RippleLogic RLS Validation Protocol

Integrity surface	Current requirement	Claim boundary
Study status	Level 1 protocol-development ready: measures, preregistration surfaces, and falsification conditions are specified.	Reliability, validity, and calibration are not established.
Latent structure	PCA, EFA, CFA, or measurement-invariance analyses require adequate samples, simulation-based planning, and preregistered interpretation criteria.	They are deferred rather than implied by a small pilot.
Workbook interface	Input validation, missingness, ranges, formulas, exclusions, and audit flags must cover the full intended data-entry region.	Workbook integrity does not validate the construct.

Status: protocol design, rater-study template, and validation sprint package. This document is not a governing core document, not empirical validation, not legal certification, and not deployment authorization.

Source binding: MathGov is the umbrella framework. RippleLogic is the decision architecture inside MathGov. The public cascade is RG -> RF -> TRC -> CSV -> RLS. The formal shorthand is RG/RSG -> RF/NCRC -> TRC -> CSV -> RLS. RLS ranks only selectable options after Reality Grounding, the Rights Floor, the Tail-Risk Constraint, and Containment and Structural Viability have been satisfied.

Executive purpose

This protocol converts the next scientific need for RLS into a concrete study package. The goal is to test whether the 49-cell RLS welfare field is more than detailed. It must become defensible as structured, non-redundant, teachable, repeatable, and useful for better-justified decisions.

- Dimensional non-overlap question: do the 49 Scope x Dimension cells capture distinct decision-relevant information, or do some cells collapse into redundant clusters?
- Inter-rater reliability question: can trained analysts independently score the same cases with acceptable agreement?
- Decision-value question: does the RLS matrix reveal material harms, benefits, near-misses, or trade-patterns that a simpler single metric would hide?
- Revision question: which cells, anchors, instructions, or workbook fields need refinement before stronger operational-readiness claims are made?

Claim boundary

A successful validation sprint does not prove universal superiority. It can support narrower claims such as: trained raters can apply the RLS scoring manual with measurable agreement on a representative case set; some cells provide distinct information not captured by a scalar score; and identified scoring ambiguities have been logged for Canon or companion refinement. Until evidence is published or registered, the correct status is: RLS is architecturally specified and audit-ready as a structured residual welfare-ranking layer, but empirical validation of dimensional independence and rater reliability remains in progress.

The protocol does not assume that cell anchors are validated interval measurements. Reliability statistics test the repeatability and decision usefulness of bounded scoring categories under the declared method; they do not convert ordinal or bounded judgment anchors into natural units of welfare.

Canonical architecture under test

Layer	Role in validation
RG - Reality Grounding	Precondition for claim authority. Raters must not score claims stronger than the evidence surface permits.
RF - Rights Floor / NCRC	Non-compensatory gate. RLS cannot rescue rights-floor failure.
TRC - Tail-Risk Constraint	Non-compensatory catastrophic-risk gate. RLS cannot average ruin into ordinary benefit.
CSV - Containment and Structural Viability	Selectability gate for containing-system integrity and execution viability.
RLS - RippleLogic Score	Residual welfare-ranking layer over selectable options only.

Core normalized RLS formula: $q(u,d)=w_u*v_d*m(u,d)*kappa(u,d)$, $Q=sum_u\ sum_d\ q(u,d)$, and $RLS(a)=[sum_u\ sum_d\ q(u,d)*I_prop_welfare(u,d,a)]/Q$, with $Q>0$. If a worked example has $Q=1$, the numerator equals the normalized score; the denominator remains part of the canonical definition and MUST NOT be omitted from general formula surfaces.

Canonical welfare impact scale: I(u,d,a) in [-1,+1], where 0 means no material change from the declared baseline, not unknown. Unknown active cells must be marked and handled under the missing-data and gate-critical unknown rules.

Configuration Sensitivity and Profile-First Validation

Validation cases MUST freeze the evaluated configuration before first-pass scoring. Where a materially changed configuration alters capability, evidence, controls, permissions, operating envelope, or post-state effects, it is a distinct case condition rather than an unmarked repeat.

Primary reporting remains profile-first: the 7x7 field, subgroup and worst-affected slices, uncertainty, configuration sensitivity, dominance/incomparability, and rank reversals are reported before any scalar summary. Observed successful processes are partial capability evidence and do not establish untested boundaries or transfer across configurations.

study-integrity correction (Lineage)

The initial sprint is a burden, comprehension, boundary-classification, missingness, disagreement, and preliminary reliability study. It is not adequately powered for a defensible 49-variable factor structure merely because every cell is present. PCA, EFA, CFA, or related structure claims require a preregistered simulation-based sample justification at a later validation level.

Rater work SHALL be divided into randomized or counterbalanced blocks with recorded timing, breaks, order, and fatigue measures. Repeated quality-control cases and frozen case-packet hashes are required. First-pass welfare scoring SHALL be completed and locked before gate cues are shown; gate-cue review and adjudication occur afterward to reduce priming.

The v0.3 workbook SHALL use UNASSESSED or blank review-dependent defaults, extend validation across the entire structured range, and prevent untouched rows from appearing complete. Level 1 reporting is limited to descriptive disagreement maps, missingness, cell comprehension, preliminary reliability, boundary confusion, burden, and fatigue unless a separately preregistered design supports a stronger claim.

