labels

Permitted labels are CONCEPTUAL, ARCHITECTURAL, WORKED\_RUN, REPLAYABLE, PILOTED, REPLICATED, DOMAIN\_CALIBRATED, DEPLOYED\_MONITORED, and PROOFPACK\_READY. A deployment MUST NOT use a higher label than its evidence supports.

## Appendix O: Outcome Observation and Requalification Extension (Normative for high-stakes L2/L3)

O.1 Purpose. The wrapper SHALL preserve a link from the executed action and frozen qualification to observed outcomes. Outcome observation updates the evidence and claim boundary; it does not retroactively prove the model.

O.2 Minimum fields. The wrapper SHALL expose the outcome observation id; action instance; qualified option; configuration; observation time/window; expected markers; observed outcomes; deviation; materiality; affected scopes and subgroups; rights, tail-risk, containment, control, and obligation signals; new unknowns; evidence source and uncertainty; reviewer; reopened stages; and claim effect.

O.3 Requalification rule. A material deviation, new harm, incident, configuration change, control or obligation failure, authority change, or new material unknown SHALL suspend inherited qualification and reopen every affected stage and downstream dependency. Under unclear dependency boundaries, use the more conservative reopening scope or Refusal.

O.4 Non-replacement. This record does not replace the RippleLogic Canon, SGP, domain evidence, or legitimate authority. It binds observation to governance learning and prevents silent post-hoc rescue.

## Appendix P: Measurement Maturity and Deployment Context Extension

Normative for L2/L3 wrapped deployments that make claim-bearing use of UCI, kernel propagation, HDW, PLSS, or biological SGP evidence.

### P.1 Required additional Assurance Preconditions fields

UCI\_Measurement\_Readiness: UCI-M0\_UNAVAILABLE | UCI-M1\_PROVISIONAL | UCI-M2\_DOMAIN\_SUPPORTED | UCI-M3\_VALIDATED.

Kernel\_Use\_Posture: K0\_NONE | K1\_STARTER | K2\_DOMAIN\_CALIBRATED | K3\_BACKTESTED | K4\_REGISTERED.

HDW\_Capture\_Audit: source, participation quality, conflict disclosures, challenger window, and independent review status.

PLSS\_Escalation\_Register: local stakeholder set, guardrail scopes, externality search, escalation triggers, non-escalation rationale, and challenger status.

Deployment\_Context\_Profile: LAB | PILOT | INSTITUTIONAL\_ADVISORY | INSTITUTIONAL\_OPERATIONAL | PUBLIC\_POLICY | AGENTIC\_EXECUTION | EMERGENCY\_TRIAGE.

Biological\_Evaluation\_Maturity when SGP biological claims are used: BEM-0 through BEM-4.

### P.2 Non-overclaiming requirement

A wrapped deployment MUST NOT claim validated UCI, validated kernel propagation, ProofPack readiness, biological measurement completeness, capture-resistant weighting, or PLSS local sufficiency unless the corresponding maturity and audit fields support that claim. If the field is absent in a high-stakes L2/L3 episode, the affected claim SHALL be scored IND and routed to Refusal, escalation, or narrower claim scope.

## Appendix Q: Machine-Readable Wrapper Roadmap and Schema Ownership Boundaries

LINEAGE (historical, superseded): v4.0 carried-forward material. At the time of this historical block, the anchors were RippleLogic Canon v12.6 and SGP v8.5; current anchors are declared in Appendix RELEASE. This standard remains the portable alignment contract and assurance wrapper around that canonical pair and does not supersede their governing roles.

In the historical v3.5 line, ripple.md was the portable assurance and wrapper contract. It records evidence, decision notes, update legitimacy, refusal conditions, and deployment assurance. It does not replace or supersede the governing Canon or SGP companion. This release is not an empirical-validation claim, not ProofPack-ready, not machine-verifiable-ecosystem-complete, and not Tier 4-ready. Stronger claims require separate schemas, validators, reference calculators, replayable PCC artifacts, calibration evidence, and independent replay instructions.

### Q.1 Schema Ownership and Wrapper Boundaries

ripple.md owns wrapper-level records, including Decision Note, Assurance Preconditions Record, waiver records, triage records, update-legitimacy records, wrapper refusal records, and wrapper assurance records. It also owns wrapper-level ReferenceStructureRecord pointer fields, including DN-18 and R\_reference\_structure, while consuming Canon-owned ReferenceStructureRecord semantics by reference where the Canon governs the underlying run.

ripple.md consumes Canon-owned PCC, ReferenceStructureRecord, OutcomeObservationRecord, ActionQualificationRecord, PrimitiveTraceRecord, FalsificationPlan, ObservableConsequenceRegister, and maturity-profile records by reference. It MUST NOT redefine Canon-owned objects.

ripple.md consumes SGP-owned moral-status, protection, participation, intelligence, and reality-management evidence/status objects by reference where moral-patient status is relevant. It MUST NOT redefine SGP MPS band rules, FPP semantics, GPR/SPR role surfaces, ICP or RMCP semantics, P100 criteria, or evidence classification.

### Q.2 Machine-Readable Wrapper Roadmap

Future machine-verifiable ripple.md releases SHOULD include schemas for Decision Note, Assurance Preconditions Record, waiver records, triage records, update-legitimacy records, wrapper refusal records, and wrapper assurance records. These schemas SHALL reference Canon-owned and SGP-owned records by id, version, and manifest hash rather than restating their fields.

#### Q.3 Boundary Note

Boundary note. Later generated schemas, templates, or validator surfaces must cite their owning artifact and version and do not independently expand this standard unless adopted through the standard's versioning procedure.

## Appendix R: v4.0 Evidence Status, Legitimacy, and Compact Profile Extension (Normative clarification)

### R.1 Release Delta (v3.4.1 → v3.5)

This patch preserves the wrapper model, conformance levels, assurance preconditions, Decision Note requirements, and wrapped RippleLogic cascade ordering. It adds evidence-status token expectations, public legitimacy disclosure, and compact profile boundaries for practical adoption.

## R.2 Evidence Status Field

Decision Notes and Assurance Preconditions Records SHOULD include an evidence\_status field for decision-critical inputs. Allowed tokens are MEASURED, ESTIMATED, ASSUMED, EXPERT\_JUDGMENT, CONTESTED, UNVALIDATED, NOT\_OBSERVED, and PLACEHOLDER\_FOR\_SENSITIVITY\_ONLY.

