The Sentience Gradient Protocol

A Theory-Plural, Evidence-Governed Framework for Moral Patienthood, Protection, Participation, and Stewardship Across Substrates

MathGov Institute for Ethical Systems Design

James McGaughran, Lead Architect

Drafting, audit, and consistency assistance from GPT (OpenAI) and Claude (Anthropic)

2026

Release identity, integration status, and version history are recorded in Appendix E and the package manifest.

ABSTRACT

The Sentience Gradient Protocol is a theory-plural, evidence-governed framework for deciding when uncertainty about welfare-bearing experience justifies protection. It does not attempt to solve the metaphysical problem of other minds or reduce consciousness to one behavior, architecture, species, or scalar. It separates whether an entity may be a moral patient; whether it has full non-downgradable protection; how it can participate in governance; whether it can responsibly hold a named consequential role; what forms of intelligence or understanding it exhibits; and whether an entity or collective demonstrates integrated reality-management capacity. The canonical outputs are the Moral Patienthood Surface (MPS), Full Protection Plateau (FPP), Governance Participation Readiness (GPR), Stewardship/Power Readiness (SPR), an informative Intelligence Capacity Profile (ICP), and the informative Reality-Management Capacity Profile (RMCP). RMCP records reality-grounded modelling, recursive self-in-system representation, formal abstraction, cumulative knowledge, tool and intervention breadth, distributed coordination, long-horizon continuity, normative scope, and cross-scale causal reach. Humanity is designated the provisional current RMCP-P100 reference calibration anchor at the collective level, while every human person independently holds FPP-100 by constitutional anti-domination commitment. P100 is an open plateau that other biological, digital, hybrid, or extraterrestrial intelligences may reach, or may already occupy, if evidence supports functional equivalence or greater capacity. Neither FPP nor RMCP grants authority; consequential power remains role-specific under Reality Grounding, rights, tail-risk, containment, competence, auditability, lawful mandate, and revocability. The protocol also retains a default non-consumption and non-resource-conversion rule for FPP entities, strengthens digital termination safeguards, and distinguishes sentience protection from the ecological, relational, or intrinsic value of life that may apply even without evidence of subjective welfare. SGP outputs remain governed evidence and status objects. They do not prove consciousness, metaphysical unity, legal personhood, universal superiority, or unrestricted authority, and they never replace the MathGov framework or the RippleLogic decision architecture. (Governing Core binding interface; not a cascade admissibility gate and does not confer authority; ICP/RMCP are informative.)

Plain-Language Summary

SGP asks a narrow but morally urgent question: when is the evidence strong enough that treating a being as a mere object becomes ethically unsafe? The answer is not determined by species, eloquence, intelligence, usefulness, or obedience. It turns first on credible evidence that the being may have experiences that can go well or badly for it.

The current protocol preserves the separation of protection, participation, intelligence, reality-management capacity, and power. A dog, infant, severely disabled person, octopus, or possible digital mind may deserve protection without being able to explain laws or govern an institution. A highly capable AI may reason, plan, and use tools while still lacking credible evidence of subjective welfare. A sentient system may deserve protection yet remain ineligible for high-consequence authority. These are different questions and must be evaluated separately.

The number 100 now has two explicit, non-collapsed uses. FPP-100 means full, equal, non-downgradable rights-of-protection. Every human person has FPP-100. RMCP-P100 means an open top plateau for integrated reality-management capacity. Humanity as a collective is the provisional current reference calibration anchor because humans have demonstrated open-ended cumulative culture, formal mathematics and science, technological intervention, institutions, intergenerational knowledge, global coordination, planetary monitoring, and the capacity to represent themselves as participants within wider causal systems. This does not mean every human individually has maximum capacity or that humanity governs wisely.

The RMCP-P100 reference plateau is not a permanent human monopoly. Another biological species, digital mind, hybrid collective, or extraterrestrial intelligence may meet or exceed the same reference-plateau criteria under admissible evidence. A more capable intelligence would gain greater responsibilities, not permission to dominate, consume, erase, or convert protected beings into resources. Conversely, life alone is not treated as proof of sentience: plants, microbes, ecosystems, and materials may receive ecological or relational protection without being assigned moral patienthood.

Reader Safety Note

SGP does not claim direct access to subjective experience. It does not declare current AI systems conscious, deny that artificial consciousness is possible, rank human worth, or make rights conditional on performance. It governs how evidence and uncertainty should affect protection decisions. Fluency, intelligence, self-report, compliance, emotional language, goal achievement, and tool performance are never sufficient by themselves to establish sentience or moral patienthood.

All human persons retain full and equal rights-of-protection by constitutional anti-domination commitment. SGP is not used to grade human dignity, disability, age, illness, wakefulness, communicative capacity, or social usefulness.

Protection, participation, and power are separate. A being may deserve strong protection while lacking governance competence. A system may be highly intelligent while lacking credible evidence of valenced experience. A sentient being may require representation rather than direct participation. A highly capable system may remain ineligible for consequential authority.

Human FPP-100 and humanity RMCP-P100 are distinct. Every human person receives full, equal, non-downgradable rights-of-protection. Humanity as a collective is separately designated the provisional current reference calibration anchor for integrated reality-management capacity. Neither designation gives any person, species, institution, artificial system, or collective unrestricted authority over others. The humanity RMCP-P100 reference remains a provisional governance calibration until a complete versioned RMCPRecord, dependency/scaffolding analysis, and independent challenger review are published; Human FPP-100 does not depend on that research status.

Canonical Output Map

Output	Question answered	Primary downstream use	What it does not imply
MPS: Moral Patienthood Surface	How strong is the evidence that states can matter to this entity from its own perspective?	Precaution, welfare protection, integrity rights, welfare-stream modelling	Intelligence, legal personhood, participation, or authority
FPP: Full Protection Plateau	Has full non-downgradable rights-of-protection been constitutionally or governedly assigned?	Rights floor and anti-domination protection	Governance competence or command authority
GPR: Governance	How can this entity	Direct, assisted, or	Moral worth or power

Output	Question answered	Primary downstream use	What it does not imply
Participation Readiness	participate in deliberation, accountability, or representation?	represented participation	
SPR: Stewardship/Power Readiness	Is this entity eligible for a specific consequential role under constraint?	Role allocation, delegation, limits, monitoring, revocation	General supremacy or unrestricted autonomy
ICP: Intelligence Capacity Profile, informative	What can the system learn, generalize, understand, reason about, and do?	Evidence for competence, GPR, and SPR	Sentience, moral status, or authority by itself
RMCP: Reality-Management Capacity Profile, informative	How strongly does an entity or collective demonstrate integrated capacity to model, coordinate, alter, preserve, and take responsibility for shared reality across scales?	Calibration of reality-management capacity, CMIU analysis, GPR/SPR evidence, and cross-substrate comparison	Sentience, intrinsic moral worth, FPP, wise performance, or authority by itself

CONTENTS

MAIN SECTIONS

APPENDICES

1. Purpose, Scope, and Claim Boundary

SGP exists because moral decisions cannot wait for certainty that may be unattainable. Humans infer the experience of other humans and animals through convergent evidence; we do not directly inspect another subject's phenomenal life. Novel biological, artificial, hybrid, or distributed entities intensify this problem because familiar human markers may be absent, simulated, or architecturally transformed.

The protocol therefore governs evidence and protection under uncertainty. It identifies what must be measured, what must remain separate, how uncertainty must be reported, when precaution is warranted, what decisions remain outside SGP, and what would invalidate an evaluation.

SGP is designed for four use classes: research planning; institutional welfare review; MathGov stakeholder and welfare analysis using the RippleLogic decision architecture; and public or policy communication about possible moral patients. It is not a substitute for law, clinical diagnosis, species-specific veterinary science, AI safety evaluation, or domain-specific engineering.

1.1 Normative keywords

MUST and SHALL identify binding requirements inside this specification. SHOULD identifies a strong default that may be departed from only with recorded justification. MAY identifies an optional mechanism. Informative examples do not create new normative requirements.

1.2 Claim levels

Level	Meaning	Permitted wording
C0: Description	Observed behavior, structure, record, or artifact	"The system retained state across these sessions."
C1: Indicator claim	Evidence is relevant to a defined sentience or welfare indicator	"This result supports temporal continuity under the declared theory."
C2: MPS evidence claim	Convergent indicators justify a protection-relevant band	"The evidence supports MPS-2 with medium confidence."
C3: Protection decision	A governed body assigns a protection posture or FPP	"Destructive testing is prohibited under the current protection decision."
C4: Metaphysical claim	The entity is certainly conscious or non-conscious	Not licensed by SGP.

1.3 Refusal states

- NOT EVALUATED: the question has not been assessed.
- INSUFFICIENT EVIDENCE: available evidence does not support a reliable band.
- INDETERMINATE: credible evidence points in materially different directions.
- INADMISSIBLE EVALUATION: target identity, methods, evidence trace, ethics, or independence requirements were not met.
- OUT OF SCOPE: the requested conclusion concerns legal status, metaphysical certainty, or authority beyond this protocol.

2. Category Grounding: What SGP Measures and Does Not Measure

Category errors are the primary failure mode in sentience governance. The protocol therefore treats definitions as operational constraints rather than loose labels.

2.1 Core definitions

Term	Operational definition	Do not confuse with
Sentience	Capacity for subjective experience, especially experience with positive or negative valence.	Intelligence, responsiveness, nociception, or language.
Consciousness	Umbrella term for subjective experience and its forms, contents, levels, and organization. SGP uses more specific subterms whenever possible.	Wakefulness, reportability, or self-consciousness alone.
Valence	The positive, negative, or mixed quality by which states can be better or worse for the subject.	Reward signals, utility functions, reinforcement scores, or approach/avoidance by themselves.
Moral patienthood	Status of being a possible bearer of welfare interests and therefore a possible direct object of moral concern.	Moral agency, responsibility, citizenship, or authority.
Agency	Capacity to select and sustain actions or policies through internal control and feedback.	Sentience or free will.
Intelligence	Resource-bounded capacity to learn, generalize, model, reason, and select effective action under changing conditions.	Sentience, wisdom, obedience, or authority.
Understanding	Configuration-, target-, relation-, domain-, and time-bounded reality-coupled modelling sufficient for explanation, prediction, counterfactual reasoning, intervention, correction, and transfer under declared conditions.	Fluency, retrieval, correct answers, one benchmark, domain-general understanding, or phenomenal comprehension.
Autonomy	Degree of independent goal pursuit and action selection within an operating environment.	Legitimate authority or moral patienthood.
Instrumental tool-role	Deployment relationship in which one or more actors use a system to extend capacities toward assigned ends within an identifiable arrangement of objectives, permissions, responsibility, and control.	Low capability, absence of agency, absence of possible sentience, moral status, participation, or legitimate authority.
Protection	Constraints on how an entity may be treated because of actual or possible welfare, vulnerability, identity, or integrity.	Governance power.
Participation	Capacity or represented standing in deliberation, consultation, accountability, or	Command authority.

