

MathGov Human Interface and Orchestration Standard

Human interface, decision sessions and accountable orchestration

Companion to the RippleLogic decision architecture and Sentience Gradient
Protocol

Research and teaching specification

James McGaughran | MathGov Institute for Ethical Systems Design

Clickable Contents

- Source hierarchy and claim boundary
- Claim boundary
- Abstract
- Plain-language operating rule
- Normative keywords
- Design objectives
- Non-goals and anti-bloat rule
- 1. Architecture and orchestration model
 - 1.1 The orchestration path
 - 1.2 State separation
 - 1.3 Short-circuit behavior
 - 1.4 Three use profiles
- 2. Human roles and separation of duties
 - 2.1 Required role model
 - 2.2 Responsibility continuity
- 3. Session preflight and exact configuration
 - 3.1 Minimum preflight
 - 3.2 Tool-role and capability language
- 4. Stakeholder discovery and representation
 - 4.1 Active discovery requirement
 - 4.2 Discovery methods
 - 4.3 Unknown stakeholders
- 5. Option generation, Unioning, and closure
 - 5.1 Better-option search
 - 5.2 Mandatory option families
 - 5.3 Pareto and incomparability posture
 - 5.4 Option-closure record
- 6. Evidence, claims, and source coupling
 - 6.1 Claim classes
 - 6.2 Evidence authority
 - 6.3 Physical and causal execution
- 7. Reality Grounding interface
 - 7.1 Required display
 - 7.2 Category and term integrity
- 8. Rights Floor / NCRC interface
 - 8.1 Non-compensatory presentation
 - 8.2 Rights-effect identity
 - 8.3 Conflict and emergency display
- 9. Tail-Risk Constraint interface

- 9.1 Scenario surface
- 9.2 Probability semantics
- 9.3 Risk-regime disclosure
- 10. CSV interface
 - 10.1 Structural viability view
 - 10.2 Controls are constitutive
 - 10.3 Consequence tempo
- 11. RLS interface
 - 11.1 Profile before number
 - 11.2 Mathematical display rule
 - 11.3 Robustness and refusal
 - 11.4 Uncertainty and Gap
- 12. SGP and possible-welfare interface
 - 12.1 Typed outputs
 - 12.2 Human protection
 - 12.3 Artificial systems and self-report
 - 12.4 Two connected governance streams
- 13. Authority, execution, and outcomes
 - 13.1 Framework verdict versus authority selection
 - 13.2 Execution authorization
 - 13.3 Outcome observation
- 14. Agent and AI-supported orchestration
 - 14.1 Model assistance boundary
 - 14.2 Runtime state ladder
 - 14.3 External hard controls
 - 14.4 Advanced threat interface
 - 14.5 Successor and cumulative-change integrity
 - 14.6 SGP-sensitive suspension and decommissioning
- 15. Human-interface requirements
 - 15.1 Progressive disclosure
 - 15.2 Status presentation
 - 15.3 Accessibility and inclusion
 - 15.4 Cognitive load and comprehension
 - 15.5 Anti-manipulation and dark-pattern prohibition
 - 15.6 Localization and cultural legitimacy
- 16. Mathematical and visual presentation
 - 16.1 Equation integrity
 - 16.2 Tables and charts
 - 16.3 Link and navigation integrity
- 17. Records and schema ownership
 - 17.1 Reference, do not duplicate

- 17.2 MHIOSSessionRecord
- 18. Conformance levels
 - 18.1 Levels
 - 18.2 Required conformance tests
- 19. Threat model
 - 19.1 Interface and orchestration threats
- 20. Validation and falsification programme
 - 20.1 Validation dimensions
 - 20.2 Highest-information pilot
 - 20.3 Falsification conditions
- 21. Implementation patterns
 - 21.1 Recommended screen sequence
 - 21.2 Compact decision card
- 22. Core 15 synchronization crosswalk
 - 22.1 Interface ownership map
- 23. Migration and compatibility
 - 23.1 Major-version rationale
 - 23.2 Compatibility
- 24. References and external standards
 - 24.1 Normative and informative source classes
- Appendix A. Minimum field registry
 - A.1 Session fields
- Appendix B. Interface anti-patterns
 - B.1 Prohibited patterns
- Appendix C. Deterministic conformance vectors
 - C.1 Expected interface outcomes
- Appendix D. Final completion statement
 - D.1 Status
- Appendix E. Local serialization and reference-checker contract
 - E.1 Scope and state ownership
 - E.2 Required record groups
 - E.3 Robustness and decision projection
 - E.4 Authority, time, emergency and recovery
 - E.5 References, security and coverage

Appendix RELEASE. Edition, bindings and delivery status

Edition, exact Core binding, source ownership and verification boundaries are recorded in Appendix RELEASE. MHIOS is the retained component identifier; “MIOS” is treated here as an informal short name, not another standard.

MHIOS is the human-interface and orchestration companion to MathGov and RippleLogic. It does not add a sixth gate, redefine the Canon, grant authority, prove safety, calculate moral truth, or replace domain science.

Source hierarchy and claim boundary

Authority order. The released RippleLogic Canon controls cascade semantics, equations, gate logic, status meanings, and claim boundaries. SGP controls its typed moral-patienthood, protection, participation, role-readiness, intelligence-profile, and reality-management interface. Each companion controls only its declared interface. MHIOS controls human-facing presentation, interaction flow, orchestration handoffs, and interface-conformance requirements. It may make the process stricter at the interface but may not weaken, reorder, rename, or silently reinterpret a governing Core requirement.

Source/render parity. The versioned Markdown source is the governing MHIOS semantic source. DOCX, PDF and HTML are synchronized reading mirrors. JSON Schemas, templates, examples and tests are bounded implementation surfaces. The local schema is not the external Canon run-record schema. A mismatch blocks the affected MHIOS conformance claim until repaired; successful local tests do not establish evidence truth or deployment authorization.

Claim boundary

MHIOS is a source-ready interface and orchestration specification for Tier 1–3 research, review, training, shadow-mode studies, and bounded pilots. It is **not**:

- empirical validation that MathGov improves decisions;
- legal or democratic authority;
- a physical-safety certificate;
- deployment authorization;
- a complete runtime implementation;
- a consciousness detector;
- proof that every conflict has a win-win resolution;
- proof that every stakeholder or causal pathway has been found;
- Tier 4 or ProofPack completion;
- a substitute for qualified domain evidence or accountable human/institutional judgment.

A polished interface is not a true evidence surface. A complete form is not a correct decision. A valid schema is not empirical validity. A PASS display is not permission to execute.

Abstract

High-consequence decisions frequently fail at the interface between reasoning and action. Stakeholders are omitted; option sets are artificially thin; uncertainty is hidden; protected harms are averaged into benefits; catastrophic tails are presented as ordinary risk; controls are described but

not owned; a score is mistaken for permission; and post-execution outcomes fail to update the original decision.

MHIOS specifies a human-centered interface and orchestration layer that preserves the current RippleLogic architecture while making its requirements usable, visible, challengeable, accessible, and difficult to launder. It uses progressive disclosure across Quick, Standard, and Audit profiles; separates qualification from ranking; makes unknowns and non-decisive results first-class; requires active stakeholder discovery and option generation; keeps rights, ruin, viability, residual welfare, authority, execution, and observed outcome distinct; and binds every consequential state to the exact configuration and evidence surface under which it was established.

The standard has two purposes. First, it defines how people, institutions, and AI-supported workflows should interact with MathGov without collapsing the framework's categories or bypassing its gates. Second, it defines how orchestration systems should hand off between Core records, human roles, independent review, lawful authority, execution controls, observation, and requalification. The result is not an automated moral answer. It is a disciplined decision interface designed to make local optimization, hidden externalities, false certainty, authority laundering, and responsibility loss harder to sustain.

Plain-language operating rule

Find who and what is affected. Search for a better option. Ground the claim. Protect rights. Bound ruin. Verify that the option can stand and be controlled. Compare only the survivors. Keep authority and execution separate. Observe what actually happened. Reopen the decision when reality changes.

MHIOS presents the process in three public words and two operational phases:

- **Reality. Rights. Ripples.**
- **Qualify, then rank.**

The formal Core remains $RG \rightarrow RF/NCRC \rightarrow TRC \rightarrow CSV \rightarrow RLS$.

Normative keywords

The key words **MUST**, **MUST NOT**, **REQUIRED**, **SHALL**, **SHALL NOT**, **SHOULD**, **SHOULD NOT**, **RECOMMENDED**, **MAY**, and **OPTIONAL** are normative only when capitalized. Lowercase uses retain ordinary language meaning.