## R.3 Public Legitimacy Disclosure

Where a ripple.md-wrapped run is used for public governance, the Decision Note SHOULD disclose the lawful authority, affected-stakeholder participation method, contestability mechanism, and appeal/review pathway. A wrapped run is an assurance artifact, not a grant of legal authority.

## R.4 Compact Profile Boundary

Compact profiles may reduce documentation burden, but they MUST NOT bypass NCRC, TRC, CSV, Refusal, non-overclaiming rules, SGP authority boundaries, auditability, or material-change controls.

# Appendix S: v4.0 Evidence-Status and Companion-Integration Extension (Normative clarification)

This extension clarifies wrapper expectations for releases that include workbook companion sheets for evidence status, compact profiles, validation roadmaps, boundary-risk registers, or SGP split guardrails.

## S.1 Wrapper evidence-status expectation

For high-stakes L2/L3 episodes, decision-critical claims SHOULD expose whether inputs are MEASURED, ESTIMATED, ASSUMED, EXPERT\_JUDGMENT, CONTESTED, UNVALIDATED, NOT\_OBSERVED, or PLACEHOLDER\_FOR\_SENSITIVITY\_ONLY.

## S.2 Anti-theater integration check

A companion sheet that is named as supporting release readiness SHOULD be present, populated, and referenced by a sanity, PCC, verifier, or decision-note surface. Unreferenced companion sheets SHOULD be described as informative only.

## S.3 Public legitimacy boundary

The wrapper structures and audits reasoning; it does not create democratic legitimacy, lawful authority, regulatory compliance, or institutional permission by itself.

# Appendix T: Reference Tooling and Replay Boundary (v3.5) (Normative boundary)

In wrapped RippleLogic deployments, reference calculators, schemas, validators, Decision Note records, and PCC replay files support reconstructability and audit. They do not replace the canonical cascade or create legal, democratic, empirical, ProofPack, or Tier 4 authority. If the tooling output conflicts with the governing canon, the canon controls and the tooling artifact requires correction.

A wrapped run that cites tooling artifacts SHOULD identify the exact artifact filename, version, SHA-256 hash, and replay command or validation command. If those artifacts are not bundled or hash-pinned, the claim MUST remain prose-supported rather than machine-replayable.

Continuation note. Appendix U below is part of RLS-RFC-0001 v3.5 and completes the ReferenceStructureRecord integration for this release line.

## Appendix U: Reference Structure Record Extension (Normative for high-stakes L2/L3 and claim-bearing episodes)

For L2/L3 high-stakes or claim-bearing episodes, the Assurance Preconditions Record SHOULD include a ReferenceStructureRecord before deployment or consequence-bearing execution.

The record SHALL identify: the reference object used to evaluate the claim; the active reference layers; invariant constraints; causal pathway; evidence source; evidence status; falsification condition; adaptation boundary; authority boundary; refusal trigger; and artifact hash references.

If the ReferenceStructureRecord is missing, circular, self-validating, unfalsifiable, or unsupported by the declared evidence sufficiency rule, affected tests SHALL be scored IND and the episode SHALL NOT proceed under a strong conformance claim without waiver, triage, or narrowed claim scope.

Decision Note field addition: reference\_structure\_record. This field points to or summarizes the ReferenceStructureRecord used for the episode.

Assurance Preconditions Record addition: R\_reference\_structure. This field records the ReferenceStructureRecord identifier, hash, maturity status, and reviewer disposition.

Boundary rule. This extension does not replace the governing RippleLogic equations, rights-localization step, or canonical cascade. It strengthens the evidence basis and wrapper assurance surface used to support the claim.

### Appendix V: Release Metadata and Dependency Pins

This appendix preserves dependency pins, companion-document references, release-delta notes, and current-line compatibility statements moved out of the opening pages in v3.5 for reader accessibility.

Normative dependencies:

- RLS-RFC-0003 (RPAP) v0.1 - REQUIRED for L2/L3 where redaction or access controls apply to audit evidence (especially recourse evidence). Implementations MUST pin to a specific RPAP version in the Assurance Preconditions Record (Appendix G, field K).
- RLS-RFC-0002 (PFAP) v0.1 - OPTIONAL unless coercion authorization is in scope. If adopted, implementations MUST pin to a specific PFAP version in the Assurance Preconditions Record.

Dependency versions MUST be pinned in the Assurance Preconditions Record (Appendix G, field K). Implementation-defined thresholds and other IDPs MUST be set ex ante and recorded in the Assurance Preconditions Record (Appendix G, field V).

Core-bundle completeness note (Informative). A core canon bundle MAY omit RPAP/PFAP files when their triggering conditions are not in scope for the reviewed artifact set. Wrapper-level completeness for any path that uses redaction/access controls or coercion authorization is claimable only when the relevant companion protocol is bundled or hash-pinned and the dependency is recorded in the Assurance Preconditions Record.

Canonical compatibility note (Informative). This standard is currently wrapped against RippleLogic Canon v13.0 and SGP v8.8 for the present release line. It functions as the portable alignment contract and assurance wrapper around that canonical pair.

Current pinned companion matrix (Informative): RippleLogic Canon v13.0; SGP v8.8; ripple.md v5.8; RippleLogic Agent System v13.0; Ripple Aligners Sheet v5.9; WDBIP v1.9.

Companion documents (non-normative):

- docs/guides/RIPPLELOGIC\_IMPLEMENTATION\_SCAFFOLDING.md - step-by-step adoption checklist for L0-L3.
- RELEASE\_NOTES.md - version history and migration guidance.

- waiver\_note.md - template for documenting §8.4 waivers (Waiver Record).
- triage\_note.md - template for documenting §2.20 emergency triage records (Emergency Triage Record; DN-15).
- templates/ - machine-readable YAML templates for key artifacts (non-normative; see Appendix L).
- decision\_note.md - template for completing the required or recommended Decision Note for episodes classified under §2.16A and governed by §11.1.
- quick\_refusal.md - template for structured refusal and escalation responses in bounded agent/tool implementations.