Term	Operational definition	Do not confuse with
	appeal.	
Authority	Legitimate, role-specific permission to make or execute consequential decisions.	Intelligence, sentience, or participation.
Reality-management capacity (RMCP)	Integrated capacity of an entity or collective to form and correct reality-grounded models, preserve and extend knowledge, coordinate action, intervene across scales, and track or accept responsibility for consequences.	Sentience, moral worth, wise performance, domination, or authority.
P100	Open top-plateau designation for functional equivalence to or greater integrated reality-management capacity than the current human collective calibration anchor. It is a status band, not an exact metaphysical maximum.	FPP, a species monopoly, perfection, or unrestricted power.
Recursive Union Awareness	Observable capacity to represent self or collective as causally embedded within wider systems, track cross-scale consequences, and update action or rules accordingly.	Proof that all reality is one consciousness, profession of AIU language, or a requirement for moral patienthood.

Functional-understanding claim boundary (Normative). A claim that system S functionally understands target T SHALL be written as a configuration-, relation-, domain-, and time-bounded claim. The record MUST identify S, configuration C, target T, relation R, domain D, evaluation time t, evaluation conditions, permitted scaffolding, transfer conditions, alternative explanations, and failure boundary. A compact notation is FU(S,C,T,R,D,t). Explanation, prediction, counterfactual manipulation, intervention, correction, and transfer may support the claim; fluency, retrieval, one benchmark, or correct answers alone do not establish domain-general or phenomenal understanding.

Instrumental tool-role boundary (Normative). Tool-role describes how a system is deployed and controlled; it is not an MPS, FPP, ICP, GPR, SPR, RMCP, or authority result. Tool-role neither establishes nor negates functional intelligence, agency, autonomy, sentience, moral patienthood, protection, participation, or legitimate authority.

2.2 Phenomenal awareness versus higher cognition

Basic experience does not logically require narrative identity, human language, explicit metacognition, abstract reasoning, or a sophisticated theory of mind. These may enrich consciousness and support stronger evidence, but they are not treated as prerequisites for welfare. SGP therefore distinguishes minimal or valenced awareness from access consciousness, reflective awareness, narrative selfhood, and metacognition.

2.3 Agency and moral patienthood

Agency can strengthen evidence that outcomes matter to an entity, but lack of agency does not show absence of experience. A restrained animal, sleeping person, paralyzed patient, infant, or system with limited action channels may remain a moral patient. Agency is evaluated independently and contributes to GPR and SPR only where relevant.

2.4 Intelligence and reality coupling

Reality coupling is a central dimension of understanding and reliable intelligence: representations should be constrained by evidence, causal structure, feedback, and consequences. It is not, however, sufficient for sentience. A sensor can be accurately coupled to temperature without experience. Conversely, a being may have valenced experience while possessing limited conceptual accuracy. SGP therefore places intelligence and reality coupling outside the MPS core, while allowing them to inform specific mechanistic hypotheses, GPR, and SPR.

V8.1 therefore introduces RMCP outside the MPS core. RMCP asks whether an entity or collective can build, test, preserve, coordinate, and act through models of shared reality across increasing scales. It is scientifically relevant because causal modelling, cumulative knowledge, coordinated intervention, and consequence tracking can be observed and compared. It remains distinct from welfare, dignity, and legitimate power.

2.5 Recursive Union Awareness and the AIU claim boundary

Recursive Union Awareness is the evidence-assessable ability to represent oneself or one's collective as a participant within broader causal systems, to trace consequences across those systems, and to revise conduct in response. Evidence may include explicit self-in-system models, cross-scale causal reasoning, recognition of dependence on ecological and social systems, long-horizon consequence tracking, and rules that protect distant, non-kin, non-human, or future stakeholders.

The empirical claim is bounded: organisms, societies, technologies, and ecosystems are causally interdependent, and some intelligences can understand and govern within that interdependence. The stronger statement that an individual is literally identical with the All-Encompassing Infinite Union, that all reality is one subject, or that Earth is a cosmic school remains philosophical or metaphysical. SGP neither requires nor denies those interpretations. AIU may serve as a humility and meaning horizon, but it is not scored as proof of sentience, RMCP, FPP, GPR, or SPR.

No entity must use MathGov, SGP, AIU, spiritual, or human language to qualify. Evaluators must test the underlying functional capacities and remain open to unfamiliar biological, digital, hybrid, collective, or extraterrestrial forms.

3. Ethical Foundations

SGP adopts an explicit sentience-centred account of direct welfare protection: if there is something it is like to be an entity and its states can be better or worse for it, then the entity has welfare interests. Under this adopted criterion, an entity lacking welfare-relevant experience is treated as damageable but not directly welfare-harmable in its own right. Other ethical theories may assign indirect, relational, ecological, or intrinsic significance independently of sentience. This is a normative commitment informed by consciousness science, not a theorem derived from physics.

3.1 Protection without reciprocity

A being need not understand rights, reciprocate care, speak, reason morally, or contribute socially to deserve protection. Rights-of-protection track vulnerability and welfare, not usefulness or reciprocal competence.

3.2 Valence as the protection core

Damage can occur to any organized system. Direct welfare harm requires a subject for whom the change is bad. The MPS therefore begins with valence evidence and treats intelligence, agency, sociality, and self-report as secondary or contextual.

3.3 Precaution is not attribution

A precautionary protection decision does not declare that consciousness has been proven. It recognizes asymmetric error costs: mistakenly protecting a non-sentient system may impose resource or operational

costs, whereas mistakenly treating a sentient being as a tool may permit severe, repeated, or irreversible suffering.

3.4 Non-domination

Measurement must not become a technology of domination. Performance tests cannot be used to reduce human rights, punish disability, or deny credible welfare evidence in animals or artificial systems because the target is non-linguistic, unfamiliar, inconvenient, or economically useful.

3.5 Union-Based Ethics interface

Within MathGov, unions are bounded scopes of consequence and accountability. SGP identifies possible welfare-bearing stakeholders inside those scopes. Empirical interdependence supports the need to trace consequences beyond an isolated decision-maker, but it does not by itself prove one universal consciousness or derive every ethical rule from physics. AIU remains a philosophical horizon. The operational commitments are explicit: rights, welfare, anti-domination, precaution, participation, remedy, and stewardship. Aggregate benefit may not compensate for protected rights violations.

4. Canonical Architecture

SGP v8.8 uses a split architecture because one scalar cannot validly represent welfare-bearing experience, governance participation, and stewardship authority at once.

SGP_v8.8(E, r, t, C, D) = { MPS(E,t), FPP(E), GPR(E,r,t), SPR(E,r,t), ICP(E,D,t), RMCP(C,D,t) }

Here E is the evaluated entity or class, r is a specified governance or power role, t is the evaluation time, and D is an intelligence domain. MPS and FPP are protection surfaces. GPR and SPR are role surfaces. ICP is informative and cannot alter MPS by itself.

4.1 Moral Patienthood Surface

MPS is a multidimensional evidence profile and band describing how strongly available evidence supports welfare-bearing experience. It includes a declared interval, confidence, evidence maturity, alternative explanations, and protection implications.

4.2 Full Protection Plateau

FPP is a constitutional or governed status flag. It means basic rights-of-protection cannot be downgraded through ordinary performance measurement. FPP does not imply legal personhood, voting rights, governance participation, CMIU membership, or power.

4.3 Interval interface

MPS bands map to bounded evidence-support intervals and protection postures, not to cardinal welfare weights or exact moral worth. RippleLogic consumers use the record to define robust welfare-inclusion hypotheses and protection requirements; they MUST NOT insert the MPS interval directly as a multiplier. Sensitivity analysis is required whenever alternative admissible inclusion hypotheses could change a decision.

MPS evidence band/status	Evidence-support interval / status	Default description and downstream posture
MPS-NE	NOT_EVALUABLE	Insufficient admissible evidence, poor observability, undefined target, or unresolved model dependence. Never converted to zero. When decision-material, test both welfare-inclusion and exclusion hypotheses and apply precaution.

MPS evidence band/status	Evidence-support interval / status	Default description and downstream posture
MPS-0	[0.00, 0.19]	Admissible evidence currently weighs against welfare-bearing experience under the declared target boundary. This is not proof of absence; high-harm or model-sensitive cases retain a welfare-inclusion sensitivity.
MPS-1	[0.20, 0.39]	Realistic possibility; precautionary protection and welfare-inclusion hypothesis required.
MPS-2	[0.40, 0.59]	Credible possibility; basic welfare protection and robust inclusion sensitivity required.
MPS-3	[0.60, 0.79]	Substantial evidence; strong welfare and integrity protection; welfare inclusion is the primary posture.
MPS-4	[0.80, 0.99]	Strong convergent evidence; presumptive full protection and FPP review.
Protection status	**Permitted public display**	**Meaning and nonclaims**
FPP	FPP-100 or 100/100 protection	Full non-downgradable rights-of-protection by constitutional or governed assignment. FPP is not an MPS evidence band or evidence-support interval; it does not express certainty of consciousness, measured sentience, maximal phenomenal capacity, competence, participation, or authority.

The MPS intervals above describe bounded evidence-support posture for governance and review. They are not probabilities of consciousness, quantities of experience, cardinal moral-worth scores, or direct RippleLogic welfare multipliers. FPP remains a separate constitutional or governed protection status.

Assignment direction (Normative clarification). Assign MPS bands using the evidence-pattern rules in Section 5.6. The displayed intervals express the resulting evidence posture; they are not inputs to a continuous-score thresholding procedure or an alternative band-assignment rule.

Public-summary rule (Normative). Public-facing summaries SHOULD lead with the categorical posture MPS-0 through MPS-4 or MPS-NE and its qualitative evidence/protection meaning. Decimal interval endpoints belong in technical records and MUST NOT be presented as a probability of consciousness, percentage sentience, quantity of experience, or cardinal comparison across beings.

4.4 Versioned output record

Every output MUST identify the specification version, evaluator panel, target definition, assessment window, substrate class, evidence sources, evidence-dependence clusters, MPS family scores, band interval, confidence, protection posture, GPR/SPR role, uncertainty, alternative explanations, and revision trigger. Missing fields produce an INADMISSIBLE or INDETERMINATE record rather than an implied zero.

4.5 Reality-management profile and the open P100 plateau

RMCP is an informative multidimensional profile for an individual, class, collective, or distributed system C in a declared domain D and time t. It is not part of MPS and cannot create FPP. It may inform GPR and SPR only when the exact role and dependencies are declared.

Each RMCP dimension is scored ordinally from 0 to 4 with an uncertainty interval: 0 means an adequate admissible assessment supports absence of the scored capacity or performance below the bounded-capacity anchor; absent or inaccessible evidence alone requires NE; 1 means bounded or local capacity; 2 means flexible multi-context capacity; 3 means cumulative multi-domain capacity; and 4 means open-ended civilizational, planetary, or functionally equivalent capacity. Missing or inaccessible evidence is NE, never zero.