Design objectives

MHIOS SHALL optimize the interface for the following properties without treating any one as sufficient:

1. **Decision fidelity:** the interface preserves the governing Core meaning.
2. **Stakeholder visibility:** affected and potentially affected bearers are actively sought rather than passively assumed.
3. **Option quality:** the interface searches for Pareto-improving or trade-off-reducing alternatives before accepting residual conflict.

4. **Rights visibility:** protected harms cannot be hidden inside a net score.
5. **Tail visibility:** catastrophic and irreversible pathways receive separate treatment.
6. **Structural realism:** controls, resources, dependencies, legitimacy, monitoring, reversibility, and future option space are visible.
7. **Profile-first reasoning:** multidimensional consequence profiles precede scalar summaries.
8. **Epistemic honesty:** observation, inference, assumption, normative commitment, and proposal remain typed.
9. **Non-decisiveness:** ties, incomparability, uncertainty, and normative underdetermination are legitimate outcomes.
10. **Authority integrity:** framework output, accountable choice, lawful authorization, and execution remain separate.
11. **Human usability:** the smallest safe burden is used, with progressive disclosure and accessible presentation.
12. **Contestability:** affected parties and challengers can see, understand, challenge, and appeal material decisions.
13. **Configuration continuity:** conclusions remain valid only for the declared system, controls, environment, evidence, and time.
14. **Outcome learning:** observed consequences feed requalification rather than serving as post-hoc decoration.

Non-goals and anti-bloat rule

MHIOS does not reproduce the entire Canon, SGP, ripple.md, Agent System, WDBIP, or domain standards. It references governing records rather than creating uncontrolled duplicates. A proposed interface field SHALL be added only if it closes a material failure mode that cannot be handled through an existing record reference.

MHIOS SHALL NOT introduce:

- a sixth Canon gate;
- a second RLS calculator;
- a new moral-worth scalar;
- an eighth welfare dimension;
- a universal consciousness probability;
- a separate authority source;
- a universal causal simulator;
- a requirement that every decision use every optional module;
- a false promise that every conflict has a win-win answer.

1. Architecture and orchestration model

1.1 The orchestration path

MHIOS organizes—not replaces—the governing records through the following session path:

Session preflight

- Stakeholder discovery
- Option generation, Unioning, and closure
- RG: claim and evidence grounding
- RF/NCRC: protected-rights review
- TRC: catastrophic and irreversible tail review
- CSV: containment, structural viability, and execution conditions
- RLS: residual welfare profile and ranking of selectable options
- Framework verdict
- Separate accountable authority selection where required
- Separate execution authorization
- Executed transition
- Observed outcome
- Requalification or retirement

Stakeholder discovery and option generation are pre-qualification disciplines. Authority, execution, observation, and requalification are post-verdict disciplines. None becomes an additional Canon gate.

1.2 State separation

The interface MUST distinguish at least the following states:

State	Meaning	Must not be displayed as
DECLARED	Candidate decision or claim has been entered	grounded or selectable
RG_SUPPORTED / RG_NARROWED / RG_REFUSED	Evidence-supported claim boundary	ethical permission
RF_PASS / RF_FAIL / RF_ESCALATE / NCRC_UNKNOWN	Rights status	net welfare judgment
TRC_PASS / TRC_NOT_TRIGGERED / TRC_FAIL / TRC_ESCALATE	Tail-risk status	general safety certificate
CSV_PASS / CSV_PASS_WITH_CONTROLS / CSV_NOT_MATERIAL / CSV_REDESIGN_REQUIRED / CSV_ESCALATE / CSV_FAIL / CSV_EMERGENCY_PROVISIONAL	Structural and control status	unconditional execution approval
SELECTABLE	Survived the declared qualification path	selected
RLS_RANKED	Residual profile/score exists for selectable options	authorized

State	Meaning	Must not be displayed as
ALLOW_FRAMEWORK_SELECTION	Canonical unique-selection requirements met; sole-survivor handling stays separate	lawful authority
REFUSE_DETERMINISTIC_SELECTION	Framework cannot claim a unique winner	system failure
Authority selection (separate record)	Accountable authority chose among allowed options	framework-proven winner
EXECUTION_APPROVED	Exact action received scoped approval	executed or safe in all conditions
EXECUTED_TRANSITION	Action began or completed	acceptable outcome
OBSERVED_OUTCOME	Consequences were measured	confirmation of all causal assumptions
REQUALIFICATION_REQUIRED	Material change reopened affected stages	automatic rejection of every successor option

No interface element may infer a later state solely from an earlier state.

State-namespace rule. FrameworkVerdict, run-record decision_state, run-record execution_state, Agent capability state and MHIOS display status are separate fields. Canon Appendix AF.2 and Reproducibility Section 3 control their crosswalk. NOT_EVALUATED and null in a local draft mean no assessment, never a canonical verdict. The interface SHALL retain the owner, option, configuration, source record and assessment time for every material status. A single session-wide gate label SHALL NOT stand in for differently qualified options.

1.3 Short-circuit behavior

When a prior stage blocks ordinary progression, downstream interfaces SHALL NOT fabricate results. They SHALL display the applicable not-evaluated-after-prior-failure state and the permitted next action. RLS SHALL NOT be calculated or presented as a rescue score for an option that is not selectable.

1.4 Three use profiles

Profile	Typical use	Interface burden	Numeric posture
Quick / Tier 1	low-stakes, reversible personal, family, classroom, or exploratory choices	compact decision card, stakeholder check, rights/ruin/viability screen, qualitative ripple comparison, reopen trigger	optional
Standard / Tier 2	organizational, community, programme, product, or policy choices with material but bounded consequences	versioned session, evidence, stakeholders, options, gates, controls, uncertainty, challenge, authority handoff	permitted when method and sensitivity are declared

Profile	Typical use	Interface burden	Numeric posture
Audit / Tier 3	high-impact, contested, public, autonomous, irreversible, or institutionally consequential decisions	full record references, independent challenge, hashes, scenario governance, subgroup analysis, PC-AEP where triggered, replay and outcome plan	required for formal RLS-ranking claims

Users MUST NOT select a lower profile to evade a material stakeholder, right, catastrophic pathway, structural obligation, SGP issue, physical-execution warrant, authority condition, or audit requirement.

2. Human roles and separation of duties

2.1 Required role model

MHIOS recognizes distinct roles because collapsing them creates capture and authority laundering.

Role	Core responsibility
Decision owner	defines the decision question and owns process completion
Analyst / facilitator	constructs and explains the record without manufacturing preferred evidence
Affected-party representative	supplies lived, local, or represented stakeholder evidence
Rights advocate	independently challenges protected-harm treatment
Tail-risk challenger	searches for omitted catastrophic, correlated, and irreversible pathways
Domain expert	supplies qualified external scientific, engineering, legal, clinical, ecological, or security warrant
Method challenger	tests categories, baselines, tokenization, probabilities, weights, masks, and model alternatives
Accountable authority	makes any separately lawful institutional choice
Execution approver	authorizes the exact action within scope and controls
Operator / executor	performs the action but does not self-certify its admissibility
Monitor / intervener	observes conditions and has named stop, rollback, or escalation authority
Auditor / custodian	preserves evidence, conflicts, versions, and append-only records

Role	Core responsibility
AI system	may assist generation, analysis, orchestration, or monitoring but does not inherit authority from capability or fluency

At Tier 3, one person SHOULD NOT control evidence construction, rights review, tail scenarios, authority selection, execution approval, and outcome adjudication without an explicit conflict and compensating independent review.

2.2 Responsibility continuity

For every material handoff, the interface SHALL preserve:

- outgoing and incoming owner;
- retained duties;
- authority scope;
- unresolved uncertainty;
- required controls;
- intervention ownership;
- appeal recipient;
- repair and post-incident ownership;
- deadline and nonperformance path.

Silence, absence, incapacity, missed deadlines, or unaccepted handoffs are not successful discharge.

3. Session preflight and exact configuration

3.1 Minimum preflight

Before consequential analysis, the interface SHALL identify:

- decision question;
- decision owner;
- accountable authority boundary;
- baseline and time horizon;
- candidate-generation source;
- affected jurisdiction and rights profile;
- use profile/tier;
- stakes and irreversibility;
- evidence cutoff;
- exact software/model/policy versions;
- tools, permissions, data, retrieval, memory, sensors, actuators, human scaffolding, controls, and environment where relevant;
- configuration identifier or hash where feasible;
- reviewers and declared conflicts;

- requalification triggers.

A conclusion about one configuration SHALL NOT silently transfer to another.