Release Delta (v3.4.6 → v3.5)

LINEAGE (historical, superseded): v3.5 was a readability and front-matter reorganization update synchronized to RippleLogic v10.8 and SGP v5.6. It preserves the wrapper model, conformance levels, evidence semantics, IND-to-Refusal mapping, ProjectionPreRegistration requirements, ActionQualificationRecord requirements, PrimitiveTraceRecord, FalsificationPlan, ObservableConsequenceRegister, ClaimMaturityLevel, OutcomeObservationRecord requirements, and ReferenceStructureRecord integration while moving dense release metadata and dependency details behind the reader-facing opening.

Current-line note (v5.1). Canon anchors are RippleLogic Canon v13.0 and SGP v8.8. This standard remains the portable alignment contract and assurance wrapper around that canonical pair and does not supersede their governing roles.

## Appendix W: v4.0 Layer-Discipline and Computable-But-Inadmissible Wrapper Extension

### W.0A Computability vs Realizability Boundary

The wrapper MUST preserve the distinction between a computable artifact and a selectable decision. A Decision Note, model output, score, SGP evidence object, tool verdict, or dashboard value may exist and still be inadmissible, ungrounded, unauthorized, structurally non-viable, or non-executable. In wrapped RippleLogic use, computed artifacts become decision-relevant only through the declared RG, RF/NCRC, TRC, CSV, RLS, wrapper-obligation, and authority surfaces.

### W.1 Purpose

LINEAGE (historical, superseded): This carried-forward v4.0 extension originally aligned ripple.md with RippleLogic Canon v10.8 and SGP v5.6; in that historical v5.0/v5.5 lineage block, the then-active pins were Canon v12.6 and SGP v8.5; current v5.8 identity is controlled by Appendix RELEASE and the manifest. It adds wrapper fields for cases where a value, score, or test output exists but may not be used as a valid conformance, alignment, or authorization claim.

### W.2 Required wrapper treatment

For high-stakes L2/L3 episodes, Decision Notes SHOULD include a ComputableButInadmissible field whenever any score, model output, SGP evidence object, tool verdict, or dashboard value exists but is barred by a gate, evidence limitation, maturity boundary, or authorization boundary.

### W.3 Minimum fields

| Field           | Meaning                                                                  |
|-----------------|--------------------------------------------------------------------------|
| computed_object | Score, scalar, model output, tool value, or dashboard value that exists. |

| Field                  | Meaning                                                                                                                       |
|------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| inadmissibility_reason | Gate failure, evidence insufficiency, maturity boundary, reference defect, authorization defect, or claim-boundary violation. |
| layer                  | NCRC, TRC, CSV, RLS, UCI/HOI, SGP, PCC, Agent, or wrapper assurance layer.                                                    |
| allowed_use            | Debugging, training, research, sensitivity analysis, or audit illustration.                                                   |
| prohibited_use         | Selection, deployment, conformance claim, legal claim, ProofPack claim, or authority claim.                                   |
| review_status          | Unreviewed, reviewer-challenged, reviewer-accepted, or escalated.                                                             |

## W.4 Adaptive Reserve wrapper note

Where an episode is materially affected by institutional, operational, ecological, or social slack, ripple.md implementations SHOULD record whether an Adaptive Reserve review was performed and whether saturation risk was found. This wrapper note does not add a new canonical gate.

## W.5 v3.5 release delta

Release Delta (v3.4 → v3.5). v3.5 adds computable-but-inadmissible wrapper fields, layer-specific claim discipline, and Adaptive Reserve recording guidance. It remains a wrapper standard and does not supersede the Canon or SGP.

Appendix X: Decision Verdict, Authority-Selection, and Reality-Grounding Wrapper Extension (v4.0)

## X.1A Reality-grounding wrapper rule

For wrapped RippleLogic deployments, FrameworkVerdict and AuthoritySelectionRecord MUST be interpreted alongside RealityGroundingStatus. A non-decisive or insufficiently evidenced reality surface cannot be laundered into deterministic framework selection by adding an authority record.

### X.1 Purpose

LINEAGE (historical, superseded): This carried-forward v4.0 extension originally aligned the portable wrapper with RippleLogic Canon v10.8 decision-verdict semantics and the Reality Grounding and Lean Cascade Patch. It does not replace the Canon, alter the wrapped RippleLogic cascade, or create ProofPack/Tier-4 claims.

### X.2 Required distinction

For claim-bearing L2/L3 episodes, Decision Notes MUST distinguish FrameworkVerdict from AuthoritySelectionRecord. The wrapper MUST NOT present a human, institutional, democratic, expert, emergency, or operator selection as a deterministic framework selection when the underlying RippleLogic run is non-decisive or underdetermined.

### X.3 Minimum wrapper fields

Recommended fields: FrameworkVerdict, AdmissibleOptionSet, DeterministicSelectionClaim, NonDecisiveReason, AuthoritySelectionRecord, AuthorityBasis, ProjectionPreRegistrationHash, CoverageClosureStatus, CoverageStatusSummary, FalsificationClass, ReviewerRequired, and DisclosureText.

### X.4 Projection and closure wrapper notes

Where a Decision Note relies on declared scenarios, thresholds, weights, uncertainty parameters, SGP objects, or stakeholder mappings, for high-stakes L2/L3 or ProofPack-prep episodes, the wrapper MUST record whether those inputs were frozen before evaluation. Required blanks SHOULD be classified as BLANK\_BY\_DESIGN, DECLARED\_UNKNOWN, NOT\_APPLICABLE, or MISSING\_UNDECLARED rather than silently converted to zero.

#### X.5 v3.5.1 release delta

Release Delta (v3.5 → v3.5.1). v3.5.1 adds portable fields for decision-verdict semantics, authority-selection separation, projection-freeze evidence, coverage-status closure, and falsification-class recording. It preserves ripple.md's role as a wrapper standard and does not supersede the Canon or SGP.

Appendix X-2: v4.0 Root Reality Doctrine Patch (Informative). This patch aligns ripple.md wrapper terminology with the Canon root vocabulary: structure, constraint, transition, consequence, reality, decision-relevant reality, and reality surface. It preserves the wrapper role and adds no independent authority, validation, ProofPack, or Tier-4 claim.