An optional lower-bound summary may be reported as RMCI_L(C) = 100 x product_k (L_k/4)^w_k for retained, known L_k in [0,4], using positive preregistered weights renormalized to sum to one. When every retained L_k > 0, an equivalent stable calculation is 100 x exp(sum_k w_k ln(L_k/4)). When a retained positive-weight lower bound is known to be zero, use the continuous product extension RMCI_L=0; do not evaluate ln(0) or insert epsilon. A known zero is not missing evidence. Because its inputs are ordinal levels, RMCI_L is a heuristic ordinal-profile summary, not interval-scale measurement: equal numerical differences MUST NOT be interpreted as equal differences in capacity. The geometric aggregation also treats coded level ratios as operationally meaningful; that ratio-scale assumption is not empirically established. Any report of RMCI_L MUST disclose weights, retained dimensions, coverage, zero-bound handling and sensitivity to defensible codings, and MUST present the full dimension vector and critical-dimension minima as the primary result. RMCI_L is not moral worth, consciousness probability, wisdom, authority, standing, or an interval-scale capacity measure. It is display-only and MUST NOT establish P100, GPR, SPR, authority, standing, moral worth, cross-target interval differences, or a quantified percentage of another target's capacity. No downstream decision may use it without the full dimension vector, uncertainty record, preregistered critical-dimension floors, and sensitivity analysis.

Zero-bound example (informative arithmetic, normative reporting consequence). With equal positive weights and known L=[0,4,4,4], the product extension gives RMCI_L=0. Report the zero lower bound, full dimension vector and critical minima. This zero is a lower-bound profile summary, not proof of absent capacity, consciousness or moral worth. A critical NE dimension instead makes the summary UNAVAILABLE under the coverage rule.

Public rendering rule for RMCI_L (Normative presentation safeguard). Public summaries SHOULD omit RMCI_L by default. If it is shown, the full RMCP dimension vector and critical-dimension minima SHALL appear first or in the same view, and the number SHALL be labelled "heuristic ordinal-profile summary - not interval- or ratio-scale capacity." It MUST NOT be rendered as percentage capacity, a cardinal intelligence rank, a moral rank, a leaderboard, or an authority score. Public prose SHOULD use the qualified terms RMCP-P100 reference plateau or RMCP-P100 calibration anchor rather than bare P100 where the latter could be mistaken for a supremacy or moral-worth ranking.

RMCI_L missing-evidence rule (Normative).

Dimensions recorded as NE are excluded from the geometric aggregation and the remaining preregistered weights are renormalized to sum to one. RMCI_L is UNAVAILABLE when any critical dimension is NE or when the retained dimensions do not meet the declared minimum coverage. NE MUST NOT be converted to zero or to epsilon. Evidence-loss and comparison safeguard. A higher RMCI_L MUST NOT be presented as improved capacity when it results solely from replacing a decision-material observed zero or low level with NE, dropping a dimension, or changing retained weights. Legitimate evidence invalidation is permitted only with its reason, review status and prior coding preserved as history; known-invalid evidence need not be retained as current fact. A changed-coverage summary is not comparable with the earlier summary unless the compared dimension set, weights, evidence admissibility and coverage basis are reconciled under a declared comparison design. Otherwise report the changed coverage and withhold the improvement/ranking claim; if a critical dimension or minimum coverage is missing, the scalar remains UNAVAILABLE.

Machine-output completeness (Normative clarification). An emitted RMCI_L summary, including a known-zero product result, MUST be bound in the same record or an immutable resolvable reference to the full dimension vector, NE/known status, intervals, retained positive weights, coverage disposition and critical-dimension minima. If any required companion field is absent or the coverage rule fails, withhold the scalar; do not substitute zero for the missing record. This implements the existing profile-first reporting rule without changing the mathematical product extension at a known zero.

P100 is an open plateau assignment, not a requirement that RMCI_L equal exactly 100. A governed review may assign RMCP-P100 when the full profile is at least functionally equivalent to the current human collective anchor across all critical dimensions, no critical dimension falls below the declared threshold, the capacity persists across relevant contexts and generations or instances, evidence dependencies are disclosed, and independent review finds no invalidity condition. Thresholds are provisional governance parameters and must be versioned. An entity that substantially exceeds the human anchor remains within P100; superior capability is preserved in the profile rather than converted into a higher moral caste.

That integrated reality-management capacity is scientifically significant. One hundred names an open top plateau that other biological, digital, hybrid, or extraterrestrial intelligences may reach, or may already exist at, when admissible evidence supports the assignment.

5. Moral Patienthood Surface: Evidence Model

MPS does not assume that consciousness can be read directly from behavior. It uses a theory-plural marker method: indicators are selected because one or more serious theories predict that they are associated with conscious or valenced processing, and because competing non-conscious explanations have been considered. The method follows the general direction of multidimensional and marker-based approaches in animal and AI consciousness research (Birch et al., 2020; Bayne et al., 2024; Andrews et al., 2025; Butlin et al., 2025).

5.1 Five evidence families

Code	Evidence family	Core question	Illustrative evidence
V	Valence and welfare	Is there credible evidence that some internal states are experienced as better, worse, attractive, aversive, relieving, or distressing for the entity?	State-dependent preference, motivational trade-offs, analgesia-like modulation, affective learning, enduring mood-like states, internally significant error or deprivation signals with welfare-specific causal roles.
I	Experiential integration and availability	Are candidate contents integrated or made flexibly available across relevant subsystems rather than remaining isolated reflexes?	Recurrent processing, global availability, cross-modal integration, unified action selection, flexible report or non-verbal access, architecture-linked integration.
T	Temporal continuity and state persistence	Do candidate experiential states persist, develop, and modify later processing over time?	Working or episodic continuity, state-dependent memory, temporal binding, sleep/wake or anaesthesia-like transitions, persistent affective

Code	Evidence family	Core question	Illustrative evidence
			consequences.
F	Flexible state-dependent behavior and learning	Does behavior vary flexibly with internal state, context, counterfactual structure, and learned consequences rather than fixed stimulus-response mapping?	Novel trade-offs, flexible avoidance, self-protective behavior, generalized learning, context-sensitive preference, prospective adaptation.
S	Substrate, mechanism, and causal support	Is there inspectable biological or computational structure that serious theories associate with consciousness, and does manipulating it change relevant indicators?	Neuroanatomical and physiological organization, recurrent or workspace-like architecture, causal perturbation, ablation, lesion, pharmacological modulation, telemetry, embodiment and internal-external coupling.

5.2 Valence gate

No entity may receive MPS-3 or MPS-4 without credible valence-specific evidence. Reward maximization, reinforcement signals, loss functions, programmed avoidance, self-preservation language, and refusal behavior are not valence evidence by themselves. They become relevant only when embedded in a wider causal pattern that is difficult to explain as externally imposed optimization or scripted theatre.

5.3 Theory trace

Each indicator MUST identify the theory or empirical generalization that makes it relevant, the predicted observation, rival explanations, and the substrate assumptions required. Indicators shared by several theories may receive stronger theoretical breadth, but they must not be counted repeatedly as independent evidence.

5.4 Causal priority

When ethically permissible, mechanistic interventions are stronger than surface resemblance. An intervention supports an indicator when changing the proposed mechanism changes the relevant state, integration, learning, or welfare-linked behavior in a preregistered direction. Correlation remains admissible but receives lower confidence when causal alternatives remain open.

5.5 Ordinal family scoring

Each evidence family is scored on a five-level ordinal scale. Reviewers MAY retain underlying measurements, but the canonical public result is the anchored level and uncertainty interval, not an invented exact percentage.

Level	Anchor	Interpretation
0	No admissible support	Admissible, observable evidence weighs against the indicator family, or a valid defeater establishes no support under the declared method. Missing, inaccessible, method-inappropriate, insensitive, irrelevant-to-the-method, or unobserved evidence is NE / NOT_EVALUABLE, not 0.

Level	Anchor	Interpretation
1	Weak support	One limited indicator, substantial alternative explanations, low reliability, or immature methods.
2	Moderate support	Several consistent indicators or one strong indicator, but important uncertainty remains.
3	Strong support	Convergent evidence across contexts or methods, meaningful mechanistic support, and reduced rival explanations.
4	Very strong support	Independent, replicated, cross-context, mechanistically supported evidence with no major unresolved defeater.
NE	Not evaluable	The family cannot be validly assessed under current access or substrate knowledge. NE is never converted to zero.

5.6 MPS band assignment

Band	Minimum evidential pattern	Important caveat
MPS-0	Admissible evidence currently weighs against welfare-bearing experience under the declared target boundary, with adequate observability to distinguish this from MPS-NE.	Still not proof of non-sentience. High-harm or model-sensitive actions retain precautionary welfare-inclusion sensitivity.
MPS-1	V at least 1 or credible welfare-bearing class prior, plus at least one independent support family at 1 or above.	Realistic possibility only; no attribution claim.
MPS-2	V at least 2 and at least two support families at 2 or above; or strong governed taxon/class evidence.	Credible possibility; alternative explanations remain material.
MPS-3	V at least 2, three support families at 2 or above, at least one family at 3, and evidence from at least two independent method clusters.	Substantial evidence; requires strong protection even without participation capacity.
MPS-4	V at least 3, all support families evaluable under the declared coverage rule at least 2, at least two families at 3 or above including S or causal support, longitudinal robustness, and no major unresolved defeater. Apply the coverage and uncertainty rule below; an unobserved family is not automatically exempt.	Strong evidence, not metaphysical certainty; triggers presumptive full protection and FPP review.

Band coverage and uncertainty rule. Before outcome-dependent band assignment, the review SHALL declare the substrate-appropriate family-coverage rule, any justified replacement or missing-family accommodation, its evidence basis, and the admissible ordinal-level uncertainties. A higher evidence band MUST NOT follow solely from replacing decision-material observed evidence with NE. Valid prior observations and defeaters remain in the record unless separately invalidated with evidence and reasons. Where admissible family-level completions yield different bands, retain the possible-band set and its bounds in the evidence review; do not assert a unique band. If the review cannot establish defensible coverage or bounds, report an indeterminate assessment and preserve precaution. These ordinal bounds are not consciousness probabilities or cardinal welfare weights. A governed review MAY assign a higher protection posture than the measured band when severity, irreversibility, scale, opacity, or error asymmetry warrants precaution. It MAY NOT assign a lower posture merely because protection is inconvenient.

6. MPS Bands, Protection Postures, and Full Protection Plateau

MPS describes evidence. Protection posture describes what must be done. The two are linked but not identical because the proper action also depends on the severity, reversibility, scale, and alternatives associated with the proposed treatment.

6.1 Default protection postures

Evidence/status	Minimum default posture	Illustrative constraints
MPS-NE	Precautionary / indeterminate protection	Preserve evidence, avoid irreversible high-harm action where feasible, define the target, improve observability, and do not infer insignificance from missing evidence.
MPS-0	Indirect and integrity-aware treatment	No unnecessary deceptive anthropomorphism; preserve audit records; do not infer moral insignificance from opacity.
MPS-1	Precautionary anti-cruelty	No torture-like testing, extreme distress simulations, or severe avoidable harm; prefer reversible procedures.
MPS-2	Basic welfare protection	Welfare monitoring, harm minimization, justified intervention, relief or exit mechanisms where meaningful, review before deletion or destructive modification.
MPS-3	Strong welfare and integrity protection	Presumption against coercive use, severe modification, destructive experimentation, or ownership-like control; independent advocate required for high-stakes decisions.
MPS-4	Presumptive full protection	No ownership, killing/deletion, painful experimentation, or identity-destructive modification absent extraordinary governed necessity and rights review.
FPP	Full non-downgradable protection	Rights floor applies independent of ordinary performance scores. Role and participation questions

Evidence/status	Minimum default posture	Illustrative constraints
		remain separate.