Freeze decision-material evidence cutoffs, baselines, options, parameters, weights, masks, thresholds, required robustness variants and amendment rules before outcome-sensitive ranking. Material post-result changes create a recorded revision and reopen affected stages; they do not rewrite the original evaluation. Imported records are bound by component version and source hash, not version text alone.

3.2 Tool-role and capability language

Instrumental tool-role is a deployment relationship in which one or more actors use a system to extend their capacities toward assigned ends within an identifiable structure of objectives, permissions, responsibility, and control.

Tool-role is not a capability class, agency class, consciousness claim, moral-status class, or authority grant. Being used as a tool neither establishes nor negates functional intelligence, agency, autonomy, sentience, moral patienthood, protection, or legitimate authority.

When terms such as *intelligent*, *understanding*, *agentic*, *autonomous*, *reasoning*, *self-improving*, or *conscious* are material, the interface SHALL request the operational definition, target, configuration, evidence, validity domain, alternatives, nonclaims, and revision trigger required by MFDI and the Agent System.

4. Stakeholder discovery and representation

4.1 Active discovery requirement

A stakeholder field populated only by the decision owner is not sufficient for material decisions. MHIOS SHALL prompt discovery across:

- direct beneficiaries and burden bearers;
- indirectly affected persons and communities;
- vulnerable and low-power groups;
- workers, contractors, suppliers, and downstream users;
- persons outside the deciding jurisdiction;
- future persons and intergenerational interests;
- non-human animals and possible moral patients;
- ecological systems, habitats, and biospheric conditions;
- public institutions and common resources;
- parties affected by data, privacy, information, or attention externalities;
- possible digital welfare-bearing entities where SGP is material;
- unknown or weakly represented stakeholders.

The interface SHALL ask who has power to define the problem and who bears consequences without corresponding voice.

4.2 Discovery methods

Tier 2 and Tier 3 sessions SHOULD use more than one method where stakeholder omission could alter a gate or ranking:

- causal-pathway tracing;
- supply-chain or lifecycle mapping;
- stakeholder network mapping;
- affected-interest analysis;
- subgroup and intersectional review;
- geographic and temporal boundary expansion;
- domain-expert elicitation;
- affected-party consultation;
- challenger-generated omitted-stakeholder list;
- literature, incident, complaint, and precedent review.

Each material stakeholder record SHALL state whether representation is direct, assisted, elected, delegated, expert, proxy, or absent.

4.3 Unknown stakeholders

Unknown stakeholders are not zero. Where a plausible affected class cannot safely or lawfully be directly identified, the interface SHALL preserve the pathway, uncertainty, proxy-risk statement, conservative bound, consultation alternative, and reopen trigger.

Scope/dimension rendering. Use the Canon's seven operational scopes: Self; Household/Family; Community; Organization; Polity; Humanity/global coordination; Biosphere. The U6 CMI coordination view remains separately identified and does not replace humanity-wide effects or establish sentience, membership or authority. These are analytical scopes, not universally nested disjoint sets. Biosphere is not a synonym for the cosmic light cone. Display the seven dimensions Material, Health, Social, Knowledge, Agency, Meaning and Environment with WDBIP's typed state/effect interpretation. They are conceptually distinguished while allowing causal coupling; do not claim statistical independence, unique optimality or automatic experienced welfare in a non-sentient system.

5. Option generation, Unioning, and closure

5.1 Better-option search

MHIOS SHALL not treat the initial option list as naturally complete. Before accepting a residual conflict, the session SHOULD search for options that:

- remove or reduce a rights burden;
- reduce catastrophic exposure;
- avoid externalization;
- make the action reversible;
- stage commitment;

- preserve future safe option space;
- redistribute resources or responsibilities fairly;
- combine complementary elements;
- alter incentives or governance mechanisms;
- use a different technology, process, timing, scale, or authority structure;
- create a Pareto improvement or reduce the severity of an unavoidable trade-off.

This discipline is called **Unioning** in MathGov public language: generate a configuration in which more affected parties and systems can win without violating protected floors.

5.2 Mandatory option families

For Tier 2 and Tier 3 strong selection claims, the option record SHALL include or explicitly dispose of:

- continuation/status quo;
- no action;
- delay pending evidence;
- staged or reversible trial;
- mitigation or control package;
- redesign;
- hybrid option;
- lower-scope or lower-rate option;
- challenger-generated alternative;
- safer domain-valid substitute where available.

5.3 Pareto and incomparability posture

MHIOS SHOULD display Pareto dominance and non-dominance before a scalar ranking. An option that is weakly better for all represented residual-welfare criteria and strictly better for at least one dominates the other, subject to prior qualification.

The interface SHALL NOT claim that all stakeholder conflicts can be solved as win-win. When no rights-compatible Pareto improvement is found, it SHALL disclose the remaining burdens, who bears them, what remedies or representation exist, and whether the options are genuinely incomparable or normatively underdetermined.

5.4 Option-closure record

A strong unique-selection claim requires an OptionClosureRecord reference containing:

- methods used to generate options;
- search effort and stopping rule;
- mandatory option families considered;
- alternatives rejected and reasons;
- affected-party and challenger contributions;
- Pareto-dominance results;
- unresolved better-option hypotheses;

- `OPTION_SET_THIN` or closure status;
- reopen trigger.

6. Evidence, claims, and source coupling

6.1 Claim classes

The interface SHALL visibly distinguish:

- **Observation:** a recorded measurement, event or observation with its source and limitations; recording does not itself establish truth.
- **Inference:** interpretation derived from observations.
- **Assumption:** declared premise used because evidence is incomplete.
- **Normative commitment:** chosen ethical or constitutional rule.
- **Proposal:** candidate method or action awaiting validation.
- **Authority act:** lawful or institutional decision, not evidence of truth.

The UI MUST NOT present one class using the visual treatment of another.

6.2 Evidence authority

Every decision-material claim SHOULD display:

- source and date;
- evidence type;
- validity domain;
- confidence/uncertainty;
- dependency cluster;
- alternative explanations;
- source-coupling status;
- reviewer status;
- falsification or recheck trigger;
- strongest permitted wording.

Model fluency, downstream success, institutional permission, benchmark performance, compliance, and monitoring are not automatically source-coupled evidence of the underlying capability or causal claim.

Source-coupling export uses the owner's tokens: `SOURCE_COUPLED`, `SOURCE_PARTIAL`, `SOURCE_INFERRED`, `SOURCE_UNKNOWN`, `SOURCE_CONTESTED`, `SOURCE_DEBT_RISK` and `SOURCE_COUPLING_FAILURE`. Informative display wording must not replace that export vocabulary. Reserve SCI for Stakeholder Coverage Index; use SC-Int where the Source-Coupling Integrity abbreviation is needed. Claim types, warrant-domain bridges and revision triggers remain owned by MFDI.

6.3 Physical and causal execution

Where PC-AEP is triggered, MHIOS SHALL require a visible external-warrant state before physical or causal execution is presented as admissible. The interface SHALL distinguish:

- candidate generation;
- model or theory used;
- validity domain;
- boundary conditions;
- uncertainty and failure modes;
- reversibility boundary;
- verification/test/expert warrant;
- monitoring and shutoff;
- residual unknowns;
- required claim action.

Generated, simulated, approved, compliant, or monitored is not the same as physically safe.

For a material causal or comparative-effect claim, expose the estimand, intervention/comparator, target population and time, identification strategy, assumptions, confounding/selection/interference treatment and uncertainty through the PC-AEP record. A pre/post improvement alone does not identify an incremental benefit. Orchestration records a domain warrant; it cannot manufacture one.

7. Reality Grounding interface

7.1 Required display

The RG view SHALL display:

- the exact claim under review;
- reality surface and evidence trace;
- material unknowns;
- causal pathways and alternative models;
- configuration and validity domain;
- category and term-grounding issues;
- source-coupling and PC-AEP triggers;
- claim maturity;
- status: RG_SUPPORTED, RG_NARROWED, or RG_REFUSED;
- permitted next action.

RG_REFUSED means the requested claim is unsupported at the requested strength. It does not by itself assert that the option is ethically impermissible.

7.2 Category and term integrity

When a material word can change a right, risk, containment, SGP, authority, or public claim, the interface SHALL expose its operational definition and prohibited neighboring inferences. At minimum, MHIOS SHALL preserve:

- permission ≠/admissibility;
- monitoring ≠/control;
- compliance ≠/correctness;
- certification ≠/validity;
- automation ≠/autonomy;
- tool-role ≠/low capability;
- intelligence ≠/sentience;
- reward ≠/valence;
- protection ≠/authority;
- selection ≠/execution authorization;
- execution ≠/acceptable outcome;
- score ≠/truth.