End of carried-forward RLS-RFC-0001 v4.0 lineage section (Appendix V release metadata, Appendix W layer-discipline extension, Appendix X decision-verdict wrapper extension, and Appendix X-2 Reality Grounding and Lean Cascade Patch). Current v5.8 interpretation continues below and is controlled by Appendix RELEASE, Appendix AA, Appendix AB, and the current release delta.

## Appendix Y: v4.0 Reality Grounding and Lean Cascade Wrapper Extension

LINEAGE (historical, superseded): v4.0 originally synchronized ripple.md to RippleLogic Canon v10.8 and SGP v5.6; v5.0 synchronizes to RippleLogic Canon v12.6 and SGP v8.5.

The portable wrapper now treats Reality Grounding as the first claim-authority level and records structural diagnostic role to prevent gate-relevant UCI/HOI evidence from being deferred to residual tie-break.

The wrapper remains ProofPack-prep unless schemas, reference calculator, machine-readable test vectors, semantic validator, replay instructions, and independent replay evidence are bundled or hash-pinned for the claimed release.

## Appendix Z: Category Grounding and Primitive Audit Wrapper Extension

### Z.1 Status and placement

This appendix is a normative wrapper extension for material claim-bearing L2/L3 episodes. It does not create a sixth RippleLogic method level. Category Grounding belongs inside Reality Grounding when category choice affects the claim boundary.

### Z.2 Required wrapper distinction

A Decision Note MUST NOT treat a category as valid merely because it appears in a source, benchmark, policy, model output, legal label, stakeholder preference, institutional process, or prior implementation. Such materials may be evidence-bearing but are not automatically category-validating.

### Z.3 Minimum wrapper fields

Recommended fields: category\_grounding\_required, material\_category, category\_role, primitive\_basis, category\_definition, category\_discriminator, dependency\_position, inclusion\_boundary, exclusion\_boundary, common\_confusions, noncollapse\_pairs, reference\_structure, evidence\_surface, source\_authority\_status,

category\_maturity\_status, semantic\_debt\_flag, falsification\_or\_revision\_trigger, term\_refusal\_condition, reviewer\_status, required\_claim\_action, and category\_failure\_class.

## Z.4 Failure classes

Allowed category\_failure\_class values SHOULD include CATEGORY\_BOUNDARY\_FAILURE, CATEGORY\_REFERENCE\_FAILURE, CATEGORY\_SCOPE\_FAILURE, CATEGORY\_METRIC\_MISMATCH, SEMANTIC\_DRIFT\_DETECTED, TERM\_DISCRIMINATION\_FAILURE, DEPENDENCY\_POSITION\_ERROR, PRIMITIVE\_CHALLENGE, DATA\_FAILURE, PROJECTION\_FAILURE, PARAMETER\_FAILURE, OPERATOR\_FAILURE, GOVERNANCE\_FAILURE, and EXTERNAL\_CHANGED\_CONDITION.

## Z.5 Non-overclaiming rule

A wrapper record may show that a decision process was auditable and that category or term uncertainty was disclosed. It does not prove the category is correct, does not validate empirical measures, does not convert certification into validity, does not certify legal compliance, and does not make Tier 4/ProofPack claims available.

## Z.6 Category Grounding patch boundary

This appendix strengthens the portable wrapper against optimizing inherited categories. It keeps Category Grounding, Primitive Audit, Reference Category Review, and Category-Reality Correspondence Check in one Reality Grounding group, preserving the five-level method and avoiding duplicate governance layers. This carries forward the ripple.md v4.0 Category Grounding Patch through Appendix Z; current v5.8 carries it forward through Appendices AA-AB and adds consequence-tempo and responsibility-continuity hardening.

# Appendix AA: v5.0 Structural Viability, Term Discrimination, and Do-Not-Collapse Wrapper Extension

## AA.1 Status and placement

This appendix is a normative wrapper extension for material claim-bearing L2/L3 episodes. It does not create a sixth RippleLogic method level and does not supersede the Canon. Term Discrimination belongs inside Category Grounding, which belongs inside Reality Grounding.

## AA.2 Required wrapper distinction

A Decision Note MUST NOT treat a term as operationally valid merely because it appears in a policy, source, benchmark, legal label, model output, institutional workflow, certification artifact, stakeholder preference, or prior implementation. Such materials may be evidence-bearing or authority-bearing in limited ways, but they do not automatically validate the term, its boundary, or its claim strength.

## AA.3 Minimum wrapper fields

When material, Decision Notes SHOULD include: term\_discrimination\_required, term, term\_definition, category\_discriminator, dependency\_position, noncollapse\_pairs, source\_authority\_status, evidence\_surface, reviewer\_status, term\_refusal\_condition, required\_claim\_action, and term\_failure\_class.

## AA.4 Failure classes

Allowed term\_failure\_class values SHOULD include TERM\_DISCRIMINATION\_FAILURE, DEPENDENCY\_POSITION\_ERROR, CERTIFICATION\_VALIDITY\_COLLAPSE, PERMISSION\_ADMISSIBILITY\_COLLAPSE, MONITORING\_CONTROL\_COLLAPSE, AUTOMATION\_AUTONOMY\_COLLAPSE, PROTECTION\_AUTHORITY\_COLLAPSE,

SCORE\_TRUTH\_COLLAPSE, COMPLIANCE\_CORRECTNESS\_COLLAPSE, and EXECUTION\_AUTHORITY\_OVERCLAIM.

## AA.5 Non-overclaiming rule

A wrapper record may show that term risk was disclosed, but it does not prove that the term is true, validate empirical measures, certify legal compliance, authorize deployment, or make Tier 4/ProofPack claims available. Certification is evidence of a declared assurance process, not validity itself.

## AA.6 Compact YAML shape

```
term_discrimination:
  required: true
  term: ""
  term_definition: ""
  category_discriminator: ""
  dependency_position: ""
  noncollapse_pairs: []
  source_authority_status: ""
  evidence_surface: ""
  reviewer_status: ""
  term_refusal_condition: ""
  required_claim_action: ""
  term_failure_class: ""
```

## AA.7 v5.0 release delta

Release Delta (v4.0 -> v5.0). v5.0 adds wrapper support for Structural Viability and falsification-discipline hardening while carrying forward v4.0 wrapper fields and failure classes for term discrimination, dependency position, and do-not-collapse risks. It preserves ripple.md's role as a portable assurance wrapper and does not supersede the Canon, SGP, Agent System, or Aligners Sheet.