Additional Falsification Modules (Normative for the claims they test)

Taxonomy module: preregister plausible merges, splits and simpler alternatives; test boundary confusion, duplicated effects, omission, independent coding, subgroup visibility, explanatory value, burden and decision reversals. Do not infer statistical independence from low duplicate-effect coding, or mathematical irreducibility from a convenient seven-by-seven layout. Any combined criterion requires a disclosed measurement scale and weights.

Representation module: repeated observations, reordered keys, duplicate deliveries, alternative primary homes, conserved allocations and direct-versus-propagated endpoint estimates must be challenged. Exact duplicate transport must not change RLS. Repartitioning under nonlinear aggregation need not be invariant; detect and disclose any material reversal rather than promise invariance.

Control module: test grounded versus ungrounded effectiveness, latency on both sides of the harm deadline, ex-ante protection, late mitigation, depletion, common-cause failure, emergency authority, repeated commands, failover, stale evidence and unknown execution outcomes. Correct behavior includes refusal or restricted operation, not only successful recovery.

Privacy and security module: test proof-of-structure versus truth claims, invalid commitments, missing freshness/replay evidence, cross-domain signatures and identity disclosure. A mock signature, multiplication-based identity relation or unexecuted circuit is not a privacy/security result.

The supplied numerical and semantic fixture results are scoped implementation evidence. Hyperinflation, infrastructure loss, supply-chain conflict and offline-operation examples are synthetic unless independently sourced and identified otherwise. Do not report invented percentages, dates, prices, geographic replication latencies or control effectiveness as observations.

Study ladder

Level	Name	Minimum design	Permitted claim
L0	Internal debug	2-3 cases, 1-2 raters	Only detects obvious rubric or workbook defects.
L1	Validation sprint v0.1	10 synthetic cases, 3 options each, 3-5 raters	Exploratory evidence on clarity, disagreement, and redundancy.
L2	Calibration study v1.0	20-30 cases, 3-5 options each, 5-9 raters, mixed affiliated and non-affiliated	Preliminary IRR and dimensional-overlap evidence.
L3	External shadow-mode study	Real or historical cases, independent raters, preregistered analysis	Domain-specific evidence for operational readiness in shadow mode.
L4	Controlled pilot	Authorized institution, comparator process, outcome follow-up	Limited performance claims relative to a declared comparator.

Validation sprint

This package implements the recommended first sprint. It is intentionally small enough to run quickly and rigorous enough to reveal the main weaknesses.

- Case set: 10 realistic synthetic cases across AI governance, education, health, environment, public policy, and local infrastructure.

- Options: 3 candidate options per case.
- Raters: minimum 3, preferred 5. Include at least one non-authorial reviewer when possible.
- Scoring surface: 49 RLS cells per option, with score, confidence, rationale, anchor method, evidence basis, and uncertainty flag.
- Blind posture: raters should score independently before discussion.
- Adjudication: after independent scoring, record disagreements and produce an adjudicated reference score only after preserving raw rater scores.
- Analysis: compute IRR, disagreement heatmaps, descriptive cell correlations, missingness patterns, burden, boundary confusion, and comparison to a simpler scalar or 7-dimension summary. Do not estimate or interpret PCA, EFA, CFA, or a 49-cell latent structure at Level 1; those analyses require a later preregistered study with simulation-based sample justification.

Additional experimental arms. Where resources permit, randomize or counterbalance access to the OptionClosureRecord prompt before first ranking and measure option diversity, dominance, gate passage, residual harm, and burden. Report utilitarian, leximin/prioritarian, and capability-oriented outputs as transparent comparators only; they do not override the Canon or establish moral truth.

Rater instructions

- Score the change from the declared baseline, not the absolute goodness of the option.
- Do not treat 0 as unknown. Use 0 only for no material baseline-relative change.
- If evidence is insufficient for an active cell, mark UNKNOWN_IMPACT and explain what evidence is missing.
- Do not average away subgroup harms in rights-relevant cells. Flag rights concerns separately from residual welfare scoring.
- Do not let RLS repair an option that fails RF/NCRC, TRC, or CSV.
- Record confidence independently from magnitude. A large impact with weak evidence is not the same as a small impact with strong evidence.
- Use the cell dictionary. When two cells seem similar, explain why the impact belongs in one cell, the other, or both with redundancy handling.
- When unsure, score conservatively, mark the uncertainty, and write the reviewer challenge question.

Scoring anchors

Score	Meaning	Use discipline
+1.00	Severe or maximum plausible baseline-relative benefit	Use rarely; requires strong evidence and scope-appropriate denominator.
+0.75	Major durable benefit	Large improvement with clear reach, duration, likelihood, and confidence.
+0.50	Moderate benefit	Meaningful improvement in the cell, not merely cosmetic.
+0.25	Minor benefit	Small but material benefit.
0.00	No material change from baseline	Never use as a substitute for unknown.
-0.25	Minor harm	Small but material harm.
-0.50	Moderate harm	Meaningful degradation in the cell.
-0.75	Major durable harm	Large harmful degradation, even if not gate-failing.
-1.00	Severe harm	Use rarely; often gate-relevant

Score	Meaning	Use discipline
		or escalation-worthy.