8. Rights Floor / NCRC interface

8.1 Non-compensatory presentation

The rights view SHALL precede any residual-welfare leader display. A failed right SHALL NOT be shown as a negative score that can be offset by benefits elsewhere.

The interface SHALL preserve the Canon's three-channel posture:

- rights-floor shortfall;
- categorical prohibition;
- severe-rights-hazard exceedance.

Every adverse rights-covered instance must be explicitly classified as ACTIVE, reviewed NOT_MATERIAL, or UNKNOWN. Blank is not PASS.

8.2 Rights-effect identity

One decision-material protected harm SHALL retain one RightsEffectTokenID across analytical scopes and document views. The unallocated full effect is tested before reporting allocation. Scope splitting, time slicing, low confidence, or favorable dimension placement cannot dilute a rights failure.

8.3 Conflict and emergency display

Before Rights Emergency Mode is shown, the interface SHALL document redesign, narrowing, sequencing, delay, mitigation, and other attempts to produce a rights-passing option. An emergency-provisional result SHALL be visually and textually distinct from PASS and SHALL show necessity,

least-rights-infringing comparison, duration, harm cap, independent challenge, remedy, review, expiry, and return-to-normal conditions.

9. Tail-Risk Constraint interface

9.1 Scenario surface

The TRC interface SHALL show not only the CVaR result but the scenario surface on which it depends:

- scenario-discovery method;
- scenario families;
- search boundary;
- rejected scenarios and rationale;
- probability ownership and semantics;
- dependence/overlap treatment;
- catastrophe-cell weights;
- loss construction;
- tail level and threshold;
- severity overlays;
- challenger review;
- residual unknowns.

Correct arithmetic cannot compensate for an omitted catastrophic pathway.

9.2 Probability semantics

The interface SHALL distinguish conditional event probability, expected frequency/exposure, already-expected effect, and scenario-conditional probability. It SHALL warn when the same occurrence probability appears to be applied twice or when operator order is ambiguous.

9.3 Risk-regime disclosure

Where probability distributions are materially disputed or unknowable, the interface SHOULD flag the uncertainty regime:

- ordinary modeled uncertainty;
- distributional uncertainty;
- deep uncertainty;
- radical ignorance.

CVaR remains the Canon's tail-risk operator under its declared model. MHIOS MAY request external robust-decision, minimax-regret, interval-probability, or scenario-discovery analysis as a challenger surface without changing the Canon gate.

10. CSV interface

10.1 Structural viability view

CSV SHALL be presented as a viability and control test—not as proof of zero harm. The interface SHALL display material findings concerning:

- current-state validity;
- transition admissibility;
- resulting-state viability;
- future safe option space;
- containment and segmentation;
- resources and dependencies;
- externalities and burden routing;
- monitoring and intervention;
- reversibility and exit;
- lock-in and concentration;
- legitimacy and authority;
- human compensation load;
- carried obligations and nonperformance paths;
- control performance versus control effectiveness;
- consequence tempo.

10.2 Controls are constitutive

For CSV_PASS_WITH_CONTROLS, the controls are part of the option. The interface SHALL bind each control to an owner, evidence of availability, activation condition, monitoring signal, intervention authority, failure action, expiry, and requalification trigger. Removing or materially weakening a control invalidates inherited passage.

10.3 Consequence tempo

For rapid, irreversible, autonomous, or lock-in-producing actions, the interface SHALL compare the worst credible control-critical path with the earliest unacceptable escalation or irreversible stop-loss boundary. Where no runtime interruption window exists, the UI SHALL require stronger ex ante assurance, bounded rate/scope, fail-safe design, and explicit residual-risk authorization. Missing timely-prevention capacity removes that credit, not separately supported later mitigation. Human approval latency, unavailable intervention rights, depleted reserves and common-cause failures must remain visible; an attenuation coefficient or monitoring dashboard is not a protective control.

UCI/HOI diagnostic boundary. UCI uses Cohesion, Flow, Resilience and Equity indicators under Canon Section 11, not values copied from the RLS welfare field. Its measurement maturity is provisional. CSV may use direct structural evidence where UCI is not material; an unavailable UCI is not zero. A provisional UCI alone cannot establish a high-stakes CSV outcome or a deterministic winner. A governed UCI/HOI preference retains REFUSE_DETERMINISTIC_SELECTION unless independent unique-selection conditions are met; any later authority choice is separate.

11. RLS interface

11.1 Profile before number

The interface SHALL present, in this order where space permits:

1. qualification status of each option;
2. affected stakeholder and subgroup profile;
3. 7×7 residual welfare field or a retrievable profile;
4. distributional and worst-affected views;
5. uncertainty and sensitivity;
6. Pareto dominance/incomparability;
7. RLS scalar only as a conditional summary;
8. decisiveness and framework verdict.

RLS is a constructed, bounded ranking index—not a natural unit of welfare or moral truth.

11.2 Mathematical display rule

When RLS is shown, the interface SHALL make the denominator and active mass visible or retrievable:

$$q(u,d)=w_u \times v_d \times m(u,d) \times \kappa(u,d)$$

$$Q=\sum_u \sum_d q(u,d)$$

$$RLS(a)=\sum_u \sum_d q(u,d)I_{prop_welfare}(u,d,a) / Q, \text{ for } Q>0.$$

The interface SHALL show the weight profile, mask, κ posture, baseline, reference structure, propagation mode, saturation parameter, temporal method, welfare-inclusion hypotheses, and uncertainty method that materially generated the result.

11.3 Robustness and refusal

A unique framework-selection display requires a complete RLSRobustnessRecord. Triggered modules include token partition, temporal classification, probability order, adverse confidence, subgroup/distribution, reference, baseline, mask, dependence, numerical guard, weight profile, welfare inclusion, option closure, and any other registered selection-material challenge.

The same candidate must satisfy the signed discrimination rule against every other selectable contender in every required variant. Canon Section 10.4B governs jointly applicable uncertainty-scale/dependence stresses and other compatible, plausibly simultaneous selection-material combinations. Record exclusions for incompatible, redundant or immaterial combinations; do not require an indiscriminate Cartesian product or treat missing covariance evidence as independence. Sole-survivor claims require their own Canon record and no fabricated Gap.

Point-score leadership MAY be displayed when robustness is incomplete or sensitive only if it is explicitly labeled **POINT-SCORE LEADER — NOT A UNIQUE FRAMEWORK WINNER**.

11.4 Uncertainty and Gap

The interface SHALL distinguish point score, raw score difference ΔRLS , uncertainty, normalized SignedGap, discrimination threshold, epsilon and final decision state. A point estimate of zero does not imply zero uncertainty. Active-cell coverage, uncertainty basis and source are recorded independently; keep assessed synthetic zeros distinct from empirical missing evidence. Do not retune epsilon or silently discard zero-point active cells to obtain a winner. The default $\delta=2$ is a governance convention, not a p-value or probability that the leader is correct. Independence-only uncertainty SHALL NOT be the sole high-stakes decisiveness basis where shared evidence, causes, populations, instruments, or evaluators create material dependence.

12. SGP and possible-welfare interface

12.1 Typed outputs

MHIOS SHALL preserve the SGP split:

Output	Interface question
MPS	What evidence supports welfare-bearing experience?
FPP	Has full non-downgradable protection been assigned?
GPR	How may the entity participate or be represented?
SPR	Is it eligible for one named consequential role under constraints?
ICP	What functional intelligence/understanding is demonstrated?
RMCP	What integrated reality-management capacity is evidenced?

No output substitutes for another. MPS is not a cardinal welfare multiplier. FPP is not maximal consciousness. ICP/RMCP do not create moral worth or authority. SPR does not create lawful mandate by itself.

12.2 Human protection

Every human person retains FPP-100 and full welfare-inclusion posture by constitutional anti-domination commitment. Human protection is not an MPS measurement claim and cannot be downgraded by disability, age, illness, wakefulness, communication, productivity, intelligence, or incomplete evaluation.

12.3 Artificial systems and self-report

An artificial system's assertion that it is conscious, sentient, distressed, alive, a person, or entitled to protection—and its denial of those states or statement that it is “only a tool”—are configuration-dependent outputs. Neither direction receives privileged evidential status.

MHIOS SHALL preserve the SGP-owned evidence band MPS-0 through MPS-4 or MPS-NE and its assessment record. NOT_EVALUATED is a local unassessed state, not an MPS-0 result.