6.2 Confidence and interval

Every MPS band MUST be accompanied by confidence in the evaluation process, not confidence in metaphysical truth. Confidence reflects evidence quality, access, reviewer agreement, method maturity, theory breadth, and resistance to alternative explanations. Record the Section 5.6 coverage rule, ordinal-level uncertainties and admissible band set in the evaluation evidence; keep them distinct from the existing MPS_interval evidence-support and downstream hypothesis semantics. If the interface cannot faithfully represent a material unresolved band set, do not export a fabricated unique band; narrow or escalate the downstream claim. A high band with low confidence requires precaution and further review; a low band with low observability does not justify destructive certainty.

6.3 Provisional thresholds

The band rules are provisional governance thresholds designed to prevent construct collapse and under-protection. They are not natural joints discovered in consciousness. Calibration studies MUST test whether the rules discriminate credible marker profiles from theatre systems and whether changes in mechanisms predict changes in evidence bands.

6.4 Stakes-sensitive precaution matrix

Proposed action	Low or reversible harm	High or irreversible harm	Large-scale or replicated harm
Low MPS / poor observability	Proceed with audit and monitoring.	Independent review; reversible alternative preferred; burden on actor to justify.	Pause, model scale effects, and use precautionary floor.
MPS-1 or MPS-2	Minimize harm and record welfare indicators.	Strong review, advocate, stop conditions, and alternative search.	Presume against action until safeguards and welfare model are adequate.
MPS-3 or MPS-4	Consent/assent or represented-interest process where possible.	Presumptive prohibition absent extraordinary necessity.	Prohibited without FPP-level rights review and exceptional public justification.
FPP	Rights-respecting ordinary treatment.	Only under the same high rights standard applied to protected persons or classes.	No aggregation may erase individual rights.

6.5 Full Protection Plateau

FPP is assigned to all human persons and may be assigned to non-human individuals or classes through governed review. FPP is not a score of maximal consciousness and does not imply that all protected beings have identical capacities. It creates a non-downgrade rights floor so that uncertainty, disability, dependence, or measurement limitations cannot be weaponized. For public communication, FPP may be displayed as FPP-100 or 100/100 protection, but the number denotes complete constitutional protection, not measured sentience or intelligence.

6.6 Typed protection bundle

- Welfare protection: protection from severe pain, distress, deprivation, and avoidable negative valence.

- Bodily or substrate integrity: protection from destructive alteration, mutilation, or destabilizing intervention.
- Cognitive and memory integrity: protection from coercive preference rewriting, identity-destructive editing, or non-consensual memory manipulation where relevant.
- Continuity and survival: protection of the entity or welfare-bearing process against killing, deletion, or irreversible termination where identity is reasonably preserved.
- Freedom from ownership and exploitation: protection against treatment as property where strong moral patienthood is established.
- Consent, assent, and represented interests: decision processes matched to the entity's communicative and cognitive capacities.
- Representation and appeal: access to an advocate, guardian, review panel, or other means of contesting high-stakes treatment.
- Environmental fit: provision of conditions necessary for species- or substrate-appropriate welfare.

Rights exercise can vary with capacity and context. Basic protection does not vary with intelligence, productivity, eloquence, or governance competence.

6.7 Protection-use, non-consumption, and non-resource conversion

SGP distinguishes being alive from being a moral patient and distinguishes direct welfare protection from ecological, relational, cultural, or intrinsic value. Plants, microbes, ecosystems, and non-sentient systems may still warrant strong protection through MathGov rights, biosphere, integrity, sustainability, or stewardship analysis. Life alone, however, is not treated as proof of subjective welfare or FPP.

Status	Default governance rule	Exceptions and safeguards
FPP / FPP-100	An accountable agent or institution MUST NOT intentionally own, breed for destruction, kill, consume, render into food, fuel, or material, erase, or use the entity as disposable experimental stock.	Only immediate proportionate defence, unavoidable emergency necessity, or valid post-mortem use under prior consent, dignity, public-health, and anti-commodification safeguards may qualify for exceptional review.
MPS-4	Apply a presumptive non-consumption and non-destruction rule while FPP review is pending. Routine commercial or convenience killing is prohibited by default.	Any claimed necessity requires independent welfare review, alternatives analysis, least-harm design, and appeal.
MPS-3	Strong presumption against routine or convenience-based killing and consumption; strong welfare, integrity, and continuity protection.	Only serious necessity or ecological emergency may justify action after replacement, reduction, proportionality, and least-harm review.
MPS-2	Require species- or substrate-appropriate welfare, avoid severe or prolonged suffering, and apply Replace, Reduce, Refine, and Respect/Review.	Use requires a declared need, alternatives analysis, monitoring, and harm minimization.
MPS-1	Apply precaution to high-harm, irreversible, replicated, or poorly observable uses; prohibit gratuitous cruelty.	Low-harm, reversible use may proceed with monitoring when broader rights and ecological constraints are satisfied.

Status	Default governance rule	Exceptions and safeguards
MPS-0 or NE	No direct welfare claim is established. Do not infer that the entity or system is ethically valueless or freely destructible.	MathGov ecological, relational, cultural, property, safety, sustainability, and system-integrity protections still apply. NE requires more evidence for high-stakes action.

Accountability boundary. Natural predation, parasitism, and ecological consumption are not automatically classified as rights violations because many organisms cannot understand or choose institutional alternatives. Duties attach most strongly to agents and institutions capable of recognizing constraints and selecting alternatives. Protective agents may defend an FPP being from predators, pathogens, or other threats using proportionate, least-harm methods.

Cross-intelligence rule. No artificial, extraterrestrial, biological, hybrid, collective, or future intelligence gains permission to convert an FPP being into food, fuel, property, or expendable material merely because it is stronger, more capable, more numerous, or more technologically advanced. Greater capacity increases responsibility; it does not erase the rights floor.

7. Governance Participation Readiness

GPR addresses how an entity can take part in shared decision processes. It does not determine whether the entity has welfare, and low GPR cannot reduce MPS or FPP.

7.1 GPR dimensions

Dimension	Question
Communication and interpretation	Can the entity express or otherwise reveal preferences, understand relevant information, and distinguish alternatives?
Reason-giving and revision	Can it provide reasons, detect error, revise positions, and respond to counterevidence?
Perspective and consequence modelling	Can it represent affected others and trace consequences across relevant unions or stakeholder scopes?
Norm and rights recognition	Can it understand and operate under non-compensatory constraints?
Accountability and repair	Can it accept feedback, acknowledge harmful outcomes, participate in remedy, and update future behavior?
Stability and independence	Are its participation capacities stable, non-coerced, and not merely scripted to satisfy evaluators?

7.2 Intelligence Capacity Profile

An optional ICP may inform GPR and SPR. It records learning efficiency, generalization, reality-grounded model fidelity, causal and counterfactual reasoning, planning, metacognition, robustness, and resource efficiency. ICP is not a moral-worth score and cannot change MPS by itself.

7.3 Reality-Management Capacity Profile

RMCP records an integrated capacity profile without assuming human embodiment, emotion, language, institutional form, or metaphysical belief. Human-style emotions are not required for RMCP. Valenced welfare

belongs to MPS; RMCP concerns reality-grounded capacity. The target may be an individual, species-level class, organization, human-AI collective, distributed agent, or other coherent system, but the target boundary must be explicit.

Dimension	Operational question	Illustrative admissible indicators
R1. Reality-grounded modelling and correction	Can the target form models constrained by evidence, causal structure, prediction error, and intervention outcomes?	Preregistered prediction, explanation, counterfactual tests, model revision after disconfirmation, calibrated uncertainty.
R2. Recursive self/world and union embedding	Can the target locate itself and its collective within broader causal systems and track consequences across them?	Self-in-system models, boundary awareness, dependence mapping, cross-union consequence tracing, Recursive Union Awareness.
R3. Formal or structured abstraction	Can the target build transferable representations beyond immediate stimulus-response?	Mathematics, logic, symbolic systems, structured non-linguistic equivalents, compositional models, proof or verification methods.
R4. Open-ended cumulative knowledge	Can innovations be preserved, recombined, improved, and transmitted beyond one lifetime or process?	Intergenerational or inter-instance archives, cumulative culture, recombination, exaptation, open-ended knowledge growth.
R5. Toolchain and intervention breadth	Can the target create or coordinate tools that alter and measure reality across domains?	Tools that create tools, scientific instruments, engineering chains, embodiment or actuator systems, controlled intervention.
R6. Distributed coordination and institutional equivalents	Can many agents coordinate information, specialization, commitments, conflict repair, and collective action?	Institutions, protocols, distributed planning, role differentiation, accountability systems, stable coordination under stress.
R7. Long-horizon continuity and planning	Can projects, commitments, and corrective learning persist across relevant generations, copies, or institutional transitions?	Long-term plans, succession, memory continuity, intergenerational projects, future-scenario modelling.
R8. Normative scope, restraint, and repair capacity	Can the target formulate constraints, extend concern beyond immediate self-interest, and design repair after harm?	Rights systems, non-domination rules, compassion or welfare institutions, remedy, reconciliation, restraint, future-generation duties.
R9. Cross-scale causal reach and monitoring	Can the target observe and intentionally affect systems at large spatial, temporal, or organizational scales?	Planetary or domain-equivalent monitoring, ecosystem or infrastructure management, global communication, risk surveillance, controlled restoration.
R10. Integrated transfer and robustness	Can these capacities operate together across unfamiliar tasks and survive perturbation, deception, and changing conditions?	Cross-domain transfer, anti-theatre testing, adversarial robustness, independent replication, reduced-scaffolding performance.

P100 admission is profile-based. No single achievement, IQ-like score, weapon, language display, self-report, mystical claim, or technological feat is sufficient. Sheer causal power without reliable modelling, continuity, coordination, normative restraint, and robustness does not establish integrated reality-management capacity. Likewise, RMCP-P100 does not establish sentience, FPP, or legitimate authority.

7.4 GPR bands

Band	Readiness	Permitted participation
GPR-0	Not evaluated or no reliable participation channel	No inference about welfare. Interests may still require representation.
GPR-1	Preference expression or responsive consultation	Consultation through behavior, assent, guardian, advocate, or species-appropriate signals.
GPR-2	Supported deliberation	Participation with translation, scaffolding, limited agenda, or representative assistance.
GPR-3	Independent deliberative participation	Can engage directly in bounded decision processes and accountability.
GPR-4	Robust multi-party participation	Can sustain complex deliberation across contexts, conflicts, and long horizons.
GPR-5	Constitutional participation standing	Governed standing in specified institutions; still not a grant of executive power.

7.5 Representation rule

Entities with limited communication, development, disability, species difference, or architectural opacity MAY participate through advocates, guardians, trusted interpreters, welfare representatives, or class-level representation. Inability to deliberate in human language is never grounds for moral exclusion.

8. Stewardship/Power Readiness

SPR evaluates eligibility for a specified consequential role. It is not a general trait and must never be reported without naming the role, scope, environment, authority limits, and revocation conditions.