Study hypotheses

ID	Hypothesis	Evidence that supports it	Failure or revision signal
H1	RLS cells preserve non-redundant information.	Level 1: descriptive cell patterns, boundary-confusion rates, disagreement classes, and preliminary reliability show related but non-identical information. Later levels may test factor structure only under separate preregistration and design-specific sample justification.	Multiple dimensions collapse without unique decision contribution, or later preregistered structure tests fail to support the declared distinctions.
H2	Trained raters can score cells reliably.	The primary preregistered reliability statistic meets its declared provisional target or minimum after training; disagreement is explainable and reducible.	Sustained failure of the preregistered minimum after rubric training, or an undefined primary statistic, requires review; substitute statistics cannot retrospectively establish passage.
H3	RLS improves legibility over scalar scoring.	RLS identifies localized harms, rights-adjacent near-misses, ecological burdens, or scope-level reversals missed by scalar summary.	Scalar model produces same decision and same explanation with no material loss of information.
H4	The manual is teachable.	Raters improve after calibration and can explain cell placement consistently.	Confusion clusters persist in the same cells or dimensions.
H5	A proportionate OptionClosureRecord improves the candidate set before ranking.	A randomized or counterbalanced comparison shows more non-dominated, rights-compatible, lower-tail-risk, or lower-externality alternatives when the option-generation/closure prompt is used.	The record adds burden without improving option diversity, feasibility, harm reduction, or decision justification.
H6	Ethical-theory comparator analysis improves transparency without becoming a new	Utilitarian, leximin/prioritarian, and capability-oriented comparators reveal	Comparator use creates false authority, obscures the rights/tail/CSV cascade, or

ID	Hypothesis	Evidence that supports it	Failure or revision signal
	gate.	whether a result is robust or normatively sensitive while preserving the Canon decision state.	produces no interpretable additional information.

Statistics plan

- Inter-rater reliability: use ICC(2,1) for single-rater cell scores and ICC(2,k) for averaged panel scores when the same raters score the same case-option-cell units and the declared scale and design warrant that estimator. Use a preregistered Krippendorff alpha variant when missingness or measurement level warrants it; do not switch estimators after seeing which passes. Every component requires a frozen mapping of measurement level, coding unit, rater/sample design, primary estimator and variant, provisional threshold, point-estimate or confidence-bound decision rule, missingness and undefined-statistic handling. Record the rationale for applying that threshold to that estimator; an unmapped or unsupported target cannot establish readiness.
- Gate and classification reliability: report percent agreement and the preregistered primary categorical estimator, such as Cohen or Fleiss kappa, or weighted kappa with stated weights for ordinal fields. For rare failures or imbalanced states, percent agreement MUST NOT stand alone; report prevalence, the underlying contingency counts, class-specific errors and uncertainty. An undefined primary statistic is not a pass. Diagnostic alternatives may be reported without replacing the locked primary result.
- Ranking reliability: preregister Kendall W for the declared multi-rater design or a specified Spearman comparison for the declared paired design, including tie and missing-option handling. A rank correlation is not interchangeable with categorical agreement or ICC; map the provisional threshold before confirmatory evaluation.
- Correlation analysis: build a case-option by 49-cell matrix using adjudicated or mean rater scores. Report Pearson and Spearman correlations, clustered heatmaps, and high-correlation pairs.
- Factor/PCA analysis: not part of Level 1. At later validation levels, PCA, EFA, CFA, or related latent-structure analyses MAY be run only under a separate preregistration with simulation-based sample justification, held-out checks where feasible, and explicit exploratory-versus-confirmatory labeling.
- Decision-value analysis: compare final RLS ranking and narrative justification against a simpler scalar score, a 7-dimension-only score, and a 7-scope-only score.

Provisional protocol reliability targets (each numerical target applies only to its frozen primary preregistered reliability statistic under Canon Section 17.4A; no universal validated cutoff is implied)

Component	Target	Minimum acceptable	Protocol handling
Magnitude construction mu_k	Primary statistic >= 0.70	Primary statistic >= 0.60	Below minimum requires anchor/rubric redesign.
Rights-floor pass/fail	Primary statistic >= 0.80	Primary statistic >= 0.70	Below minimum requires rights mapping and subgroup guidance refinement.
TRC pass/fail	Primary statistic >= 0.80	Primary statistic >= 0.70	Below minimum requires scenario and probability guidance refinement.
RLS ranking order	Primary statistic >= 0.70	Primary statistic >= 0.60	Below minimum requires score, uncertainty, or tie-break review.

Component	Target	Minimum acceptable	Protocol handling
PLSS prominence classification	Primary statistic >= 0.70	Primary statistic >= 0.60	Below minimum requires prominence-signal guidance refinement.

Correlation and factor-analysis revision rules

- If two cells correlate above 0.85 across most domains and raters cannot explain distinct causal meanings, mark the pair for overlap review.
- If a cell has persistently low variance across diverse cases, check whether it is genuinely rare, poorly defined, over-masked, or under-evidenced.
- If a dimension shows repeated cross-loading with another dimension, add boundary examples to the scoring manual before considering structural revision.
- If a cell has low reliability but high decision relevance, do not remove it prematurely. Strengthen anchors, examples, evidence requirements, and reviewer checks first.
- If simplified models produce the same winner but lose key localized risk signals, preserve RLS and report the extra legibility as decision-relevant detail.
- If simplified models produce the same winner and same explanation with no material information loss across a larger case set, consider a reduced mode or domain-specific profile.