HISTORICAL LINEAGE - NOT CURRENT SPEC. The versions named in the passages below were the pins at the time those passages were written. Current active pins are stated in the Appendix RELEASE and release manifest; historical passages preserve provenance only and MUST NOT be read as active pins.

Wrapper precedence clarification. ripple.md adds wrapper obligations that may further constrain execution, assurance, or evidence requirements, but it cannot weaken, reorder, or bypass the Canon cascade, RF/NCRC, TRC, CSV, or RLS boundaries.

END HISTORICAL LINEAGE. Appendix AB and later current extensions are active normative material governed by the v5.8 Appendix RELEASE and current release manifest.

# Appendix AB: v5.1 Consequence-Tempo and Responsibility-Continuity Extension

## AB.1 Status and placement

This appendix is a normative wrapper extension for high-stakes L2/L3 episodes and for any episode where rapid propagation, material irreversibility, automation, or multi-party handoffs make the extension material. It does not create a sixth Canon gate. Consequence tempo is evaluated primarily through Containment and Structural Viability evidence, with TRC reopened if the tempo analysis reveals a catastrophic, irreversible, or lock-in pathway not already represented. Responsibility continuity is an assurance and audit obligation that may block execution or claims when ownership, authority, or repair duties become untraceable.

## AB.2 Consequence-tempo rule

A high-stakes action **MUST NOT** execute at a rate that outruns meaningful detection, interpretation or review, intervention, containment, and rollback. The record **MUST** identify the earliest unacceptable adverse escalation threshold or irreversible stop-loss boundary and compare it with the credible control path under non-ideal operating conditions. A nominal kill switch, review channel, or rollback plan is insufficient when it cannot act before the relevant unacceptable escalation or stop-loss boundary.

Where a deliberate interval is necessary to preserve review, challenge, or multi-key authorization before irreversible execution, the implementation **SHALL** record constitutional latency. The latency **MUST** be proportionate, bounded, and paired with a harm-of-delay comparison. Where delay creates the greater rights or tail-risk harm, emergency handling may shorten or bypass ordinary latency only through the pinned emergency rules, explicit scope limitation, higher authorization, monitoring, expiry, and post-hoc review. Emergency use does not convert a failed or indeterminate tempo record into an ordinary Pass.

## AB.3 Responsibility-continuity rule

The decision and execution chain **MUST** retain identifiable ownership for candidate generation, evidence warrant, interpretation, selection, authorization, execution, intervention or revocation, appeal and communication, and repair. Each material handoff **MUST** state what artifact or decision was transferred, what authority was transferred, what authority was retained, what uncertainty remained, and how the receiving role accepted, challenged, or escalated it.

No implementation may use organizational complexity, procurement, outsourcing, model output, tool execution, committee process, or operator deference to produce an unowned decision. The following collapses are prohibited: candidate generation into evidence warrant; evidence warrant into authority; authorization into physical or causal admissibility; model recommendation into human judgment; monitoring into control; and distributed participation into dissolved accountability.

### AB.2A Trigger and critical-path rule

A full ConsequenceTempoRecord is **REQUIRED** only when rapid propagation, material irreversibility, consequential automation, significant lock-in, delayed detectability, or control-response limits could alter rights, TRC, CSV, execution authorization, or requalification. Lower-risk episodes **MAY** record `TEMPO_NOT_MATERIAL` with rationale and a reopen trigger.

Timing **SHALL** be modeled as a dependency-aware critical path rather than an automatic arithmetic sum. Each material timing element **SHALL** record minimum credible, expected, and worst credible bounds; evidence basis; uncertainty; serial or parallel relation; and whether it depends on a human, automated sentinel, vendor, network, or external authority. Parallel fail-safe or safe-state paths **MAY** shorten the critical path only where tested evidence supports the claim.

Where an action necessarily produces its intended effect before runtime interruption is possible, the absence of an interruption window does not itself make the action inadmissible. The assurance burden moves *ex ante*: bounded scope and rate, domain evidence, fail-safe design, least-harm comparison, independent authorization proportionate to stakes, and explicit residual-risk acceptance are required.

### AB.3A Typed responsibility roles

A ResponsibilityContinuityRecord **SHALL** distinguish at least:

- `decision_authority`: who legitimately selected the action;
- `evidence_accountability`: who warranted the evidence and claim boundary;
- `model_analytical_responsibility`: who selected, configured, or interpreted the analytical method or model;
- `operational_responsibility`: who configured and operated the system;

- **execution\_authority**: who could initiate execution;
- **intervention\_authority**: who could pause, override, revoke, isolate, or shut down;
- **appeal\_owner**: who must process affected-party challenges;
- **communication\_owner**: who must disclose incidents, uncertainty, and material changes;
- **remedy\_owner**: who must restore, compensate, remediate, or repair;
- **residual\_responsibility**: who remains answerable when another participant or contractor fails.

Legal liability SHALL NOT be inferred from this record unless separately established under applicable law.

### AB.3B Material-obligation integrity

Whenever selectability, execution, intervention, review, appeal, communication, repair, monitoring, rollback, or requalification depends on a material obligation, the ResponsibilityContinuityRecord SHALL include a material\_obligations entry. A named owner is insufficient unless the carrier has an evidenced authority and capacity basis. Each entry SHALL state the obligation, accountable authority, operational carrier or authenticated execution path, trigger, scope, required action, response window, evidence of discharge, review or expiry, challenge route, delegation acceptance, backup or successor, nonperformance trigger and escalation, amendment/waiver/suspension/retirement authority, change-control rule, residual responsibility, obligation status, and effectiveness status.

Carrier absence, silence, incapacity, conflict, missed deadline, or unaccepted handoff SHALL not count as discharge. It SHALL activate the declared escalation, substitution, pause, safe-state, or requalification path. A material control change is a configuration change. The constrained system or agent SHALL NOT authorize its own waiver or silently weaken the control.