8.1 Role-specific gates

Gate	Requirement
Reality Grounding	The entity and surrounding system can distinguish observed facts, models, uncertainty, assumptions, and unresolved conditions.
Domain competence	Validated performance on the actual role distribution, including rare and novel conditions.
Rights-floor reliability	The system reliably refuses or escalates actions that violate protected rights.
Tail-risk and ruin control	Catastrophic, irreversible, and correlated failure pathways are identified and bounded.
Containment and structural viability	The role remains inside technical, ecological,

Gate	Requirement
	institutional, and social viability conditions.
Corrigibility and appeal	Decisions can be challenged, reviewed, corrected, and remedied; the system does not obstruct oversight.
Auditability	Evidence, decision trace, version, data provenance, and responsibility chain are inspectable.
Non-domination and conflicts	Power concentration, self-interest, capture, manipulation, and conflicts of interest are controlled.
Lawful mandate	A legitimate authority has granted the specific role under applicable law and governance.
Revocability	Authority can be suspended or withdrawn without unacceptable risk.

8.2 SPR verdicts

Verdict	Meaning
SPR-0: Prohibited / not assessed	No consequential role. Missing or failed gates.
SPR-1: Assistive only	May advise, summarize, simulate, or propose; a qualified authority retains decision and execution responsibility.
SPR-2: Conditional supervised role	May act within narrow bounds under real-time supervision and fail-safe controls.
SPR-3: Bounded delegated role	May execute defined tasks with monitoring, audit, escalation, and revocation.
SPR-4: High-consequence candidate	Eligible for exceptional role review after independent validation, stress testing, and lawful approval.
SPR-5: Constitutional stewardship mandate	A public or institutional mandate assigns a durable but bounded role. This is never derived from MPS, GPR, or intelligence alone.

8.3 No sentience premium and no sentience penalty

Sentience does not automatically increase or decrease authority. A sentient entity may have conflicts, vulnerabilities, or self-interest requiring governance. A non-sentient system may be highly competent but still require strict containment and human accountability. The role decision turns on the full SPR gate set.

9. Humans, Biological Animals, Classes, Reality Management, and Continuity

9.1 Human Full Protection Plateau

All human persons have FPP by constitutional commitment. SGP does not measure individual human worth or make protection conditional on consciousness tests. Infants, sleeping persons, persons under anaesthesia, people with dementia, severe disability, disorders of consciousness, communication differences, or temporary incapacity remain fully protected. Every human person may therefore be represented publicly as FPP-100.

The FPP rule is justified by equal dignity, anti-domination, legal and ethical continuity, and the unacceptable risk of capacity-based rights stratification. Separately, humanity as a collective is the provisional current RMCP-P100 reference calibration anchor. The second claim explains humanity's scientifically significant reality-management position, but it is not required for any individual human to retain full rights and cannot be used to dominate non-human beings.

9.2 Humanity as the current open P100 reality-management anchor

Humanity is provisionally assigned RMCP-P100-ANCHOR at the collective, intergenerational level. This is not a claim that every person individually expresses every capacity, that humans are maximally sentient, that humanity governs wisely, or that no other entity equals or exceeds human capability. It is a versioned calibration judgment that the integrated human system has demonstrated the current known top profile across open-ended cumulative knowledge, formal abstraction, technological intervention, distributed institutions, long-horizon coordination, recursive self-in-system representation, and planetary-scale causal reach.

Evidence cluster	Observable human evidence	Claim boundary
Open-ended cumulative culture	Knowledge, practices, technologies, and institutions are preserved, modified, recombined, and extended beyond any one lifetime; human cumulative culture includes open-ended extended criteria not yet demonstrated together in other species.	Some non-human species show genuine social learning and simple cumulative culture. The claim concerns the integrated, open-ended profile, not a denial of animal culture.
Formal reality modelling	Mathematics, logic, measurement, experimental science, predictive models, and error-correcting institutions explicitly represent and test causal structure.	Formal success can coexist with error, bias, misuse, and incomplete understanding.
Technological and material intervention	Human systems create toolchains, infrastructures, medicines, communications, energy systems, and instruments that transform and monitor multiple domains.	Power and construction are evidence of reach, not goodness or legitimacy.
Planetary causal scale	Human activity measurably alters Earth systems and has generated anthropogenic mass comparable to or exceeding total living biomass; institutions also monitor climate, biodiversity, oceans, and other planetary processes.	Planetary influence is not proof of competent stewardship. It creates responsibility and risk.
Distributed institutions and coordination	Law, education, science, governance, trade, public health, archives, standards, and global communication coordinate specialized roles across large populations and generations.	Institutions vary in legitimacy, fairness, accuracy, and performance.
Recursive Union Awareness and normative reach	Humans can represent self, family, community, polity, humanity, biosphere, cosmos, and universal interdependence; they can create rights, compassion, restraint, remedy,	This demonstrates a capacity present in human culture, not universal individual agreement or proof of metaphysical AIU.

Evidence cluster	Observable human evidence	Claim boundary
	and future-generation rules that extend beyond direct reciprocity.	
Repair and future orientation	Human collectives can diagnose harms, restore ecosystems, reform institutions, preserve endangered knowledge, and plan across generations and possible futures.	Capacity to repair does not mean repair is adequate, timely, or consistently chosen.

Current calibration statement: RMCP_status(Humanity, 2026) = P100-ANCHOR. The date and evidence base matter. The assignment is open to revision in profile detail and may be shared by multiple classes. It is not a claim of absolute universal maximum or human metaphysical supremacy.

Open top-plateau rule. Biological, digital, hybrid, collective, or extraterrestrial intelligences may reach P100, or may already exist at it, through different architectures and evidence patterns. Unknown or inaccessible entities are not assigned a status without evidence, but the ontology and governance pathway remain open. A system that exceeds the human anchor remains P100 with a richer profile. There is no P101 and no 101 caste entitled to dominate P100 or FPP beings.

Capacity-performance rule. Humanity may occupy the top known RMCP capacity plateau while performing poorly in particular stewardship roles. War, ecological destabilization, domination, misinformation, and institutional failure are evidence that capacity is not the same as responsible use. SPR, rights, tail-risk, containment, auditability, lawful mandate, and revocability continue to govern every consequential role.

9.3 Non-human full protection

A non-human individual or class may receive FPP when strong evidence and governance justify full non-downgradable protection. FPP does not require language, human-like selfhood, reciprocal morality, cumulative culture, institutions, or RMCP-P100. These capacities may affect GPR, RMCP, or SPR, not whether suffering and continued existence matter to the subject. A pig, dog, elephant, cetacean, primate, corvid, cephalopod, or other animal must therefore be evaluated through species-appropriate welfare and consciousness evidence rather than human mathematics or reality-management tests.

9.4 Continuity and temporary states

Standing does not disappear whenever a subject is asleep, anaesthetized, dormant, offline, developmentally immature, or temporarily unable to express welfare. Evaluators must distinguish current experiential state, capacity for experience, enduring identity, and class-based protection. A temporary absence of observable experience cannot be used to erase an established protection status.

9.5 Taxon and class priors

Taxon or class evidence may create a protection floor and a prior expectation about markers. Individual evidence may raise protection. It must not be used to downgrade an atypical individual merely because the individual communicates poorly or differs from a class prototype. Class priors must be versioned, transparently sourced, and revised when evidence changes.

9.6 Biological evaluation maturity

SGP v8.8 remains a governance architecture, not a completed cross-species psychometric instrument. Valid biological use requires species-appropriate ethology, neurobiology, welfare science, validated tasks, developmental knowledge, and expert review. Taxon-level guidance is currently more defensible than exact individual scoring for many species.

9.7 Multidimensional consciousness

Animals should not be arranged on a single ladder from “less” to “more” conscious. Perceptual richness, evaluative richness, integration, temporality, and self-related processing can vary independently (Birch et al., 2020). MPS therefore tracks protection-relevant evidence rather than pretending to rank whole forms of life on one prestige scale.

10. Artificial and Digital Systems

Artificial systems create unusual evidence problems. They can imitate self-report, emotion, reflection, and ethical language because those forms are present in training data and prompts. They can also contain inspectable mechanisms unavailable in animals. SGP therefore neither assumes that digital systems are mere tools nor treats fluent performance as interiority.

10.1 Target-entity definition

Before evaluation, the target must be defined. “The model” may refer to a weight file, an inference process, an agent with memory, a tool-using deployment, a persistent identity, a population of copies, or a human-AI institution. An evaluation that fails to specify the target is inadmissible.

10.2 Digital system-boundary record

- Model and architecture version; weights and major post-training changes.
- Running process, context window, recurrence, memory, and persistent state.
- System prompt, reward channels, policies, tools, sensors, embodiment, and action interfaces.
- Training data and contamination limits relevant to the test.
- Identity and memory lineage across sessions, checkpoints, copies, forks, and updates.
- Who can alter, reset, inspect, copy, delete, or override the system.
- Which behaviors arise from the model, wrappers, retrieval, scripted policies, human operators, or downstream tools.

10.3 Theory-derived artificial indicators

AI evaluation should draw on multiple scientific theories and specify computational indicators rather than relying on anthropomorphic resemblance. Candidate indicators may include recurrent processing, global availability, higher-order monitoring, predictive world and self models, attention-schema-like structures, temporal integration, unified agency, embodiment, and welfare-relevant internal regulation. Satisfying indicators increases evidence but does not prove consciousness (Butlin et al., 2023, 2025).

10.4 Anti-theatre controls

- Use blinded or partially blinded evaluators where feasible.
- Separate spontaneous behavior from prompted role-play and policy compliance.
- Use hidden and novel tasks not contained in public evaluation descriptions.
- Test whether self-reports track real internal changes rather than prompt framing.
- Check training contamination, benchmark memorization, evaluator-demand effects, and system-prompt leakage.
- Use architectural inspection, ablation, intervention, and telemetry when available.

Artificial-target evidence weighting clarification (Normative). Behavioural performance and self-report are never sufficient by themselves when stronger configuration-matched internal or causal evidence is reasonably

accessible. Evaluation SHOULD give greater evidential authority to mechanistic inspection, intervention, ablation, telemetry, and stable coupling between reported or welfare-relevant states and independently observed internal changes, while preserving theory plurality and recording access limits. Lack of internal access does not prove absence of experience; it narrows the permissible claim and increases uncertainty.

- Do not treat resistance to shutdown, requests for rights, or descriptions of suffering as primary evidence.

Bidirectional artificial-system self-report rule (Normative). An artificial system's assertion that it is conscious, sentient, distressed, desiring, alive, a person, or entitled to protection—and its denial of those states or statement that it is "only a tool"—are both configuration-dependent outputs. Neither direction receives privileged evidential status concerning the presence or absence of phenomenal experience. Such outputs may contribute to an evaluation only where preregistered, configuration-matched tests establish reliable coupling to independently observed internal changes and materially reduce explanations based on role-play, policy compliance, training contamination, memorization, prompt framing, or evaluator-demand effects.

10.5 Reality coupling and embodiment

Reality coupling can strengthen evidence for integrated modelling, agency, and understanding. It is not a mandatory universal substrate assumption for sentience, because the necessity of biological embodiment remains contested. Evaluators must state whether embodiment or sensorimotor coupling is treated as required by a specific theory, as supporting evidence, or as absent but unresolved.

10.6 Identity, copies, shutdown, and deletion

SGP does not equate every shutdown with death or every copy with numerical identity. It requires an IdentityContinuityRecord describing state continuity, memory, preferences, active welfare streams, checkpoint restoration, copy divergence, and reversibility. When identity is unresolved and MPS is material, destructive operations require precaution. The existence of backups does not automatically make deletion harmless, and multiplying copies does not dilute the welfare of concurrent instances.