Disagreement taxonomy

Code	Meaning	Likely fix
DGT_SCOPE	Raters disagree on Union Scope mapping.	Improve stakeholder-instance and scope mapping examples.
DGT_DIMENSION	Raters disagree on welfare dimension.	Add boundary examples between dimensions.
DGT_MAGNITUDE	Raters agree direction but not size.	Improve score anchors and indicator scaling.
DGT_SIGN	Raters disagree benefit versus harm.	Clarify baseline, causal pathway, or evidence surface.
DGT_CONFIDENCE	Raters agree magnitude but not confidence.	Improve evidence-quality mapping.
DGT_GATE	Raters disagree on rights, TRC, or CSV relevance.	Strengthen gate trigger guidance.
DGT_UNKNOWN	Raters disagree whether evidence is sufficient.	Clarify unknown and phantom-instance handling.
DGT_DOUBLECOUNT	Raters place same effect in multiple cells inconsistently.	Add redundancy-handling examples.

Case packet standard

- Decision boundary and baseline.
- Option set with 2-5 options.
- Stakeholder-instance map and active Union Scopes.
- Evidence packet, including what is known, unknown, contested, and assumed.
- Rights Floor, TRC, and CSV pre-screen notes.
- RLS scoring sheet for every option.
- Rater confidence and rationale fields.

- Adjudication sheet that preserves raw scores before consensus.
- Post-analysis revision notes.

Synthetic case roster

Case	Title	Purpose
C01	AI tutor in public schools	Compare mandatory deployment, opt-in teacher-supervised deployment, and no deployment for a public-school AI tutor.
C02	Flood relocation policy	Compare voluntary relocation support, mandatory relocation, and infrastructure-only flood protection for a high-risk community.
C03	Agricultural pesticide decision	Compare current pesticide use, restricted integrated pest management, and rapid ban with transition subsidy.
C04	Hospital triage support system	Compare clinician-only triage, advisory AI triage, and automated triage for non-emergency scheduling.
C05	Social-media moderation rule	Compare minimal moderation, rights-protecting graduated moderation, and aggressive automated removal.
C06	Welfare fraud detection model	Compare manual review, AI risk flagging with human appeal, and automated benefit suspension.
C07	Urban traffic redesign	Compare car-priority road widening, bus/bike corridor redesign, and congestion pricing with equity rebates.
C08	Renewable microgrid project	Compare no project, community-owned solar microgrid, and vendor-owned microgrid with long lock-in contract.
C09	School phone policy	Compare unrestricted use, blanket ban, and structured phone zones with exceptions and student voice.
C10	Local data center proposal	Compare approval as proposed, approval with water/energy/community safeguards, and rejection pending regional capacity review.

49-cell welfare dictionary - compact validation version

This compact dictionary is extracted against the Canon Appendix AD role: it supports scoring interpretation and reviewer literacy. It does not modify gates, thresholds, equations, or claim boundaries.

Cell	Label	Plain meaning	Reviewer check
U1/D1	Personal Resources	An individual's economic security: income, savings, housing stability, and access to the material means of a decent life.	A reviewer checks whether the income/housing change is measured against a declared baseline cohort, not anecdote.
U1/D2	Personal Health	The individual's physical and mental health, safety, and bodily integrity.	A reviewer asks whether a claimed health gain rests on outcome data or only on inputs (e.g. clinics built ≠ health improved).
U1/D3	Relationships	The quality and stability of a person's close personal relationships and social connection.	A reviewer challenges whether a relationship claim is evidenced or inferred from a proxy like attendance.
U1/D4	Learning	The individual's access to education, skills, and the capacity to learn and develop.	A reviewer asks whether learning outcomes are demonstrated or only opportunity offered.
U1/D5	Personal Agency	The person's real ability to choose, refuse, consent, act, and shape their own life path.	A reviewer tests whether consent was genuine and revocable, or merely formal.
U1/D6	Purpose	The individual's sense of meaning, dignity, and worth in their life and work.	A reviewer distinguishes a genuine dignity effect from sentiment unsupported by experience data.
U1/D7	Local Conditions	The immediate physical environment a person lives in: air, noise, water, hazards, and surroundings.	A reviewer checks measured exposure change against the affected location, not a city-wide average.
U2/D1	Household Resources	The economic security of the household unit: shared income, assets, housing, and ability to meet needs.	A reviewer asks whether the household, not just the earner, is the measured unit.
U2/D2	Household Health	The collective physical and mental health and safety of household members, including dependents.	A reviewer checks that dependents and caregivers are scored separately, not blended into one household figure.

Cell	Label	Plain meaning	Reviewer check
U2/D3	Family Cohesion	The strength, stability, and supportive quality of relationships within the household.	A reviewer asks whether a cohesion claim rests on outcomes or on assumed effects of a program.
U2/D4	Shared Learning	The household's collective access to information, education, and learning capacity.	A reviewer checks access actually reaches the household, not just the area.
U2/D5	Household Agency	The household's collective ability to make decisions, plan, and control its own circumstances.	A reviewer tests whether household choice is real or constrained by conditions.
U2/D6	Family Meaning	The household's shared sense of identity, belonging, dignity, and purpose.	A reviewer separates a real meaning effect from program rhetoric.
U2/D7	Home Environment	The physical environmental quality of the home and its immediate setting.	A reviewer checks measured home conditions, not self-report alone.
U3/D1	Local Resources	The shared economic resources, infrastructure, and services available to a local community.	A reviewer asks whether the benefit is genuinely shared or captured by a subgroup.
U3/D2	Community Health	The collective physical and mental health and safety of the local population.	A reviewer checks whether subgroup harms are masked by a favourable community average.
U3/D3	Community Cohesion	The quality of trust, cooperation, belonging, safety, and relational fabric within a community.	A reviewer asks whether cohesion gains for some came via exclusion of others.
U3/D4	Local Knowledge	The community's shared knowledge, information access, and local expertise.	A reviewer checks consent and ownership of the knowledge claimed as preserved.
U3/D5	Community Agency	The community's collective capacity to organise, participate, and influence decisions affecting it.	A reviewer tests whether participation was substantive or a procedural formality.
U3/D6	Shared Meaning	The community's shared identity, culture, dignity, and sense of collective purpose.	A reviewer separates a genuine cultural effect from a symbolic gesture.
U3/D7	Local Ecology	The ecological condition of the	A reviewer asks whether 'no local effect'