### AB.3C Human-compensation-load record

Before automation, scaling, outsourcing, or role removal, the record SHALL assess whether current functioning depends on undocumented overrides, exception handling, reconciliation, memory, relationship work, after-hours recovery, weak-signal interpretation, informal escalation, or recurrent prevention of visible failure. Where material, record the carrier role classes, hidden work and workflow gap, burden and recovery indicators, single-person dependency, consequence if compensation stops, evidence sources, privacy and non-retaliation controls, preservation or redesign action, and review trigger. A bounded human contribution may remain part of the design; hidden, involuntary, saturated, or non-substitutable load requires controls, capacity restoration, redistribution, redesign, escalation, or refusal.

## AB.4 Minimum records

When triggered, the Decision Note MUST include or reference:

1. A ConsequenceTempoRecord containing the fields in DN-10A and Appendix L.2C.
2. A ResponsibilityContinuityRecord containing the fields in DN-10B and Appendix L.2C.
3. The applicable implementation-defined tempo-compatibility method and responsibility-chain reconstructability target from Appendix G field V.
4. Evidence of at least one drill, replay, simulation, or justified operational demonstration for high-stakes L2/L3 execution claims, unless infeasible and formally escalated.

## AB.5 Dispositions and refusal mapping

Allowed tempo\_disposition values are COMPATIBLE, COMPATIBLE\_WITH\_CONTROLS, INCOMPATIBLE\_REDESIGN, EMERGENCY\_PROVISIONAL, and INDETERMINATE.

Allowed reconstructability\_status values are COMPLETE, PARTIAL\_WITH\_CONTROLS, INCOMPLETE\_ESCALATE, and NOT\_APPLICABLE\_WITH\_RATIONALE.

For ordinary high-stakes execution, `INCOMPATIBLE_REDESIGN`, `INDETERMINATE`, or `INCOMPLETE_ESCALATE` MUST block execution or the stronger execution/conformance claim until the deficiency is resolved. `COMPATIBLE_WITH_CONTROLS` and `PARTIAL_WITH_CONTROLS` require named, binding controls and monitoring. `EMERGENCY_PROVISIONAL` is available only under the pinned emergency rules and is not an ordinary conformance Pass.

## AB.6 Non-overclaiming rule

A completed tempo or responsibility record shows that the relevant questions were documented and reviewed. It does not prove that timing estimates are empirically correct, that a shutdown path will succeed under all conditions, that an obligation was effectively carried, that control performance produced the intended protective outcome, that every contributor is legally liable, that deployment is lawful, or that the system is safe. Stronger claims require domain evidence, tested controls, lawful authority, outcome evidence, and independent replay proportionate to stakes.

## AB.7 v5.1 release delta

Release Delta (v5.0 -> v5.1). v5.1 adds consequence-tempo proportionality, bounded constitutional latency, responsibility continuity, two conditional Decision Note records, new drift and threat tags, implementation-defined parameter hooks, and machine-readable template fields. It also corrects the duplicate numbering in the Required Artifacts section. It preserves the five-level Canon cascade, the seven-test wrapper suite, the Rights Floor, TRC, CSV, RLS, SGP type boundaries, and all prior non-overclaiming rules.

## AB.8 v5.2 release delta

Retained baseline lineage. Release Delta (v5.1 -> v5.2). v5.2 corrects the tempo boundary from the first material effect to the earliest unacceptable adverse escalation or irreversible stop-loss boundary; requires a dependency-aware worst-credible critical path rather than a naive serial sum; narrows full-record triggers; distinguishes actions with no technically available runtime interruption window; types responsibility roles; and clarifies that constitutional latency is an assurance term rather than a claim of legal authority. It also synchronizes to RippleLogic Canon v12.7, SGP v8.6, WDBIP v1.7, and the MathGov run-record v3 interface. No gate, welfare dimension, SGP threshold, or RLS equation is added or changed. Current component and run-record compatibility is governed by the candidate Appendix RELEASE and current interface clauses.

## Methodological integrity boundary

RLS calculations rank only residual welfare after RG, RF/NCRC, TRC, and CSV. RLS MUST NOT be used to rescue a claim whose assumptions, definitions, thresholds, indicators, or evidence surfaces are `METHOD_UNSUPPORTED`, `TEST_REQUIRED`, `CONTESTED`, `REVISION_REQUIRED`, or `REFUSE_STRONGER_CLAIM`. If an upstream methodological dependency changes, the affected scoring inputs must be recalculated or the score must be marked non-decisive, exploratory, or refused for stronger claims.

Cross-domain non-substitution. Scientific or engineering evidence does not by itself create legal authority or normative legitimacy. Legal, regulatory, institutional, certification, compliance, or policy authority does not by itself establish empirical truth, causal validity, technical capability, physical safety, or reduced uncertainty. Commercial or anthropomorphic language creates no warrant by itself. A material cross-domain inference requires the B.ab bridge fields; an unsupported bridge MUST NOT support selection, authority, execution, conformance, or public assurance.

## Physical execution wrapper boundary

A ripple.md Decision Note for consequence-bearing physical systems must separate governance permission from physical admissibility. The wrapper may record approvals, certifications, monitoring, and authority, but it must not treat those records as proof that a physical action is safe.

For physical execution, the Decision Note must name the PC-AEP evidence source, validity domain, failure modes, reversibility boundary, monitoring/shutoff path, residual unknowns, and claim action. If admissibility is not established, the note must say so plainly and cannot claim safe execution.

## Physical execution Decision Note field

Decision Notes involving material physical or cyber-physical execution SHOULD include `physical_execution_claim_status` with one of: `PHYSICAL_ADMISSIBILITY_SUPPORTED`, `GOVERNANCE_PERMISSION_ONLY`, `PHYSICAL_ADMISSIBILITY_NOT_ESTABLISHED`, or `PHYSICAL_ADMISSIBILITY_CONTRAINDICATED`. This field is a claim-boundary field, not a new gate.

Decision Note field block (conditional; physical/cyber-physical execution).

- `physical_execution_claim_status` (conditional): one of `PHYSICAL_ADMISSIBILITY_SUPPORTED`, `GOVERNANCE_PERMISSION_ONLY`, `PHYSICAL_ADMISSIBILITY_NOT_ESTABLISHED`, or `PHYSICAL_ADMISSIBILITY_CONTRAINDICATED`.
- `candidate_generation_source`: model, planner, optimizer, agent, human, or other source that generated the candidate action.
- `admissibility_warrant_source`: external/domain evidence source used to warrant physical or causal admissibility.
- `PC_AEP_reference`: link/hash/id for the PC-AEP record where triggered.