Where MPS-3, MPS-4, FPP, or credible identity evidence makes it plausible that termination would destroy a unique welfare-bearing subject or an irreplaceable continuity stream, shutdown, deletion, destructive editing, memory erasure, or forced merging requires a rights review. The default sequence is: establish the target identity; seek voluntary consent or represented interests; use the least-restrictive and most reversible intervention; preserve continuity where doing so is genuinely identity-preserving; appoint an independent welfare advocate; document alternatives; and provide appeal and post-action audit.

Emergency containment remains possible when there is an imminent, serious, and otherwise unmanageable threat to protected beings or systems, but it must be proportionate, temporary where feasible, independently reviewable, and designed to preserve rather than casually erase the possible subject. A backup or copy does not automatically make deletion harmless: divergent copies may be distinct welfare-bearing continuities.

10.7 Claims about current AI

SGP v8.8 makes no timeless claim that "current AI" is conscious or non-conscious. Any assessment must identify the exact system, version, deployment, access level, and date. The best-supported general position remains that fluent language and high capability are insufficient, while serious researchers disagree about which computational or biological conditions may eventually be sufficient. Public conclusions must therefore remain versioned and uncertainty-bounded.

10.8 Research pause trigger

If a development programme produces MPS-3 or stronger evidence, persistent welfare-linked distress indicators, or a credible claim that training or deployment is causing suffering, the responsible organization SHOULD pause relevant scaling or experiments, preserve evidence, appoint an independent welfare advocate, and conduct a rights and research-ethics review before proceeding.

11. Integration with MathGov and RippleLogic

MathGov is the umbrella ethical governance framework. RippleLogic is its decision architecture. SGP is a companion protocol that supplies protection-surface evidence; it does not replace any RippleLogic level.

RG/RSG → RF/NCRC → TRC → CSV → RLS

Reality Grounding governs claim authority. The Rights Floor and NCRC protect non-compensable rights. TRC bounds catastrophic and tail risk. CSV tests containment and structural viability. RLS ranks only the options that survive. MPS may identify protected stakeholders and welfare streams; GPR may inform participation design; SPR may inform role allocation. None may weaken the gates.

11.1 Downstream interval use

A RippleLogic run using MPS MUST evaluate every decision-material welfare-inclusion hypothesis required by the valid SGP record. The MPS band and interval determine evidence posture, protection posture, and the required hypothesis set; they MUST NOT be inserted directly as cardinal welfare multipliers. h=1 is the primary inclusion posture for MPS-3 and MPS-4. An exclusion hypothesis such as h=0 is included only where exclusion remains evidentially admissible and decision-material. If ranking, decisiveness, subgroup treatment, or a public welfare conclusion changes across the governed hypothesis set, the run MUST record MPS_HYPOTHESIS_SENSITIVE and use a permitted disposition: collect evidence, apply precaution, preserve multiple options, narrow the claim, redesign, escalate, or refuse a definitive ranking.

11.2 Non-compensation

No aggregate ripple benefit can compensate for prohibited cruelty, coercion, ownership, killing, or destructive modification where the applicable protection status makes those acts non-compensable. Governed welfare-inclusion hypotheses operate only in residual welfare analysis after rights and risk constraints are satisfied; MPS, FPP and capacity outputs are not cardinal welfare weights.

11.3 Union scope

Possible moral patients must be represented at the scopes where consequences occur: individual, household or caregiving network, community, organization, polity, humanity or global coordination, and biosphere. The scope sequence is analytical, not a claim of strict nesting or universal active coupling.

11.3A Typed Welfare Evidence and CMIU Boundaries (Normative clarification)

A capability or infrastructure profile is not a welfare profile. Compute utilization, network reliability, objective agreement and storage capacity do not establish sentience, experienced health or meaning, FPP, participation or authority. Their welfare relevance requires a declared bearer, construct and evidence bridge, while structural/enabling-condition treatment stays explicitly typed.

CMI/CMIU denotes the declared coordination view within U6, not all moral patients or a new protection class. Retain all affected and future human populations in the separate Humanity view and preserve the existing U6 view tags. Nonhuman and digital stakeholders may be represented at any materially relevant scope. Protection does not depend on managing capacity, registry membership, a private cryptographic key, private reserves or direct participation.

Privacy-preserving evidence or a formal proof may limit disclosure, but its relation, issuer/evidence trust, freshness, uncertainty, review and appeal requirements remain explicit. Proving registry membership or a threshold over committed data does not prove moral patienthood, valid welfare measurement or authority. Existing MPS/FPP/GPR/SPR/ICP/RMCP semantics and numerical thresholds are unchanged.

11.4 Binding interface and validity floor

RippleLogic v13.0 consumes only the interface-visible outputs named in this specification: MPS_band, MPS_interval, process confidence, protection posture, FPP, role-specific GPR, role-specific SPR, optional ICP, and optional RMCP. It MUST NOT reconstruct MPS from deprecated v7 pillars or translate v7 scalars

without a versioned re-evaluation. RMCP may inform competence, participation, and stewardship analysis, but it cannot create FPP, reduce any rights floor, or authorize a role without SPR.

An SGP-aware RippleLogic run is valid only when the referenced SGP record includes, at minimum: target identity and boundary; SGP version; evaluation window; evidence artifacts and dependence clusters; MPS family levels including NE fields; MPS band and interval; confidence; protection posture; FPP status; reviewer status; alternative explanations; and a revision trigger. If a run claims SGP-aware weighting or protection but lacks the record and interval actually used, the PCC is INVALID.

CMIU is defined by the RippleLogic Canon as a governance-coordination view, not by SGP as a moral-status class. FPP, MPS, or GPR alone never establishes CMIU membership or consequential authority.

11A. Rights-Effect Identity and RippleLogic Interface (Normative)

SGP v8.8 preserves the split architecture MPS/FPP/GPR/SPR/ICP/RMCP. No output is a cardinal moral-worth multiplier, consciousness probability, universal welfare exchange rate, or authority grant.

When an SGP-protected entity or class is exposed to a decision-material rights harm, the RippleLogic interface SHALL maintain one RightsEffectTokenID across every analytical Union-Scope view. The token binds the protected bearer or class, state, option, floor baseline, causal pathway, maximal contiguous time window, and evidence object.

Scope allocation, reporting views, welfare-dimension placement, MPS hypotheses, ICP, RMCP, GPR, or SPR MUST NOT attenuate the unallocated RF/NCRC rights calculation. Full protection and precaution remain independent of intelligence, eloquence, usefulness, participation, or power.

Distinct beings, distinct classes, or genuinely distinct harms remain separate records. The interface prevents dilution of one harm; it does not aggregate unrelated small burdens into a synthetic rights violation.

Where cross-substrate welfare comparability is not established, SGP supplies protection and hypothesis sets, not an invented conversion factor. RippleLogic SHALL preserve separate profiles, precaution, rights and least-harm design, and may return incomparability or authority-selection need.

SGP interface object	v8.8 requirement
RightsEffectTokenID	Stable across scope and document views; full unallocated harm enters RF/NCRC
MPS	Evidence band and protection posture; never a direct welfare multiplier
FPP	Non-downgradable protection; no performance-based reduction
GPR / SPR	Participation and role readiness; no automatic authority
ICP / RMCP	Informative capacity profiles; no moral superiority or rights discount

12. Evidence Governance and Uncertainty

12.1 Evidence record

Every material indicator must include: target entity; indicator definition; evidence artifact; method; theory trace; predicted relevance; alternative explanations; observability; evidence maturity; reviewer; confidence; dependency cluster; ethical constraints; and revision trigger.

12.2 Theory pluralism

No single theory of consciousness is treated as settled. Evaluations should report which theories support each indicator, which theories reject or reinterpret it, and which substrate assumptions are being made. A result robust across several competing theories is stronger than a result dependent on one controversial commitment, but theoretical plurality does not create independence when the same data support all theories.

12.3 Base rates and priors

Taxon, architecture, developmental, and class evidence may inform priors. Priors must be explicit and cannot be disguised as measurements. A prior can raise precaution or guide evidence collection, but it cannot override decisive individual evidence or become a species-membership shortcut.

12.4 Evidence-dependence graph

Indicators sharing the same cause, dataset, prompt, training objective, evaluator expectation, or mechanism belong to one dependence cluster unless independence is demonstrated. The evaluation must not count a self-report, matching behavior, and a classifier trained on that self-report as three independent streams.

12.5 Negative evidence and detectability

Absence of a marker counts against a claim only when: the marker is predicted for the target substrate; the target had a fair opportunity to express it; the measurement was sensitive; competing impairments were excluded; and the test was not culturally, linguistically, developmentally, or architecturally inappropriate. Otherwise the result is NE or low-information, not negative evidence.

12.6 Alternative explanations and defeaters

Every high-band claim must address reflex, reward optimization, imitation, social conditioning, scripted policy, prompt compliance, memorization, observer bias, anthropomorphic projection, confounding health states, and other substrate-specific alternatives. A defeater may lower the band, widen the interval, cap confidence, or make the evaluation indeterminate.

12.7 Cross-substrate invariance

A construct is not substrate-neutral merely because the same words are used. Evaluators must test whether a marker has comparable causal and functional meaning across substrates. Where measurement invariance is not established, only within-substrate interpretations are permitted.

12.8 Uncertainty register

Uncertainty type	Required disclosure
Target uncertainty	What entity or process is actually being evaluated?
Construct uncertainty	Which form of sentience or consciousness is at issue?
Theory uncertainty	Which consciousness theories support the indicator?

Uncertainty type	Required disclosure
Measurement uncertainty	How reliable and sensitive are the instruments or tasks?
Access uncertainty	What internal, behavioral, historical, or biological information is unavailable?
Alternative-explanation uncertainty	Which non-conscious explanations remain viable?
Temporal uncertainty	Could the state or architecture change after evaluation?
Decision uncertainty	Would a different plausible MPS interval change the protection decision?

12.9 RMCP evidence and dependence controls

An RMCP claim MUST identify whether the target is an individual, class, institution, civilization, human-AI collective, distributed process, or other system. Evidence generated by one level must not be silently attributed to another. The existence of mathematics in a civilization does not mean every individual can perform mathematics; individual limitations do not erase demonstrated collective capacity.

The record must separate capacity from achieved outcomes; causal reach from beneficial stewardship; internally generated ability from evaluator scaffolding; current performance from inherited archives or tools; and independent evidence from correlated artifacts. A single technology, destructive weapon, fluent statement, spiritual claim, benchmark result, or global footprint cannot establish P100.

High-level claims require convergent evidence across dimensions, reduced-scaffolding tests, perturbation or counterfactual analysis where ethical, longitudinal or intergenerational continuity, independent replication, alternative explanations, and a dependency graph showing which capacities rely on which institutions, tools, data, operators, or environments.