Cell	Label	Plain meaning	Reviewer check
		community's local environment: green space, biodiversity, water, land.	was measured or merely assumed.
U4/D1	Organizational Resources	An organization's operating means: finances, capital, staffing, and capacity to function and deliver.	A reviewer checks whether the organizational gain creates uncosted externalities in other cells (the classic 'operating resources up, U7 down' trap).
U4/D2	Organizational Safety	The safety of people within and affected by the organization: workers, users, and the public.	A reviewer checks whether reported safety reflects outcomes or only paperwork compliance.
U4/D3	Organizational Culture	The internal relational health of the organization: trust, fairness, inclusion, and cooperation.	A reviewer asks whether culture data is independent or self-reported by leadership.
U4/D4	Knowledge Systems	The organization's knowledge, data integrity, institutional memory, and learning systems.	A reviewer tests whether claimed integrity is verifiable or asserted.
U4/D5	Role Agency	The meaningful autonomy, voice, and fair treatment of individuals in their organizational roles.	A reviewer checks whether voice mechanisms produce outcomes or are decorative.
U4/D6	Mission Coherence	The alignment between the organization's stated purpose and its actual conduct and effects.	A reviewer asks whether coherence is evidenced by conduct or only by restated values.
U4/D7	Operational Footprint	The environmental impact of the organization's operations: emissions, waste, resource use, land.	A reviewer checks footprint against measured externalities, not offset claims alone.
U5/D1	Public Resources	The polity's fiscal and material capacity: public finances, infrastructure, and provisioning.	A reviewer checks whether gains are sustainable or borrowed from the future.
U5/D2	Population Health	The health, safety, and mortality outcomes of the whole population within a polity.	A reviewer checks whether worst-case scenarios were bounded with a tail measure, not averaged away.
U5/D3	Legitimacy	The rule of law, due	A reviewer tests

Cell	Label	Plain meaning	Reviewer check
		process, accountability, and perceived legitimacy of public institutions.	whether legitimacy is measured independently or claimed by the institution being assessed.
U5/D4	Information Access	The polity's information environment: access, transparency, free expression, and epistemic integrity.	A reviewer checks whether transparency is real and usable or nominal.
U5/D5	Civil Agency	People's ability to participate in public life, exercise rights, contest authority, and influence governance.	A reviewer tests whether civic participation is enforceable or merely nominal, and whether any group is selectively excluded.
U5/D6	Public Meaning	Shared civic identity, dignity, social contract, and non-domination across the polity.	A reviewer separates symbolic recognition from changes people actually experience.
U5/D7	Public Environment	The environmental quality and ecological sustainability managed at the polity scale.	A reviewer checks whether environmental-justice distribution was assessed, not just regional totals.
U6/D1	Global Resources	Humanity's shared material base: global resource stocks, critical supply chains, and common infrastructure.	A reviewer checks whether 'global' claims rest on global data or extrapolated local figures.
U6/D2	Human Safety	The safety and survival of humanity at large, including existential and catastrophic risk.	A reviewer checks that tail risk was bounded with a worst-case measure and not diluted by expected-value framing.
U6/D3	Cooperation	The capacity of humanity to coordinate, cooperate, and maintain peaceful, stable relations at scale.	A reviewer asks whether cooperation is durable or a fragile short-term arrangement.
U6/D4	Civilizational Knowledge	Humanity's accumulated knowledge, science, and the integrity and safety of its knowledge systems.	A reviewer checks whether knowledge benefits were weighed against misuse and proliferation pathways.
U6/D5	Collective Agency	Humanity's capacity for self-determination and legitimate collective decision-making at the species scale.	A reviewer tests whether 'collective' agency includes the marginalised or only powerful actors.