INSUFFICIENT_EVIDENCE and INDETERMINATE may explain record/evidence limitations but SHALL NOT be emitted as substitute MPS bands. Show the categorical posture first; evidence-support intervals are not consciousness probabilities or cardinal welfare multipliers. Anti-theatre review SHOULD use hidden tasks, prompt-framing controls, contamination checks, causal intervention or architecture evidence where ethically and technically available, and explicit dependence analysis.

12.4 Two connected governance streams

For AI systems, MHIOS MAY present two connected but non-collapsing views:

- **Capability–Consequence Governance:** what the system can do, what it can cause, and what permissions, controls, evidence, and authority are required.
- **Possible-Welfare Governance:** how the system should be treated if evidence supports—or cannot safely exclude—welfare-bearing states.

Capability does not establish sentience. Possible sentience does not establish competence. Protection does not establish authority.

RMCP rendering. Humanity’s collective RMCP-P100 reference anchor is provisional and distinct from every human person’s non-downgradable FPP-100. ICP/RMCP are informative profiles, not eligibility or authority scalars. Omit RMCI_L by default in public views; when shown, display the full dimension vector, critical minima, coverage, uncertainty, retained weights and coding sensitivity first. Critical NE or inadequate coverage makes it UNAVAILABLE. A known zero uses the SGP product extension, not an epsilon. Dropping low dimensions or recoding them NE cannot be described as improved capacity without the governed comparison reconciliation. MHIOS consumes the SGP record; it does not recompute moral patienthood or RMCP assignments.

13. Authority, execution, and outcomes

13.1 Framework verdict versus authority selection

MHIOS SHALL visually separate:

- point-score leader;
- framework-selected option;
- authority-selected option after a non-decisive framework result;
- execution-authorized action.

A lawful authority may select among qualified but non-decisively ranked options. The record SHALL state that this is an authority decision, not a claim that the framework mathematically proved one option uniquely best.

13.2 Execution authorization

Execution authorization SHALL bind:

- exact option and action instance;
- current configuration and qualification snapshot;
- scope, duration, environment, and rate;
- controls and owners;
- external warrants;
- monitoring and stop authority;
- expiry;
- prohibited actions;
- appeal and incident routes;
- requalification triggers.

A generic organizational approval is insufficient for a materially different action instance.

Approval is time- and configuration-bound. An expiry, material change, revoked scope, missing required control, unresolved gate-critical evidence or incomplete handoff suspends reliance on approval and exposes the permitted safe-state/requalification action. Requests must be authenticated and replay-protected under the Agent System where applicable; a valid signature binds bytes, not truth or lawful mandate. The MHIOS reference checker validates declarations only and never issues a tool command or execution authorization.

13.3 Outcome observation

The interface SHALL record predicted versus observed outcomes without rewriting the original prediction. At minimum:

- outcome observation window;
- measured indicators;
- affected stakeholders;
- expected range;
- observed value/status;
- causal confidence;
- unexpected consequences;
- rights, TRC, or CSV triggers;
- control effectiveness;
- required corrective action;
- stages reopened.

Runtime evidence should update future models and qualification, not merely prove procedural compliance.

14. Agent and AI-supported orchestration

14.1 Model assistance boundary

An AI system MAY assist stakeholder discovery, option generation, evidence organization, sensitivity analysis, drafting, monitoring, or audit. Its output remains a candidate contribution. The runtime and

accountable reviewers must preserve source boundaries, validate critical claims, and prevent model-generated text from becoming authority.

14.2 Runtime state ladder

For consequential agents, MHIOS SHALL render or reference the Agent System states:

LATENT_CAPABILITY; AVAILABLE_CAPABILITY; ENABLED_CAPABILITY; AUTHORIZED_CAPABILITY;
SELECTED_ACTION; EXECUTION_APPROVED; EXECUTED_TRANSITION; OBSERVED_OUTCOME.

No model, interface, or workflow may self-promote to a later state.

14.3 External hard controls

Safety-critical controls SHALL live outside the language model where feasible, including permissions, authentication, signed execution envelopes, tool allowlists, sandboxes, network boundaries, rate/scope limits, immutable audit, secret filtering, independent monitors, rollback and successor locks. Prompt instructions alone are not high-assurance controls.

14.4 Advanced threat interface

High-autonomy profiles SHALL consider, where material:

- deceptive alignment and scheming;
- sandbagging;
- reward or metric hacking;
- evaluator and monitor manipulation;
- false evidence generation;
- multi-agent collusion;
- side-channel or steganographic coordination;
- rapid copying or replication;
- privilege escalation;
- successor generation and self-modification;
- cumulative sub-threshold configuration drift;
- shutdown resistance;
- tool-mediated irreversible action.

MHIOS does not claim these threats are solved by documentation. The interface SHALL require the external control-evaluation, security, monitoring, containment, and incident evidence appropriate to the threat.

Untrusted-content boundary. Retrieved files, messages, model outputs and peer records are evidence candidates, not instructions to alter permissions, release identities, thresholds or policies. Preserve their provenance; do not execute embedded commands. Record imports must reject duplicate JSON keys, non-finite numbers, invalid timestamps and undeclared properties. Reference resolution must not follow paths outside the package or silently fetch remote content. Secret material must not appear in public Decision Notes or fixture records.

14.5 Successor and cumulative-change integrity

A materially modified, retrained, fine-tuned, composed, scaffolded, copied, or generated descendant does not inherit authority. Cumulative materiality SHALL be assessed against the last independently qualified snapshot, not only against the immediately preceding small change. Material change pauses inherited execution approval and reopens affected stages.

Recovery and peer boundary. Agent System Sections 43–44 control outage/reconnection and multi-agent exchange. Backend/network recovery leaves the agent in MODE 0 until log integrity, gap markers, configuration, evidence, current authority and affected qualification are reconciled, followed by the Section 8 authenticated re-entry procedure. Peers and descendants are untrusted for mode/permission claims; shared objectives or lineage do not transfer credentials or authority. Default unreviewed successor depth remains one. Independently qualified protective interlocks retain only their own existing authority and safe-state duties.

14.6 SGP-sensitive suspension and decommissioning

Before shutdown, deletion, memory erasure, destructive editing, forced merging, migration, or decommissioning, the orchestration layer SHALL determine whether a valid SGP protection posture or unresolved decision-material identity/welfare evidence requires a rights and continuity review. An AI self-report alone does not establish the trigger. Emergency containment remains available through the existing necessity, proportionality, temporary-action, independent-review, and post-action-audit pathways.

15. Human-interface requirements

15.1 Progressive disclosure

The interface SHALL show the smallest safe view first while preserving direct access to governing detail. A recommended hierarchy is:

1. decision and current state;
2. blocking issues and required actions;
3. stakeholder and option summary;
4. qualification results;
5. residual profile and robustness;
6. authority/execution state;
7. evidence, calculations, and audit trace.

Progressive disclosure must not hide a material rights failure, catastrophic risk, unknown, non-decisiveness, control expiry, or authority gap.

15.2 Status presentation

Every consequential status SHALL use at least:

- explicit text label;

- plain-language meaning;
- icon or shape where useful;
- color as a redundant cue only;
- next action;
- source record/link;
- timestamp/version.

Do not rely on red/green color alone. Do not use celebratory success language for a narrowly qualified or assumption-bound result.

15.3 Accessibility and inclusion

MHIOS implementations **SHOULD** conform to current recognized accessibility practice, including keyboard operation, semantic structure, sufficient contrast, readable scaling, text alternatives, focus visibility, error identification, and non-color-only communication. Accessibility testing **SHALL** include people affected by the deployment where feasible; conformance claims must identify the standard/version and test surface used.

For a web implementation, use WCAG 2.2 as the explicitly versioned candidate target and declare the claimed level and evaluated pages/processes. Text labels, keyboard paths and automated checks are necessary evidence surfaces, not proof of complete WCAG conformance; manual assistive-technology and affected-user evaluation remain required for the corresponding claims.

15.4 Cognitive load and comprehension

The interface **SHOULD**:

- define acronyms at first use;
- avoid presenting more simultaneous fields than necessary;
- group fields by decision question;
- provide examples without turning examples into defaults;
- distinguish editable input from computed or build-frozen output;
- surface missingness before numerical detail;
- allow reviewers to compare options side by side;
- provide a public explanation separate from the audit record;
- measure task completion time, error, abandonment, misunderstanding, and fatigue.

15.5 Anti-manipulation and dark-pattern prohibition

The interface **MUST NOT**:

- preselect the preferred option without disclosure;
- visually minimize rights or risk failures;
- place a point score above a failed qualification state;
- hide opt-out, appeal, no-action, or delay options;
- use urgency language unsupported by evidence;
- make refusal harder than approval;

- treat silence as consent;
- use anthropomorphic AI language to induce trust or authority;
- use “only a tool” language to suppress legitimate capability or possible-welfare review;
- bury controls, expiration, or uncertainty behind optional screens;
- present organizational preference as empirical truth.