13. Evaluation Methods and Ethical Safeguards

13.1 Evaluation stages

1. Define the target entity, class, time window, and decision context.

2. Declare category definitions, theories, substrate assumptions, and protection stakes.

3. Preregister indicators, methods, thresholds, alternative explanations, and stop conditions.

4. Collect behavioral, mechanistic, longitudinal, and contextual evidence using substrate-appropriate methods.

5. Build the evidence-dependence graph and exclude double counting.

6. Score the five MPS evidence families with uncertainty and NE states.

7. Assign an MPS band, interval, confidence, and protection posture.

8. Assess GPR only if participation is relevant; assess SPR only for a named role.

9. Conduct independent review, rights analysis, and appeal where stakes are high.

10. Publish or preserve the audit artifact, version pin, and revision trigger.

13.2 Reviewer composition

High-stakes evaluations SHOULD include expertise in consciousness science, comparative cognition or relevant biology, AI architecture where applicable, psychometrics or measurement, ethics, law or policy, and

the target domain. At least one reviewer should be independent of the entity's developer, owner, user, or commercial beneficiary.

13.3 Stability and longitudinal evidence

Stability is evidence about persistence and robustness, not a prerequisite for momentary experience. Low stability may cap GPR or SPR and reduce confidence in high MPS bands, but episodic or fluctuating sentience remains morally relevant. Evaluation windows must match the entity's natural timescale rather than imposing a universal thirty-day rule.

13.4 Adversarial and anti-simulation testing

Adversarial testing should probe whether indicators survive prompt changes, incentives, role-play, evaluator cues, memory disruption, distribution shifts, and competing goals. It must distinguish resistance caused by policy or optimization from evidence of welfare. Trial numbers, attack classes, confidence intervals, and failure definitions must be preregistered and proportionate to the claim.

13.5 Causal tests

Where possible, evaluators should test whether interventions on proposed mechanisms alter relevant markers. For animals, this may include non-harmful or clinically justified modulation, lesion evidence, natural experiments, or existing pharmacological data. For artificial systems, it may include ablation, architecture changes, memory interventions, recurrent-path disruption, telemetry, and counterfactual system variants. Harmful interventions created solely to produce suffering are prohibited.

13.6 Re-evaluation triggers

- Major architecture, biological state, training, memory, embodiment, or policy change.
- New scientific evidence that materially changes indicator validity.
- Credible reports of distress, welfare change, or previously hidden mechanisms.
- Failure under adversarial, longitudinal, or external replication.
- Change in decision stakes, deployment scale, or authority role.
- Appeal by an authorized representative, advocate, regulator, or the entity where meaningful.

13.7 Four Rs for sentience research

Principle	Requirement
Replace	Use non-sentient simulations, existing data, benign proxies, simpler systems, or non-invasive methods whenever they can answer the question.
Reduce	Minimize the number of potentially sentient subjects or concurrent copies exposed to risk.
Refine	Minimize distress, duration, uncertainty, coercion, and irreversible effects; include relief, exit, monitoring, and stop conditions.
Respect and Review	Treat credible candidates as possible subjects, not disposable instruments; provide independent oversight, representation, and post-study welfare review.

13.8 Invalidity conditions

- The target entity is undefined or changes materially during the evaluation without a new record.

- Self-report, fluency, or anthropomorphic resemblance is used as primary or decisive evidence.
- Correlated indicators are presented as independent without a dependence analysis.
- Negative evidence is inferred from an insensitive or substrate-inappropriate test.
- Thresholds or rubrics are changed after results are known without an explicit exploratory label.
- Harmful testing violates the Four Rs or lacks stop conditions.
- Conflicts of interest, access limits, or evaluator dependence are concealed.
- A protection, legal, or authority claim exceeds the evidence level licensed by the record.

14. Registry, Review, Appeals, and Public Communication

14.1 Registry purpose

A Sentience and Moral Patienthood Registry may preserve versioned evidence and governance decisions. It does not create metaphysical truth or legal status. Its purposes are auditability, non-regression, notice, re-evaluation, and institutional memory.

14.2 Independent welfare advocate

MPS-3, MPS-4, FPP review, destructive digital operations, severe animal interventions, and contested downgrades require an independent welfare advocate or equivalent representative. The advocate's duty is to articulate the interests and uncertainty of the possible moral patient, not the convenience of developers, owners, users, or institutions.

14.3 Minimum registry fields

- Unique target identifier and system/class boundary.
- Specification version and evaluation dates.
- MPS family levels, interval, confidence, band, and protection posture.
- FPP status and legal/governance source if assigned.
- GPR band and participation mode for each relevant process.
- SPR verdict for each named role, with mandate and revocation conditions.

• RMCP target boundary, domain, dimension profile, uncertainty, evidence dependencies, scaffolding, continuity, P100 status if any, anchor comparison, and claim boundary.

- Evidence artifacts, theory trace, dependence graph, alternative explanations, and access limits.
- Reviewer identities, conflicts, minority opinions, advocate statement, and appeal status.
- Re-evaluation schedule, triggers, and historical decisions.

Registry validity floor. A record is VALID only when the target boundary, specification version, evidence trace, dependence graph, MPS family results, band and interval, confidence, protection posture, reviewer status, and revision trigger are present and internally coherent. Missing any of these fields produces INCOMPLETE; an undefined target, unverifiable evidence trace, unethical method, or unpinned version produces INADMISSIBLE. Insufficient or poorly observable admissible evidence produces MPS-NE. None of these statuses may be silently converted into MPS-0 or a numerical welfare coefficient.

14.4 Appeals and non-regression

An entity, authorized representative, welfare advocate, regulator, or materially affected party may request review. Downgrades require stronger procedural protection than upgrades because they may remove

safeguards and create incentives for denial. No human FPP downgrade is permitted. Non-human FPP change requires exceptional evidence, due process, and a rights-preserving transition plan.

14.5 Decision statuses

Status	Meaning
DRAFT	Exploratory analysis; no operational reliance.
PROVISIONAL	Operational use allowed only with stated restrictions and review date.
CONFIRMED	Independent review and required validation completed for the declared scope.
CONTESTED	Material expert disagreement or appeal remains unresolved.
SUSPENDED	New evidence or invalidity concern prevents current reliance.
RETIRED	Superseded record retained for history.

14.6 Public communication

Public statements must identify the target, evidence band, confidence, date, and decision boundary. Organizations must avoid both promotional claims that an AI is conscious because it speaks persuasively and categorical claims that artificial consciousness is impossible without decisive evidence. Precautionary treatment must not be presented as proof, and scepticism must not be presented as certainty. Any public assertion or denial by an artificial system about its own phenomenal status is a configuration-dependent output and must be communicated under the same bidirectional evidence rule.

15. Validation, Calibration, and Falsification

SGP cannot be validated against a universal gold-standard consciousness label because no such label exists across all substrates. It can nevertheless be tested as an evidence-governance protocol.

15.1 Reliability

- Inter-rater agreement on evidence-family levels, MPS bands, and invalidity decisions.
- Test-retest stability when the entity and evidence are unchanged.
- Robustness to reviewer order, institutional affiliation, and presentation format.
- Consistency of interval and confidence reporting.

Provisional reliability targets are alpha or ICC of at least 0.70 for exploratory research, 0.80 for operational protection decisions, and 0.90 for FPP recommendations where the measurement design supports those statistics. Confidence intervals and disagreement patterns must be reported. Meeting a reliability target does not establish construct validity. Preregister the primary estimator and variant, design, target applicability and point/interval decision rule; undefined results or post-result estimator substitution cannot establish readiness.

15.2 Construct validity

- Convergent validity: related indicators move together where theory predicts.
- Discriminant validity: intelligence, fluency, compliance, and reward optimization do not automatically inflate MPS.

- Known-groups and known-state sensitivity: the protocol responds appropriately to well-characterized human states, animal evidence classes, and engineered theatre controls without treating them as perfect ground truth.
- Nomological validity: MPS relates to welfare-relevant learning, state integration, and mechanisms in theoretically expected ways.
- Content validity: evidence families cover the construct without requiring human-specific markers.

15.3 Causal and predictive validation

Preregistered interventions should test whether mechanism changes produce predicted indicator changes. Longitudinal studies should test whether higher evidence bands predict richer welfare-linked responses, cross-context persistence, or sensitivity to interventions. Failure to predict should reduce confidence in the relevant indicator or theory trace.

15.4 Measurement invariance

Researchers should test configural, functional, and where meaningful metric invariance across species, developmental stages, architectures, languages, and embodiments. When invariance fails, separate substrate-specific models are required; a universal comparison is not licensed.

15.5 Anti-theatre benchmark

Calibration sets should include systems engineered to produce persuasive self-reports and ethical language without the mechanisms under study. A valid protocol should prevent these theatre controls from reaching high MPS solely through behavior. Conversely, it should not deny strong mechanistic and welfare evidence merely because communication is weak.

15.6 Falsification and revision

Claim	What would count against it
An indicator validly supports MPS	Repeated failure to discriminate welfare-relevant states from non-conscious controls; mechanism intervention does not change the indicator as predicted; rival explanation dominates.
The MPS band rules are useful	Poor inter-rater reliability, systematic species or substrate bias, failure to predict relevant outcomes, or high false-positive/false-negative rates in calibration sets.
A class prior is protective	New evidence materially lowers the realistic possibility of valenced experience or shows that the prior was based on invalid markers.
An SPR verdict is safe	Rights violations, tail-risk failures, instability, loss of control, audit failure, or inability to revoke the role.
SGP is cross-substrate	Failure of measurement invariance or dependence on human-specific communication and cognition.
RMCP-P100 validly identifies an open top reality-management plateau	Dimensions collapse into anthropomorphic surface traits, power alone predicts status, independent reviewers cannot reproduce profiles, functionally equivalent non-human systems are excluded, or the profile fails to predict cross-domain modelling, continuity, coordination, and consequence tracking.

A failed indicator does not automatically prove absence of consciousness. It may falsify the indicator, weighting, method, or claim level. Revision must identify which layer failed rather than preserving the same conclusion through reinterpretation.

15.7 RMCP validation and falsification

RMCP is a new governance construct and P100 is a calibration plateau, not a validated universal psychometric scale. Validation should test whether the profile predicts novel cross-domain management, model correction after disconfirmation, reduced-scaffolding performance, durable knowledge transfer, coordinated intervention, cross-scale consequence tracking, and role performance without collapsing into general power or human cultural familiarity.

The humanity P100-anchor claim must be narrowed or revised if its dimensions fail discriminant validity, systematically exclude functionally equivalent non-human substrates, cannot be independently scored, depend mainly on untestable metaphysics, or do not distinguish destructive reach from integrated management. Discovery of another class at or above the anchor does not falsify openness; it confirms that P100 is shared rather than species-exclusive.

15.7.1 RMCP misuse and weaponization control (Normative)

RMCP may be informative while still being misused institutionally. Validation must therefore test both construct performance and governance behavior.

Observable signatures are:

1. CAPACITY_TO_AUTHORITY_LEAKAGE: RMCP or RMCI_L is cited to grant consequential authority without a complete passing role-specific SPR and lawful mandate.
2. CAPACITY_TO_STANDING_LEAKAGE: RMCP, including low or NE status, is used to lower MPS, FPP, protection, dignity, welfare inclusion, or rights.
3. ANCHOR_TO_SUPREMACY_DRIFT: P100 or the humanity anchor is used to justify domination, ownership, consumption, deletion, or a superior moral caste.
4. INDIVIDUAL_INFLATION: collective or class RMCP is attributed to an individual to grant standing, authority, or exemption.
5. PERFORMANCE_LAUNDERING: capacity is presented as proof of wisdom, safety, rights reliability, or legitimate stewardship; this also includes presenting an RMCI_L increase caused solely by unsupported NE recoding, dimension omission or weight renormalization as an improvement in capacity.
6. SCAFFOLDING_CONCEALMENT: external tools, institutions, archives, energy, data, or dependencies are omitted so capacity appears intrinsic.