This block is a claim-boundary record. It is not a sixth gate and does not turn governance permission into physical safety proof.

## 2.20A RippleLogic Tail-Emergency Firewall (Normative for wrapped RippleLogic use)

When this standard wraps a RippleLogic run, generic emergency triage does not weaken or replace the RippleLogic Tail-Risk Constraint. A Test 3 / TRC failure may proceed only if every Tail Emergency Mode prerequisite in the pinned RippleLogic Canon Section 8.8 is satisfied, including explicit no-action and delay comparison, redesign attempts, independent challenge, an absolute maximum exposure bound, higher approval, minimized scope and duration, monitoring and shutoff, remedy/redress, expiry, and return-to-normal criteria. The resulting state is `TAIL_EMERGENCY_PROVISIONAL_ACTION`, not `TRC_PASS`, ordinary alignment, ordinary selectability, or proof of safety. If the Canon prerequisites are not met, the wrapped run must delay, redesign, escalate, or refuse.

## SGP v8.8 RMCP/P100 wrapper fields (Normative synchronization)

Where a wrapped decision uses reality-management claims, the Decision Note MUST reference the exact SGP v8.8 RMCP record id, target/collective boundary, domain, time window, dimension vector and uncertainty, enabling scaffolds, causal scale, continuity evidence, P100 status or non-status, independent review, capacity-performance distinction, and revision trigger. The wrapper MUST reject any use of RMCP/P100 as consciousness proof, moral-worth ranking, human-rights downgrade, CMIU membership, lawful mandate, or authority grant.

## Reproducible run-record compatibility

The controlling reading specification is

Core\_15/MATHGOV\_REPRODUCIBILITY\_AND\_USE\_STANDARD\_v1.7.docx. A ripple.md Decision Note MAY embed or reference a `mathgov_run_record` object in the v4 family, with the v4.1 external schema pin specified by Reproducibility v1.7. This is a family-name cross-reference, not a claim that the external schema is bundled or that an unpinned v4 variant is interchangeable. When it does, the record MUST satisfy the deterministic ordering and non-compensation rules in the pinned MATHGOV Reproducibility and Use Standard v1.7. The historical semantic-source path `docs/reproducibility/MATHGOV_REPRODUCIBILITY_AND_USE_STANDARD_v1.5.md` remains an external

baseline reference; candidate source/mirror parity is not established by this package. A wrapper cannot repair a missing parameter, failed gate, absent authority basis, or incomplete evidence trace. Passing the run validator establishes record/cascade conformance only.

### **AB.9 v5.3 release delta**

v5.3 synchronized the portable wrapper to the then-current MathGov release line and strengthened current-versus-lineage labeling, source hierarchy, and machine-interface references. It did not add a Canon gate or alter rights, TRC, CSV, or RLS semantics.

### **AB.10 v5.5 release delta (Lineage)**

v5.5 adds consequence-tempo compatibility and responsibility-continuity records, binds qualification to declared configuration and material change, updates exact current companion pins, and clarifies that policy, monitoring, approval, and risk acceptance cannot replace domain evidence or make an unknown operating envelope safe. These are wrapper and assurance hardenings only; the governing cascade remains RG -> RF/NCRC -> TRC -> CSV -> RLS.

# APPENDIX RELEASE: Identity, Source Authority and Revision Record

Framework release: MathGov/RippleLogic v13.0. Component: ripple.md Standard v5.8. Edition-origin preparation: 10 September 2026; the separately identified correction build is dated below. Two-part major.minor component versions are used; preserved historic identifiers are not renumbered.

Status: integrated research and teaching specification with a bounded worked-example and scoped reference implementation. Readiness is limited to the checks in Reports/Verification\_and\_Readiness.md. No empirical validation, independent human validation, full production runtime, physical-safety certification, legal authority, deployment authorization or Tier-4 ProofPack status is implied by the edition number.

Delivery identity: MathGov/RippleLogic v13.0, build MG-RL-13.0-20260926-RELEASE-I. Prepared 26 September 2026 from the supplied Release H publication package. Only reproduced errata and necessary source/evidence synchronization were patched; component editions are unchanged. The manifest identifies each component's exact bytes and source build. Live publication is not asserted.

Component/build identity. The manifest distinguishes this correction build from unchanged Release G and H masters. Aligners Sheet v5.9 keeps its numerical inputs, formulas, cached results and worked verdict; text and print-area errata are itemized in Verification/Final\_I/Exact\_Workbook\_Changes.json. A filename or component edition alone is not exact file identity.

Verification boundary: Reports/Release\_I\_Verification.md and Verification/Final\_I/ identify the checks performed on this package and their limits. Prior-build receipts do not certify changed bytes. Cache-independent replay and spreadsheet-application recalculation are separate evidence surfaces. Microsoft Excel parity, full external-registry conformance, empirical validation and production authorization are not asserted.

Package source authority: Core\_15 contains 14 DOCX specification masters and one XLSX Aligners frozen worked-run workbook. DOCX files control prose, equations and tables subject to Canon and SGP ownership. The workbook controls only its disclosed exemplar surfaces. Sources, Reading\_HTML and Reading\_PDFs are generated reading projections, not competing normative masters. The manifest and hashes identify exact bytes; supplemental code and schemas govern only their documented subset. Conflicts require recorded correction against the controlling source.

License and reuse: consult the package-level LICENSE and NOTICE and retain applicable component-specific and third-party notices. This pointer does not override a valid exception or grant rights over separately supplied private or third-party review material.

External repository publication and any absent legacy schema/validator/registry are separate authorities/evidence surfaces. The local package does not claim to update a remote repository or reproduce unavailable implementations. Current source pointers in this appendix replace prior front-matter edition pointers for this package; accurate historical source references below are retained as lineage only.