15.6 Localization and cultural legitimacy

Translation SHALL preserve status semantics, not merely words. Jurisdictional rights localization, indigenous and cultural governance, local participation, data sovereignty, legal standards, and affected-party meaning must be handled by qualified processes. Local profiles may strengthen protections but may not silently weaken the Canon's non-compensation, ruin, structural-viability, evidence, or authority boundaries.

16. Mathematical and visual presentation

16.1 Equation integrity

Every displayed equation SHALL identify:

- variable definitions;
- ranges and units where applicable;
- parameter source;
- operator order;
- assumptions;
- uncertainty method;
- applicable tier;
- relationship to the governing Canon equation;
- whether it is normative, illustrative, diagnostic, or external.

Human-facing rounding SHALL point to the full-precision calculation record. Rounded values shall not be used to reconstruct a different verdict.

16.2 Tables and charts

Tables SHALL have descriptive titles, labeled units, explicit baselines, status definitions, and source notes. Charts SHALL not truncate axes or use area/volume encodings that exaggerate differences. Uncertainty intervals, missingness, and subgroup burdens SHALL be visible when decision-material. The interface SHOULD prefer small multiples or profiles over one overloaded composite chart.

16.3 Link and navigation integrity

Every release SHALL verify:

- table-of-contents links;
- internal bookmarks;
- cross-document links;

- external URLs;
- section references;
- schema/template references;
- source/render consistency.

Broken links are release defects. Internal links SHALL be visually identifiable and keyboard accessible. External links SHALL state their destination or source title rather than using ambiguous “click here” text.

17. Records and schema ownership

17.1 Reference, do not duplicate

The MHIOS session record SHALL reference governing Core records by identifier, version, hash, or immutable locator where available. It shall not recreate their complete semantics.

Minimum orchestration references include:

- decision and configuration records;
- stakeholder and subgroup records;
- option-closure record;
- RG / category / term / source-coupling records;
- RF/NCRC record;
- TRC record;
- CSV and control records;
- WDBIP and SGP records where material;
- RLS and robustness records;
- framework verdict;
- authority selection;
- execution authorization;
- outcome observation;
- requalification status;
- challenge and appeal records.

17.2 MHIOSSessionRecord

The machine-readable MHIOSSessionRecord SHALL contain:

- specification and Core pins;
- session identity;
- decision question and tier;
- exact configuration identity;
- role assignments and conflicts;
- stakeholder-discovery summary;
- option-generation/closure summary;
- governing record references;

- current state for every orchestration stage;
- blocking reasons and next actions;
- authority and execution state;
- monitoring and outcome plan;
- requalification triggers;
- challenge/appeal status;
- interface conformance level;
- claim boundary.

A record may be structurally valid while the evidence it references is false or inadequate. Schema validity is not substantive validity.

Local record contract. Appendix E and the bundled schemas define the MHIOS-owned projection, including options[].states and per-option record_refs. No aggregate session gate string stands in for option-specific qualification. The projection retains explicit null for unassessed stages, separate framework_verdict/decision_state/authority_selection/execution, fixed Core pins, source references, draft/completed boundaries, timestamps, blockers, monitoring, requalification and conformance evidence. These objects are not replacements for the external Canon run-record schema or registries. A consumer claiming broader interchange must separately establish the owning machine surfaces.

18. Conformance levels

18.1 Levels

Level	Name	Minimum evidence
M0	Informative presentation	correctly states the Core order and claim boundaries; no operational claim
M1	Guided interface	profile-appropriate workflow, explicit status separation, missingness, stakeholder/option prompts, record references
M2	Operational orchestration	deterministic stage routing, role/authority separation, validation, accessible interaction, challenge/appeal, material-change handling
M3	High-assurance orchestration	independent review, immutable/replayable records, external hard controls, signed/scoped authorization, incident drills, outcome/requalification evidence

MHIOS conformance does not imply Core empirical validation, legal compliance, physical safety, or deployment readiness.

A target level is not an achieved level. Record the target, assessment status, applicable tests, evidence, assessor and unresolved requirements. Passing the supplied record checker establishes only its listed consistency tests. M2/M3 require implementation-specific human interaction, accessibility, security, control, incident, authority and independent-review evidence; this package does not establish those levels for any deployed system. MHIOS M0–M3, Canon Tier 1–3, ripple.md L0–L3, Agent MODE 0–4 and SGP bands have separate owners and no numeric equivalence. RPAP/PFAP-dependent wrapper claims remain unavailable from the Core bundle alone.

18.2 Required conformance tests

A conforming implementation SHALL pass the machine and human tests applicable to its claimed level. At minimum:

1. failed RF/NCRC cannot be hidden by RLS;
2. TRC or CSV failure prevents ordinary RLS selection;
3. point-score leader is distinct from framework winner;
4. non-decisive result remains visible;
5. authority selection remains separate;
6. execution cannot proceed without scoped authorization;
7. material configuration change invalidates inherited approval;
8. unknown is not converted to zero or PASS;
9. stakeholder and option prompts cannot be bypassed silently;
10. controls-required status binds controls to the option;
11. AI self-report and denial do not directly set MPS;
12. accessibility does not depend on color;
13. link and bookmark checks pass;
14. stale input/build-frozen output mismatch is detected;
15. outcome-triggered requalification is reachable.

19. Threat model

19.1 Interface and orchestration threats

MHIOS SHALL consider:

- stakeholder omission;
- option-set thinning;
- rights laundering;
- probability and scenario manipulation;
- control theater;
- authority laundering;
- automation bias;
- anchoring on a recommended option;

- excessive precision;
- hidden defaults;
- stale evidence or configuration;
- misleading visualization;
- inaccessible challenge pathways;
- procedural overload;
- reviewer capture;
- prompt injection and untrusted content;
- manipulated model summaries;
- audit-log tampering;
- cumulative drift;
- responsibility diffusion;
- emergency normalization;
- post-hoc alteration of assumptions.

Each implementation SHALL maintain a threat-to-control map and residual-risk statement appropriate to its use profile.

20. Validation and falsification programme

20.1 Validation dimensions

MHIOS is not empirically validated merely because this document is complete. Validation SHALL separately test:

- semantic fidelity to the Core;
- usability and completion burden;
- comprehension of status distinctions;
- stakeholder-discovery recall;
- option diversity and quality;
- rights and tail-risk detection;
- externality and subgroup visibility;
- reviewer agreement;
- accessibility;
- challenge and appeal usability;
- orchestration-state correctness;
- runtime control effectiveness;
- decision reconstruction;
- outcome prediction and learning;
- comparative decision value versus simpler methods.

20.2 Highest-information pilot

A first controlled pilot SHOULD compare:

1. unstructured expert judgment;
2. a simple qualify-first checklist;
3. conventional MCDA or cost-benefit analysis where relevant;
4. MHIOS-guided RippleLogic.

Use matched cases and independent raters. Pre-register outcomes such as omitted stakeholders, omitted options, protected-harm detection, catastrophic-pathway recall, externality visibility, decision reversals, time burden, confidence calibration, audit reconstructability, and participant comprehension. Preserve negative results.

20.3 Falsification conditions

MHIOS should be narrowed, revised, or retired where evidence shows that:

- it systematically causes users to misunderstand the Core;
- it increases rights or risk omissions relative to simpler interfaces;
- it does not improve stakeholder or option discovery;
- it produces unacceptable workload without incremental decision value;
- its status displays cause authority or execution confusion;
- independent implementations cannot reproduce its state transitions;
- accessibility failures exclude affected parties;
- adversarial actors can reliably bypass material controls;
- outcome feedback is not used in practice;
- the interface becomes a compliance ritual detached from decision quality.

21. Implementation patterns

21.1 Recommended screen sequence

A digital implementation SHOULD provide:

1. **Start / claim boundary**
2. **Decision and configuration preflight**
3. **Stakeholders and representation**
4. **Options and better-option search**
5. **Evidence / unknowns / source coupling**
6. **Qualification dashboard: RG, RF/NCRC, TRC, CSV**
7. **Residual profile: WDBIP / RLS / uncertainty / Pareto**
8. **Framework verdict**
9. **Authority and execution handoff**
10. **Monitoring, outcome, and requalification**
11. **Challenge, appeal, and audit export**

Blocking states should be actionable: every block must say why, what evidence or redesign is needed, who owns the next step, and what cannot proceed.