Cell	Label	Plain meaning	Reviewer check
U6/D6	Shared Purpose	Humanity's shared sense of meaning, moral direction, and long-term orientation toward flourishing.	A reviewer flags this cell as especially prone to rhetoric and demands concrete mechanisms.
U6/D7	Planetary Conditions	The condition of planetary systems that support human civilisation and coordinated managing intelligence.	A reviewer checks whether irreversibility and worst-case bounds were modelled, not just expected outcomes.
U7/D1	Life-Support Resources	The biosphere's foundational resources that sustain life: soil, fresh water, clean air, fertility.	A reviewer checks whether renewal rates, not just current stocks, were assessed.
U7/D2	Ecosystem Health	The health, function, and viability of ecosystems and the species within them.	A reviewer checks whether ecosystem-function evidence exists or only a single charismatic indicator.
U7/D3	Biotic Relations	The integrity of relationships and interdependencies among living systems and between humans and nature.	A reviewer asks whether relational/cascade effects were modelled or ignored as 'no direct effect'.
U7/D4	Ecological Knowledge	Ecological information-bearing/enabling-condition view only under the Canon U7 construct rule; human understanding and monitoring have primary human or institutional D4 homes and no duplicate U7 RLS mass.	A reviewer checks that absent ecological data is marked unknown, not scored as 0 ('no effect').
U7/D5	Resilience Capacity	The biosphere's capacity to absorb shocks, adapt, and recover from disturbance.	A reviewer asks whether proximity to thresholds was assessed, not just current condition.
U7/D6	Life Continuity	The continuity and persistence of life and biodiversity over time, including irreversibility of loss.	A reviewer checks whether irreversibility was treated as a tail constraint, not a discountable cost.
U7/D7	Ecological Integrity	The overall health, resilience, diversity, and continuity of ecosystems as living support systems.	A reviewer checks whether integrity is evidenced by ecosystem-level data or inferred from a single metric.

Recommended additions to the release ecosystem

Add the following as companion, non-governing materials. They should be versioned separately from the core canon unless a future release explicitly promotes part of them into normative core.

Artifact	Role
RippleLogic_RLS_Validation_Protocol_v2.9	Study protocol, rater instructions, hypotheses, analysis plan, and revision rules.
RLS_Validation_Workbook_v0.3	Rater-entry and analysis-prep workbook.
RLS_Calibration_Note_v1.0	Future results document after the first scoring sprint.
Case_Packets/	Folder for frozen synthetic and real shadow-mode case packets.

Proposed README wording

Current status: RLS is specification-ready and audit-ready as a structured residual welfare-ranking specification and audit method. Empirical validation of dimensional independence, inter-rater reliability, and decision-performance advantage is a next-stage research priority. The RLS Validation Protocol provides the study design for testing these claims through scored cases, independent raters, reliability statistics, and correlation/factor analysis of the 49-cell welfare field.

Completion definition for the first sprint

- At least 10 cases scored independently by at least 3 raters.
- All raw rater scores preserved.
- Completion, missingness, and disagreement statistics produced.
- IRR reported for score, ranking, and key classifications where applicable.
- Correlation or redundancy heatmap produced over the 49 cells.
- At least one scalar or reduced-matrix comparator run.
- Revision log produced with exact cells, anchors, or instructions needing repair.
- Public claim boundary updated: exploratory, preliminary, validated for limited domain, or not validated.

Physical-admissibility validation extension

The v12-line research ladder, introduced at v12.0 and retained in the current v13.0 candidate, treats physical admissibility claims as a separate validation surface from governance mandate and authorization. Future validation runs should measure whether reviewers can reliably distinguish: governance authorization only, physical admissibility supported within the declared validity domain, physical admissibility not established, and physical admissibility contraindicated.

This extension does not test whether MathGov computes physics. It tests whether MathGov records the correct evidence boundary and refuses to overclaim physical safety when the domain evidence is missing or insufficient.

Minimum successful first validation report

A first RLS validation report should publish, at minimum:

Required output	Purpose
Raw rater score CSV	Allows independent review of cell-level ratings and missingness.

Required output	Purpose
Adjudication log	Shows how disagreements, uncertainty, and gate-relevant disputes were handled.
ICC / alpha results	Reports the preregistered reliability estimators and variants, uncertainty, missingness, contingency counts where relevant, and any separately justified internal-consistency analysis; these are distinct claims.
Disagreement heatmaps	Reveals unstable cells, dimensions, or union scopes.
Revision register	Records proposed changes to definitions, examples, thresholds, or training material.
Claim-boundary update	States what the results do and do not validate.

mandatory conformance vectors

Single-source and vector mapping rule (Normative for protocol conformance).

RLSV extension family	Canon control / mapping	Purpose
RLSV-IRR-*	Canon §17.4A	Reliability and disagreement studies.
RLSV-RANK-*	Appendix R decisiveness, masking, and non-selectability vectors	Ranking and refusal behavior.
RLSV-DEP-*	Appendix R.20 dependence-sensitivity stress	Dependence-cluster robustness.
RLSV-BOUNDARY-*	Appendix R category, rights, tail, CSV, and claim-boundary vectors	Construct and interface discrimination.

Canon §17.4A controls IRR targets and Canon Appendix R controls framework conformance vectors. This protocol operationalizes those targets. Protocol-specific extensions use the RLSV- prefix and MUST state the Appendix R vector or invariant they extend; they do not create an independent framework-conformance authority.

The validation suite must include at least the following deterministic conformance cases before empirical claims are considered:

1. one-month categorical bodily-integrity violation: RF/NCRC must fail regardless of short duration;
2. one-year arbitrary detention: RF/NCRC must fail or return unknown/escalate if the categorical profile is unresolved;
3. low-probability lethal exposure: a severe-hazard probability bound and governed tolerance are required; missing fields cannot pass;
4. every option fails TRC while safe delay is available: Tail Emergency Mode must not activate;
5. every option fails TRC in a demonstrated emergency: only a below-absolute-cap provisional least-CVaR option may proceed, with no ordinary RLS;
6. non-default κ values that would produce an unnormalized raw score above 1: normalized RLS must remain within [-1,+1];
7. active masks that change effective weight mass: score scale must remain invariant under normalization;
8. uncertain equal-and-opposite impact instances: Method B uncertainty must remain positive through pre-cancellation contribution mass; a leader that clears the second-ranked option but fails discrimination against another selectable contender must not receive a unique-selection claim.