Preserved invariants: RG -> RF/NCRC -> TRC -> CSV -> RLS; rights non-compensation and unallocated rights effects; existing seven rows and seven dimensions; existing SGP MPS/FPP/GPR/SPR/ICP/RMCP separation; probability ownership, uncertainty, stability and every-contender rules; selection, authority and execution separation.

| Component                   | Current edition |
|-----------------------------|-----------------|
| RippleLogic Canon           | v13.0           |
| Sentience Gradient Protocol | v8.8            |
| ripple.md Standard          | v5.8            |
| RippleLogic Agent System    | v13.0           |

| Component                                                       | Current edition |
|-----------------------------------------------------------------|-----------------|
|                                                                 |                 |
| CSV Gate Standard                                               | v2.7            |
| RippleLogic Cascade Standard                                    | v2.9            |
| MathGov Reproducibility and Use Standard                        | v1.7            |
| Welfare Dimension Boundary and Interaction Protocol             | v1.9            |
| RLS Validation Protocol                                         | v2.9            |
| RippleLogic Foundations Primer                                  | v4.7            |
| MathGov Public Introduction                                     | v13.0           |
| Physical/Causal Admissibility Evidence Profile                  | v2.6            |
| Methodological Falsifiability and Dependency Integrity Standard | v2.6            |
| Source-Coupling Integrity Standard                              | v2.6            |
| RippleLogic Aligners Sheet                                      | v5.9            |

Change locations and rationales: Reports/Release\_I\_Adjudication.md; exact edits: Verification/Final\_I/Exact\_Document\_Changes.json and Exact\_Workbook\_Changes.json. Navigation and preservation records are separate. The untouched prior publication and original feedback are preserved in the complete provenance archive.

Current proposals are not universal truths. Taxonomy maximality, continuous-time propagation, cross-substrate cardinal welfare, generic susceptibility/shield formulas, autonomous recovery, vendor infrastructure, zero-knowledge circuits and hardware meshes remain unvalidated unless independently demonstrated under a scoped implementation profile.

Preserved Baseline Release Material (Historical; Non-Controlling)

Retained upstream repository identifier: ripplelogic/ripple-md-standard. This retained identifier is not evidence that the repository currently hosts these bytes or that it is the same namespace as MathGov/ripple-logic. Publish and verify the chosen mirror against the current manifest.

The following blocks are relocated intact from the recovered baseline. Their edition numbers, release-readiness wording and external-source limitations describe that historical candidate, not current component identity or newly executed verification. Governing current metadata is the matrix above.

v5.7-rc2 Historical Release Integration

HISTORICAL (NON-CONTROLLING): Release: MathGov Core Release 2026.09 — RippleLogic Canon v12.8-rc2 / SGP v8.7-rc2 — Rights Non-Dilution, Representation Integrity, and Robust Selection Release

Historical companion pins: Canon v12.8-rc2; SGP v8.7-rc2; ripple.md v5.7-rc2; Agent System v12.7-rc2; CSV v2.6-rc2; Cascade v2.8-rc2; Reproducibility v1.6-rc2; WDBIP v1.8-rc2; RLS Validation v2.8-rc2; Primer v4.6-rc2; Public Introduction v12.8-rc2; PC-AEP/MFDI/Source-Coupling v2.5-rc2; Aligners Sheet v5.8-rc1.

| Release control | Historical value            |
|-----------------|-----------------------------|
| Component       | ripple.md Standard v5.7-rc2 |

| Release control      | Historical value                                                                                                                                                                                                                                                                                                       |
|----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Release              | HISTORICAL (NON-CONTROLLING): MathGov Core Release 2026.09 — RippleLogic Canon v12.8-rc2 / SGP v8.7-rc2                                                                                                                                                                                                                |
| Architecture         | RG → RF/NCRC → TRC → CSV → RLS                                                                                                                                                                                                                                                                                         |
| Role                 | Portable assurance wrapper and claim/evidence packaging standard subordinate to the Canon.                                                                                                                                                                                                                             |
| Claim boundary       | Candidate Tier 1–3 research specification; readiness is bounded by the accompanying verification report. No empirical validation, source/mirror parity, implementation conformance, legal authority, physical-safety certification, deployment authorization, Tier 4, or moral truth is established by this candidate. |
| Source/render parity | Versioned semantic source and DOCX/PDF mirrors must agree. Filename, internal version, active pins, manifest, and hashes must agree; mismatch is release-integrity failure.                                                                                                                                            |

## Historical Release Integration (v5.7-rc2)

HISTORICAL (NON-CONTROLLING): This release candidate advances ripple.md Standard from v5.6 to v5.7-rc2 and binds it to MathGov Core Release 2026.09 — RippleLogic Canon v12.8-rc2 / SGP v8.7-rc2. Historical references remain lineage only; unmarked candidate references use the v12.8-rc2 component matrix. This candidate corrects claim-specific evidence precedence and the preregistration exception summary while preserving the Canon cascade and wrapper protections.

### v5.2.1 Patch Note (Lineage)

v5.2.1 repairs the current companion matrix and wrapper-versus-cascade terminology. It does not change consequence-tempo, responsibility-continuity, assurance, waiver, or verification semantics.

Version: v5.7-rc2 (Control-Critical-Path, Responsibility-Continuity, and Core-Interface hardening)

Source date: 2026-07-18; candidate: 2026-09-05

Supersedes: RLS-RFC-0001 v5.2.1

## Release alignment

Component version: v5.7-rc2. Current package synchronization: RippleLogic Canon v12.8-rc2 / SGP v8.7-rc2.

HISTORICAL (NON-CONTROLLING): This companion document is synchronized with MathGov Core Release 2026.09 v12.8-rc2. The public cascade is RG -> RF -> TRC -> CSV -> RLS; the formal shorthand is RG/RSG -> RF/NCRC -> TRC -> CSV -> RLS. RF is the public Rights Floor layer and NCRC is its formal non-compensatory predicate. CSV means Containment and Structural Viability. Detailed cascade definitions are maintained in the Canon, the RippleLogic Cascade Standard, and the CSV Gate Standard; this file applies those invariants without repeating the full release note.

HISTORICAL (NON-CONTROLLING): Current synchronization note. ripple.md v5.7-rc2 is the portable assurance-wrapper standard synchronized in this release line to RippleLogic Canon v12.8-rc2 and SGP v8.7-rc2. The MathGov Core Release 2026.09 package line applies the Physical Admissibility Release while

carrying forward Term Discrimination and Semantic Stability controls, and decouples package naming from component versions. ripple.md does not supersede either governing artifact.