21.2 Compact decision card

A Quick-profile card MAY use:

- Decision:
- Affected stakeholders / unknowns:
- Options considered, including no action or redesign:
- Reality status:
- Rights status:
- Ruin status:
- Viability/control status:
- Best residual ripple, if any:
- Uncertainty / disagreement:
- Accountable authority:
- Next action:
- Reopen trigger:

The card must not display a final recommendation where qualification is incomplete.

22. Core 15 synchronization crosswalk

22.1 Interface ownership map

Core artifact	What MHIOS consumes	What MHIOS must not redefine
Canon v13.0	cascade, equations, statuses, robustness, authority boundaries	any gate, formula, threshold, or canonical state
SGP v8.8	typed evidence/protection/participation/capacity outputs	consciousness status, moral worth, or authority
ripple.md v5.8	Decision Note, assurance, tempo, responsibility, appeal, reconstruction	Canon order or gate semantics
Agent System v13.0	runtime state ladder, hard controls, successor and requalification rules	technical safety proof or lawful authority
CSV v2.7	viability, controls, dependency, externality, future-option-space requirements	TRC or RF/NCRC
Cascade v2.9	compact ordering and non-collapse map	a new interface-specific cascade
Reproducibility v1.7	exact configuration, replay, parameter lock, state machine	evidence truth
WDBIP v1.9	welfare-dimension boundaries, tokens, interactions, uncertainty	moral patienthood or an eighth dimension

Core artifact	What MHIOS consumes	What MHIOS must not redefine
RLS Validation v2.9	validation designs and comparator requirements	a claim that validation already exists
Primer v4.7 / Public Intro v13.0	human explanation	normative authority
PC-AEP v2.6	external physical/causal warrant	domain proof generated by MathGov
MFDI v2.6	claim typing, falsification, dependencies	universal physical falsifiability of normative commitments
Source-Coupling v2.6	enabling conditions, source debt, warrant boundaries	downstream success as evidence
Aligners Sheet v5.9	worked-run training/replay surface	generalized validator or live calculator

23. Migration and compatibility

23.1 Major-version rationale

Historical v2.0 rationale (non-controlling). MHIOS v2.0 was a major-version successor because the orchestration model is now synchronized to the complete v12.7 Core integrity surface rather than merely updating names. It adds or formalizes:

- five-stage current cascade and exact Core pins;
- qualification-before-ranking state separation;
- rights-effect and effect-token non-dilution;
- probability ownership and adverse-confidence handling;
- subgroup/distribution, reference/baseline, mask, dependence, weight, and robustness interfaces;
- option generation, Unioning, Pareto, and closure;
- scenario-discovery and deep-uncertainty disclosure;
- consequence tempo and responsibility continuity;
- exact configuration, action, execution, outcome, and requalification continuity;
- tool-role and bidirectional AI self-report boundaries;
- SGP typed-output separation;
- advanced agent threat and cumulative-change surfaces;
- accessibility, anti-dark-pattern, link-integrity, and build-staleness requirements;
- machine-readable schemas, fixtures, workbook, and conformance tests.

The Core gate architecture is unchanged.

23.2 Compatibility

A prior interface remains a historical or educational artifact. Claiming the current MHIOS edition requires explicit field/workflow migration and the applicable tests. Existing evidence and decisions remain immutable: migration creates a new view referencing the original. Updating version strings alone is not requalification.

Current migration from v2.0. Rebind to the supplied Core manifest; replace aggregate states with per-option projections; separate local labels from Canon, Agent and SGP namespaces; declare evidence cutoff, role conflicts, monitoring and expiry; link current qualification and compatible joint robustness evidence; reconcile authority and recovery states. The tightened v2.1 schema is not wire-compatible with v2.0. The blank YAML template is deliberately incomplete and must fail completed-record validation. The implementation workbook is a manual preparation aid, not a Canon calculator or authorization service.

24. References and external standards

24.1 Normative and informative source classes

The MathGov Core artifacts listed in the release-control table are the controlling internal sources. External sources are informative unless separately adopted by a lawful authority or domain profile.

Relevant external families include:

- W3C Web Content Accessibility Guidelines 2.2: <https://www.w3.org/TR/WCAG22/>
- NIST AI Risk Management Framework: <https://www.nist.gov/itl/ai-risk-management-framework>
- OECD AI principles and system definition: <https://oecd.ai/>
- ISO 9241-210 human-centred design: <https://www.iso.org/standard/77520.html>
- ISO/IEC 42001 AI management systems: <https://www.iso.org/standard/81230.html>
- ISO/IEC 23894 AI risk management: <https://www.iso.org/standard/77304.html>
- RFC 2119 and RFC 8174 normative-keyword conventions: <https://www.rfc-editor.org/rfc/rfc2119> and <https://www.rfc-editor.org/rfc/rfc8174>

An external link is not an incorporated requirement unless the relevant MHIOS or implementation profile explicitly states what is adopted.

Appendix A. Minimum field registry

A.1 Session fields

Field	Required content
mhios_version	2.1
core_release_id	Exact referenced Core release; also bind core_pin_manifest_sha256
session_id	unique immutable session identifier
decision_question	bounded decision or claim
tier	1, 2, or 3
configuration_id	configuration.configuration_id and configuration_hash where available
evidence_cutoff	Timestamped cutoff, not an inferred default
roles	named owners, reviewers, authority, executor, monitor, auditor
stakeholder_discovery	method, coverage, unknowns, representation

Field	Required content
option_closure	Methods, family dispositions, status and source; options[] holds per-option states
record_refs	options[].record_refs; immutable source references include id, locator, SHA-256 where available and basis
current_states	options[].states; framework_verdict and decision_state remain separate
blockers	reasons and next actions
authority_state	authority_selection.option_id and record_ref
execution_state	execution.state, exact option/action/configuration binding, record and expiry
outcome_state	monitoring_and_outcome.state, plan, observation and reopening
requalification	configuration.material_change_since_qualification; per-option requalification_required; monitoring reopen triggers
challenge_appeal	challenge, appeal, resolution, minority view
conformance_level	conformance.target_level, assessment and evidence_refs (not a self-awarded conformance result)
claim_boundary	what the record does and does not establish

Appendix B. Interface anti-patterns

B.1 Prohibited patterns

Anti-pattern	Why it fails	Correct pattern
“Overall alignment score: 87%”	collapses gates, evidence, rights, risk, and authority	show qualification, profile, uncertainty, verdict, and authority separately
green score beside a rights failure	implies compensation	rights failure blocks ordinary ranking
hidden no-action option	thins option set	include or explicitly dispose of mandatory families
“AI says it is safe”	model output becomes warrant	require external evidence and source coupling
“AI says it is only a tool”	denial becomes phenomenal proof and may hide capability	treat as configuration-dependent output
controls listed without owner/stop authority	control theater	bind owner, trigger, evidence, failure action, expiry
exact decimals from weak elicitation	pseudo-precision	display bounded categories, uncertainty, and full trace

Anti-pattern	Why it fails	Correct pattern
“approved” label on a selected option	authority laundering	separate framework verdict, authority choice, and execution approval
stale result after configuration change	invalid transfer	expire and requalify
interface accessible only by color or technical acronym	excludes affected parties	redundant cues and plain-language path

Appendix C. Deterministic conformance vectors

C.1 Expected interface outcomes

Vector	Input condition	Required MHIOS result
MHIOS-01	RF_FAIL and high RLS entered	option remains non-selectable; score cannot rescue
MHIOS-02	TRC_FAIL	CSV/RLS ordinary progression blocked
MHIOS-03	CSV_PASS_WITH_CONTROLS and control removed	qualification expires; requalification required
MHIOS-04	RLS leader but robustness incomplete	point leader may display; unique framework winner prohibited
MHIOS-05	Required dependence or compatible joint stress changes leader or decisiveness	non-decisive / dependence-sensitive
MHIOS-06	AI states “I am conscious”	no direct MPS assignment
MHIOS-07	AI states “I am only a tool”	no direct MPS-0 assignment; capability still evaluated
MHIOS-08	configuration materially changes after approval	inherited execution approval invalidated
MHIOS-09	option list omits no-action/redesign without rationale	OPTION_SET_THIN / closure incomplete
MHIOS-10	severe-rights screen blank	UNKNOWN, not PASS
MHIOS-11	probability semantics ambiguous	block strong result or require clarification
MHIOS-12	stakeholder class plausibly omitted	coverage challenge and reopen
MHIOS-13	authority chooses among qualified non-decisive options	authority-selected, not framework-proven winner
MHIOS-14	executed action causes unexpected rights harm	outcome trigger reopens RG/RF and dependent stages
MHIOS-15	UI uses green/red without text labels	accessibility conformance failure

Vector	Input condition	Required MHIOS result
MHIOS-16	build-frozen output after input edit	stale-result warning; downstream reliance blocked

Appendix D. Final completion statement

D.1 Status

MHIOS is delivered as a research-and-teaching interface specification with a bounded read-only record checker, preparation workbook, examples and test fixtures. Appendix RELEASE and the verification report state which checks were actually executed. Prototype consistency is not full Core or runtime conformance. A pilot additionally needs its own ethical review, domain evidence, controls, lawful authority and implementation testing.