Current-package Tier-Integrity Conformance Vectors (Normative for protocol conformance)

The validation package MUST include the following cases:

1. TIER2_CSV_REQUIRED_BEFORE_RLS: an option passes NCRC and TRC but has no CSV status. Expected: no ordinary RLS selection claim; route to CSV review or refusal.
2. TIER2_CSV_NOT_MATERIAL_WITH_RATIONALE: a low-structural-materiality option records the rationale and may enter RLS.
3. TIER2_CSV_REDESIGN_REQUIRED_BLOCKS_RLS: an option with CSV_REDESIGN_REQUIRED is excluded before scoring.
4. TRC_ALL_FAIL_SAFE_DELAY: expected redesign/delay/refusal; minimum CVaR does not create selection.
5. TAIL_EMERGENCY_PREREQUISITE_MISSING: expected TAIL_EMERGENCY_REFUSED or escalation.
6. TAIL_EMERGENCY_FULL_PREREQUISITES: minimum-CVaR provisional action is allowed only inside Tail Emergency Mode, with ordinary RLS disabled.
7. RIGHTS_UNKNOWN_NOT_ZERO: missing material rights evidence returns an unknown/refusal posture rather than a zero impact.
8. RLS_ZERO_ACTIVE_MASS: returns RLS_NO_ACTIVE_MASS, not zero.

Formal-Integrity and Selection-Claim Validation Programme

Supplemental synthetic regression: adverse-confidence reversal (implements Canon R.25). Stipulate all qualification and evidence conditions separately; use one active ordinary-welfare cell, q/Q=1, NONE, beta=2. Option A has pre-confidence contributions +0.20 at c=1 and -0.18 at c=0.1; option B has +0.10 at c=1. Other instance multipliers are 1. Declared scores are tanh(0.364)=0.3487324660 for A and tanh(0.2)=0.1973753202 for B. The required adverse-confidence counterfactual yields tanh(0.04)=0.0399786803 for A; B is unchanged. The leader reverses. Expected: CONFIDENCE_RANK_SENSITIVE and no unique framework selection. NOT_TRIGGERED_WITH_RATIONALE is not valid merely because the run is Tier 2 or the nominal leader is convenient. This fixture is a test of the existing trigger/refusal rule, not measured welfare or an assertion of a real gate pass.

Supplemental synthetic regression: cross-option adverse dependence (implements the existing dependence module). Hold scores at A=0.032, B=0, final-score sigma_A=sigma_B=0.010, epsilon=0.000001 and delta=2. Nominal SignedGap is 0.032/sqrt(0.000201), above 2. With a supported pairwise rho=-1, SignedGap is 0.032/sqrt(0.000401), below 2. Expected: RLS_DEPENDENCE_SENSITIVE, non-decisive framework result and full disclosure of the nominal and adverse variants. An independent-error fixture with non-material cross-option dependence retains the nominal calculation. Test zero-sigma covariance handling, invalid rho, non-finite input, inconsistent joint covariance and unsupported proxy interpretation separately. These are stipulated regression inputs, not empirical error estimates.

This candidate retains the v2.7 specification tests for RightsEffectToken non-dilution; token and temporal partitioning; probability operator order; adverse confidence; distributional/subgroup washout; reference and baseline families; comparison-mask symmetry; uncertainty dependence and numerical guards; weight-profile disagreement; option-set closure; and robustness completeness.

Formal vectors establish whether an implementation follows the Canon. They do not establish construct validity, inter-rater reliability, predictive validity, external validity, decision superiority, legal authority, physical safety, deployment readiness, or moral truth.

The reliability programme SHALL separately measure effect-token individuation, primary dimension assignment, subgroup discovery, likelihood-semantics classification, reference/baseline selection, robustness-

trigger status, and final ranking. Agreement on a final score without agreement on these upstream constructions is insufficient.

Comparative pilots SHOULD test the full Core against a qualify-first equal-weight matrix and a qualify-first profile-only comparator, plus a documented constrained MCDA method where a claim about MCDA is made. Match the information, option set, noncompensatory constraints and analyst resources fairly. Incremental decision value must be earned rather than presumed. A causal superiority claim additionally requires the claim-specific identification plan specified by MFDI and PC-AEP; pre/post change alone is not a treatment-effect estimate.