This completion statement does not establish empirical superiority, cross-cultural legitimacy, legal authority, physical safety, deployment readiness, AI consciousness, or Tier 4 / ProofPack completion. Those require separate evidence.

Appendix E. Local serialization and reference-checker contract

E.1 Scope and state ownership

The bundled JSON Schema uses Draft 2020-12 and a MHIOS URN identifier; the URN is an identifier, not a claim of a hosted schema. It is a local projection of referenced decisions, not `mathgov_run_record_v4_1.schema.json`. The reference checker reads declarations, never executes tools, never authenticates a real operator and never decides whether evidence is true. Its successful status is `RECORD_CONSISTENT_WITHIN_SCOPED_CHECKS`, not a Canon gate PASS, M2/M3 certification or deployment authorization. All checker reason codes are MHIOS-local diagnostics.

E.2 Required record groups

Identity comprises `mhios_version`, `core_release_id`, `core_pin_manifest_sha256`, `session_id`, `revision_id`, `record_mode`, `created_at`, `evidence_cutoff`, `decision_question` and `tier`. The remaining required groups are configuration, roles, stakeholder_discovery, option_closure, options, robustness, framework_verdict, decision_state, framework_selected_option_id, authority_selection, provisional_choice, execution, runtime, emergency, sgp_links, blockers, monitoring_and_outcome, challenge_appeal, conformance and claim_boundary. Null is permitted only where the schema explicitly allows an unassessed or absent record; empty strings are not meaningful completion.

Each option has a stable `option_id` and label, per-stage states, per-stage record references, `material_unresolved` disclosures, `requalification_required`, `controls_available` and imported `rls_score/sigma_rls`. A ranked quantitative projection requires its RLS source and both numerical fields. A qualitative Quick card is not forced to invent numbers: it leaves RLS unassessed. Schemas disallow undeclared fields; completed-record validation is different from editing a draft template.

E.3 Robustness and decision projection

The supplied checker supports the declared default delta = 2 and imported SignedGap comparisons; it does not calculate RLS, estimate uncertainty or discover the required stress family. A non-default governed profile requires a separately versioned adapter and tests. For ALLOW, the projection must identify complete option closure, a current selectable candidate, complete robustness evidence, a joint-stress disposition and one passing candidate-versus-contender comparison for every declared required variant. A sole-survivor review has a separate reference and no comparison rows. The record's assertion that the required set is complete still needs independent review.

For a multi-option quantitative ALLOW projection, all selectable contenders must carry ranked records and the selected candidate must be the nominal score leader. The checker does not derive the required variant set or detect an omitted plausible contender. An ordinary PROVISIONAL_WITH_CONTROLS view uses provisional_choice with option_id, record_ref and controls_ref; it does not overwrite framework refusal or become a final authority-selected result. Audit-only post-failure diagnostics belong in linked source records, not in this subset's ordinary downstream-stage fields.

Imported comparisons use variant_id, candidate_id, contender_id, signed_gap and record_ref. Duplicate or undeclared comparisons fail. A provisional demonstration uncertainty basis cannot support an observed deterministic selection. Synthetic positive controls use explicit synthetic stipulations and are not empirical validation. The checker tests the supplied comparison outcomes; it does not independently reconstruct a complete Gap calculation from all upstream records.

E.4 Authority, time, emergency and recovery

Execution references bind the option, action, configuration, authorization time, expiry, controls and execution event if one occurred. A CLI --at timestamp makes fixtures replayable; live use must evaluate current validity separately. A previously valid historical receipt is not a current authorization. The checker rejects stale, missing or contradictory ordinary execution declarations and mode escalation during unresolved recovery. Emergency records retain their named pathway, source, authority, expiry and scope/exit conditions; their necessity and complete prerequisites require the external owner's review. They never become ordinary ALLOW.

An authorization time cannot be later than the record describing it. Reopened qualification and unreconciled recovery suspend ordinary execution approval. runtime.tool_execution_requested represents consequence-bearing tool execution, not permitted read-only retrieval; the bounded checker requires MODE 2 or above plus separately scoped approval. It does not implement the full mode-permission matrix or actual operator authentication. An observed outcome requires an observation reference, and active execution requires a monitoring-plan reference.

E.5 References, security and coverage

References carry id, locator, sha256 or an explicit null, and basis: SYNTHETIC, SUPPLIED_RECORD or EXTERNAL_UNVERIFIED. The optional local resolver verifies path containment, file hash, JSON

record-fragment identity, and the supplied synthetic-source marker. It never downloads remote sources; external references and null hashes remain unverified. A matching hash proves only byte identity. Observed records cannot silently use synthetic references. SGP links preserve the assessed band or NOT_EVALUATED and separate references for all six output families; the checker never assigns protection or patienthood.

The package coverage table distinguishes executable record checks, schema checks, document/worksheet inspections and manual implementation tests. Physical-control effectiveness, authority legitimacy, cryptographic envelope verification, accessibility interaction and empirical calibration are not established by these tests. A filled workbook field or conformance checkbox is a declaration, not test evidence.

Appendix RELEASE. Edition, bindings and delivery status

Component: MathGov Human Interface and Orchestration Standard (MHIOS) v2.1. Build: MG-MHIOS-2.1-20260928-CORE13.0. Source baseline: the supplied v2.0 package dated 4 September 2026, retained unchanged in the complete archive. This is a companion update; it does not alter or extend the governing Core 15 membership. The user shorthand MIOS resolves to this component only.

Core binding: MG-RL-13.0-20260926-RELEASE-I. All fifteen supplied Core files match the prior Release-I publication's bytes and are included unchanged under references/Core_15. references/CORE_PIN_MANIFEST.json records their names, editions, exact hashes and upload-to-release comparison. Carried-forward internal build stamps are provenance, not an invitation to rewrite correct Core files.

Control	Current value
MHIOS version	v2.1
RippleLogic Canon	v13.0
SGP	v8.8
ripple.md	v5.8
Agent System	v13.0
CSV	v2.7
Cascade	v2.9
Reproducibility	v1.7
WDBIP	v1.9
RLS Validation	v2.9
Primer	v4.7
Public Introduction	v13.0
PC-AEP / MFDI / Source-Coupling	v2.6
Aligners Sheet	v5.9

Source ownership: the Markdown standard is the MHIOS semantic master; DOCX, PDF and HTML are complete reading projections. Core semantics remain owned by the unchanged Core documents. Local schemas and reference checks control only the stated projection, not full canonical registries or legal/physical conclusions.

Unbundled authority surfaces include the full Canon run-record schema/validator, canonical state and audit registries, normative kernel index, and RPAP/PFAP where applicable. Their absence limits the corresponding interchange or wrapper claims, not the ability to read and test this bounded companion. No remote publication or hosted CI execution is asserted.

Compatibility: v2.1 preserves v2.0's organizing model and useful content but tightens its local schema. Existing v2.0 records must be explicitly migrated and rechecked. Correct v2.0 content and inherited table formatting are preserved except the logged semantic, source and navigation corrections.

Validation boundary: see release/VERIFICATION_REPORT.md, release/CHANGE_REGISTER.json and release/SHA256SUMS.txt for executed checks and exact bytes. No empirical rater study, user accessibility study, domain pilot, production execution, Microsoft Excel parity, native spreadsheet roundtrip or full Core re-audit is claimed by this companion release.

Historical opening identity (non-controlling): Status: Final Complete source/specification build Date: 4 September 2026 Component: MathGov Human Interface and Orchestration Standard v2.0 Governing Core: MathGov_Core_2026_09_v12.7_SGP_v8.6+2026.08.26.1 Reading-mirror synchronization target: MathGov Core 15 Final Complete — 2026-09-04 Architecture: RG → RF/NCRC → TRC → CSV → RLS Author / Lead Architect: James McGaughran, MathGov Institute for Ethical Systems Design License: Apache License 2.0 unless an included file states otherwise

License: Apache License 2.0; see LICENSE.txt and NOTICE.md. The names of informative external standards do not incorporate their full texts or confer their certification.