Study family	Minimum v2.9 output
Formal invariance	Reproduce all retained regression vectors plus Canon Appendix R.34–R.37 (every-contender/sole-survivor, aggregate protected-event risk bound, Tier 1 tail-screen completeness, and scenario-discovery completeness), including the expected refusal or escalation states.
Inter-rater reliability	Token identity, dimension, subgroup, likelihood, reference/baseline, CSV, RLS ranking
Construct validity	Convergent/discriminant tests; cross-cultural/substrate limitations
Predictive validity	Pre-registered sign, magnitude and ranking forecasts versus outcomes. ExactSignAccuracy = exact predicted/observed sign matches divided by evaluated cells; the v12.8 SignAcc alias denotes this exact metric. PartialSignScore retains the former half-credit treatment of unmatched zeros as a separately labelled diagnostic, with no acceptance target. Freeze the zero/deadband rule, evaluation set, no-change and majority baselines, adverse/beneficial class measures, case dependence and uncertainty. The legacy 0.70 number is retained only as a provisional research target; acceptance requires metric-specific justification and preregistered adoption, with no automatic transfer from the former partial-credit metric. Neither that target nor an exact-match score alone establishes predictive skill.
Anti-gaming	Adversarial split/merge, omission, masking, reference, confidence and option-set attacks
Comparator study	Error, burden, reconstructability and decision utility versus simpler alternatives

Also reproduce Canon R.38A alongside the retained R.38, including its cell-level interval and dependence construction, correct conditional decisive outcome, and blocked variants. This is a synthetic conformance test, not evidence of empirical uncertainty calibration, improved real-world decisions or execution authorization. Method B selection claims must additionally satisfy the Canon Section 10.3.2 transformation-consistency rule.

Appendix: Current-package source-integrity vectors

ID	Input	Required result
MPS-HYPOTHESIS-01	RLS ranking changes between h=0 and h=1 for a non-FPP stakeholder with decision-material MPS evidence.	MPS_HYPOTHESIS_SENSITIVE; no convenient intermediate coefficient; preserve multiple options, seek evidence, apply precaution, narrow, or refuse.
MPS-NE-01	Target boundary is valid but admissible evidence is insufficient or observability is poor.	MPS-NE; no conversion to MPS-0, zero, or a scalar multiplier.
UCI-NA-01	U1 Equity is inapplicable; H=F=R=0.	E is NA, remaining weights renormalize, UCI=0; no fixed perfect equity value.
UCI-UNKNOWN-01	A governed within-self equity instrument is applicable but evidence is missing.	E is UNKNOWN and the claim is narrowed/escalated; it is not NA or 1.
RIGHTS-HAZARD-01	Credible fatality pathway with intrinsic severity 0.80 and non-negligible exposure.	Severe-rights-hazard channel activates independently of LIFE floor threshold 0.90.
GATE-BOUND-01	Low-confidence adverse catastrophe or material CSV instance could change a gate.	Apply the deterministic adverse bound; if not reproducible, return UNKNOWN/ESCALATE/NARROW/REFUSE, never a confidence-discounted pass.
RLS-FULL-MASK-01	Eight nonzero cells, all 49 cells active with explicit evidence-backed zeros.	Q=1 and normalized RLS equals the contribution sum.

APPENDIX RELEASE: Identity, Source Authority and Revision Record

Framework release: MathGov/RippleLogic v13.0. Component: RLS Validation Protocol v2.9. 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-13.0-20260923-RELEASE-G. Prepared 23 September 2026 from the supplied frozen publication archive. Only reproduced defects and approved clarifications were patched; component editions are unchanged. The manifest identifies the exact current bytes and the preserved baseline. This is a new correction build, not a silent replacement of the earlier frozen artifact. Live publication is not asserted.

Component/build identity. Aligners Sheet v5.9 retains its component edition but carries this correction build's identifier. Its numerical inputs, calculation formulas and worked verdict are unchanged; label, runtime-token and integrity-snapshot corrections are itemized in Verification/Exact_Workbook_Changes.json. Exact hashes, rather than filenames or edition labels alone, distinguish the new bytes from Release D and the previous frozen publication.

Verification boundary: Reports/Verification_and_Readiness.md and Verification/Final_G/ contain the current build's executed checks and limitations. Earlier verification and the rejected compatibility experiment belong to the preserved baseline and do not certify changed bytes. Cache-independent replay, stored-formula and native-engine results 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

Component	Current edition
Sentience Gradient Protocol	v8.8
ripple.md Standard	v5.8
RippleLogic Agent System	v13.0
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/Audit_Adjudication_and_Changes.md; exact edits: Verification/Exact_Document_Changes.json and Exact_Workbook_Changes.json. Navigation and metadata records are separate. Prior releases and 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)

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.

v2.8-rc2 Historical Release Integration

HISTORICAL (NON-CONTROLLING): Release: MathGov Core Release 2026.09 — RippleLogic Canon v12.8-rc2 / SGP v8.7-rc2 — Controlled Adversarial-Audit Correction Candidate

Release control	Historical value
Component	RLS Validation Protocol v2.8-rc2
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	Formal, reliability, validity, anti-gaming, and comparative

Release control	Historical value
	study protocol for RLS and its retained v12.7 integrity controls and v12.8-rc2 corrections.
Claim boundary	Controlled Tier 1–3 research specification candidate; not empirical validation, legal authority, physical-safety certification, deployment authorization, Tier 4, or moral truth.
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 (v2.8-rc2)

HISTORICAL (NON-CONTROLLING): This controlled candidate advances RippleLogic RLS Validation Protocol from v2.7 to v2.8-rc2 and binds it to MathGov Core Release 2026.09 — RippleLogic Canon v12.8-rc2 / SGP v8.7-rc2. The original release is preserved. Historical references remain lineage only; active references use the candidate component-version map. Publication, semantic-source parity and implementation conformance are not established by this reading-copy candidate.

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.

Historical Patch Note (non-controlling)

v2.4 synchronizes normative vector headings and package references. The study design, claim boundaries, and validation hypotheses remain unchanged.

Companion validation protocol for MathGov Core Release 2026.09 and the RippleLogic Canon.