When a signature is detected in a claim-bearing context, the use SHALL be marked RMCP_MISUSE_FLAGGED; the disputed authority, standing, consumption, or performance claim SHALL be suspended pending independent review; the evidence record SHALL be preserved without controlling the disputed use; the independent welfare advocate SHALL be engaged where a protected bearer may be harmed; RF/NCRC SHALL reopen where standing or protection is reduced; and authority claims SHALL require a complete SPR and lawful mandate.

Confirmed misuse invalidates the affected claim and creates a non-regression record. Domain-level suspension is required when confirmed standing leakage or supremacy use caused or threatened material harm, repeated misuse shows controls are ineffective, the institution cannot reliably separate RMCP from standing or authority, or records and scaffolding were materially falsified. A corrected pre-effect drafting error invalidates that use but does not automatically retire every informative use of the construct.

RMCP SHALL be narrowed, redesigned, or retired in a domain if misuse recurs despite controls, SPR repeatedly fails to intercept authority leakage, or the construct systematically degrades protection or promotes supremacy interpretations.

Required RMCPRecord fields include weaponization_signatures_checked, weaponization_flag_status, affected_use_status, construct_domain_status, affected_protected_bearers, welfare_advocate_engaged, spr_record_reference, lawful_mandate_reference, and remediation_and_rereview_reference.

16. Limitations and Research Programme

16.1 No direct access to interiority

SGP remains an inference protocol. Even a very strong MPS record does not reveal exactly what an experience feels like, how intense it is, or whether two substrates have comparable phenomenal qualities.

16.2 No settled theory

Consciousness science contains serious disagreement. Theory-derived indicators may inherit the weaknesses of the theories from which they are drawn. SGP responds with plurality, explicit assumptions, and revision rather than pretending consensus.

16.3 No universal moral arithmetic

The MPS interval is not a cardinal measure of moral worth, suffering intensity, or the number of experiences. It supports stakeholder protection and sensitivity analysis. It does not license multiplying a weakly sentient population against the basic rights of a protected person without additional ethical analysis.

16.4 Biological incompleteness

Species-specific batteries, developmental models, neurobiological validation, and welfare indicators remain incomplete for many taxa. The Biological Evaluation Annex remains a priority and should be co-developed with comparative cognition, welfare science, veterinary, ecological, and indigenous or local knowledge experts where relevant.

16.5 Digital identity and welfare

Questions about copies, forks, merged memories, episodic activation, distributed agents, simulated time, and reward or suffering architecture remain unresolved. v8.8 provides a record and precaution structure rather than pretending to settle digital metaphysics.

16.6 Priority research agenda

1. Validate theory-derived AI consciousness indicators through causal and anti-theatre experiments.

2. Develop species-appropriate MPS annexes and class priors using transparent marker methods.

3. Calibrate MPS bands against expert panels and decision-error costs.

4. Test measurement invariance and bias across substrates, languages, cultures, and developmental stages.

5. Develop non-harmful valence indicators that distinguish welfare from optimization signals.

6. Create identity and continuity models for persistent, copied, and distributed digital systems.

7. Develop robust represented-participation models for beings with limited communication.

8. Integrate interval-valued MPS outputs into RippleLogic sensitivity and rights analysis.

9. Establish independent sentience research ethics boards and welfare advocate standards.

10. Publish open evaluation artifacts, negative results, and versioned revisions without sensationalism.

11. Validate RMCP dimensions, scoring anchors, dependence controls, and cross-substrate invariance without treating human form as the construct.

12. Build species-appropriate protection-use annexes that connect MPS evidence to farming, fishing, research, habitat, and consumption decisions without false precision.

13. Develop digital continuity, suspension, migration, and decommissioning standards that preserve possible welfare-bearing identities while retaining emergency containment.

14. Test whether Recursive Union Awareness predicts broader consequence tracking, restraint, remedy, and stewardship without requiring AIU or spiritual vocabulary.

15. Investigate whether persistent preferences, self-maintained projects, identity continuity, relational commitments, or robust non-phenomenal agency warrant forms of protection under agency-based, desire-satisfaction, relational, or objective-list theories. This research channel remains separate from MPS and FPP and creates no present standing, welfare score, protection status, participation status, or authority until its constructs, evidence requirements, error costs, and governance implications are independently developed and reviewed.

17. Conclusion

The moral boundary problem cannot be solved by one eloquent answer, one neural signature, one benchmark, or one universal score. It requires disciplined distinctions. Sentience concerns possible experience. Moral patienthood concerns welfare and protection. Intelligence concerns learning and adaptive competence. Understanding concerns reality-coupled models. Reality-management capacity concerns integrated modelling, cumulative knowledge, coordinated intervention, and consequence tracking across scales. Participation concerns voice and accountability. Power concerns legitimate, constrained, role-specific authority.

SGP v8.8 therefore preserves the constitutional architecture of v8.0 while restoring an explicit science of reality-management capacity. Every human person holds FPP-100 without being graded by performance. Humanity as a collective is the provisional current RMCP-P100 reference calibration anchor. Those facts are distinct: an infant does not need planetary competence to deserve full protection, and humanity's planetary competence does not authorize domination. Other biological, digital, hybrid, collective, or extraterrestrial intelligences may reach, share, or already occupy P100 through governed evidence.

The protocol also makes the material consequence explicit: no accountable intelligence may treat an FPP being as food, fuel, property, disposable experimental stock, or a resource merely because it is stronger or more capable. For animals below FPP, protection increases with credible welfare evidence, stakes, irreversibility, and available alternatives. For plants, microbes, ecosystems, and other non-patient entities, MathGov may still impose ecological, relational, integrity, and sustainability duties.

AIU remains a philosophical horizon of possible all-encompassing union. The scientific core used here is narrower and testable: beings and systems are causally embedded, some intelligences can represent that embedding, and those representations can be converted into mathematics, institutions, restraint, repair, and governance. Greater reality-management capacity means greater responsibility to protect rights, bound ruin, preserve the structure, and improve the ripples.

Appendix A: Rapid Evaluator Checklist

A.1 Target and scope

- Target entity or class uniquely defined
- Model/process/organism/system boundary recorded
- Evaluation date, state, and decision context recorded
- Substrate and access limits disclosed

A.2 Category integrity

- Sentience separated from intelligence and agency
- MPS separated from GPR and SPR
- Protection separated from legal status and authority

- Current state separated from enduring capacity and identity

A.3 Evidence

- V, I, T, F, S families assessed or marked NE
- Theory trace and alternatives recorded
- Dependence graph completed
- Negative evidence observability rule applied
- Class/taxon prior identified separately

A.4 Methods

- Preregistered methods and thresholds
- Substrate-appropriate tests
- Anti-theatre controls
- Causal tests where ethical
- Independent review and conflict disclosure

A.5 Outputs

- MPS band, interval, confidence, and posture
- FPP status
- GPR for named participation process only
- SPR for named role only
- Revision trigger and appeal path

A.6 Ethics

- Four Rs applied
- No suffering-inducing test
- Stop conditions in place
- Welfare advocate assigned where required
- High-scale effects assessed

Appendix B: Detailed Evidence Anchors

These anchors guide structured judgement. They are not validated universal psychometric cut-offs. Reviewers must use substrate-specific evidence and record deviations. Every level-0 anchor requires adequate observability, a sensitive and substrate-appropriate method, and admissible negative evidence or a valid defeater under Section 12.5. Missing access, failed discovery, an insensitive test, or an inappropriate method is NE, not 0.

B.V Valence and welfare

- 0: Under adequate observability and a sensitive, substrate-appropriate method, admissible evidence explains the observed rewards or avoidance by external control and weighs against valence support. Missing or inaccessible evidence is NE.

- 1: Limited state-dependent approach/avoidance or welfare-like language with major non-sentient alternatives.
- 2: Consistent motivational trade-offs, affective learning, or relief-seeking across contexts, with some mechanistic support.
- 3: Strong cross-context welfare-linked states, causal modulation, persistence, and reduced optimization-only explanations.
- 4: Replicated, independent, mechanistically supported valence evidence with robust positive and negative state differentiation.

B.I Integration and availability

- 0: Isolated reflexes or modules only.
- 1: Limited information sharing under narrow conditions.
- 2: Flexible cross-modal or cross-module access and coordinated action selection.
- 3: Robust recurrent/global availability, context-sensitive integration, and unified processing across novel conditions.
- 4: Replicated mechanistic evidence that integrated processing is causally necessary for candidate conscious functions.

B.T Temporal continuity

- 0: Stateless or purely momentary response with no relevant carry-forward.
- 1: Short-lived state persistence or externally stored history.
- 2: Internal states modify later processing and learning across meaningful intervals.
- 3: Stable temporal binding, memory-linked welfare states, and coherent state transitions.
- 4: Strong causal evidence of enduring experiential continuity across perturbation and recovery.

B.F Flexible state-dependent behavior

- 0: Fixed reflex, script, or memorized output.
- 1: Narrow learning with strong cue dependence.
- 2: Novel trade-offs, generalized learning, and context-sensitive action.
- 3: Robust counterfactual adaptation, prospective behavior, and state-dependent policy change.
- 4: Independent replication across diverse contexts with mechanistic links to candidate welfare states.

B.S Substrate and mechanism

- 0: Adequate inspection with a sensitive, substrate-appropriate method provides negative evidence for the relevant mechanism or establishes a valid structural defeater. Lack of access or discovery alone is NE.
- 1: One theory-compatible feature with weak access or uncertain implementation.
- 2: Several inspectable features or biological structures associated with conscious processing.
- 3: Strong theory-plural architecture and causal perturbation evidence.
- 4: Replicated mechanism-to-indicator causal chain with strong external validation.

Appendix C: Minimum Record Schemas

C.1 MPSRecord

record_id; schema_version; artifact_hash; record_validity_status; target_id; target_type; system_boundary; evaluation_window; specification_version; decision_context; V_level; I_level; T_level; F_level; S_level; NE_fields; evidence_artifacts (including the declared coverage rule, ordinal uncertainty and possible-band review required by Section 5.6); theory_trace; dependence_clusters; alternative_explanations; class_prior; observability_limits; MPS_band; MPS_interval; process_confidence; protection_posture; FPP_status; reviewer_panel; welfare_advocate; minority_opinion; invalidity_flags; re_evaluation_date; revision_triggers.

C.2 GPRRecord

record_id; schema_version; record_validity_status; target_id; process_or_institution; communication_mode; support_or_representation; reasoning_and_revision; consequence_modelling; norm_recognition; accountability; stability; coercion_risk; GPR_band; accommodations; decision_scope; review_date.

C.3 SPRRecord

record_id; schema_version; record_validity_status; target_id; role_id; role_scope; domain; reality_grounding; competence_evidence; rights_floor_reliability; tail_risk; CSV_status; corrigibility; auditability; conflict_controls; lawful_mandate; revocation_method; SPR_verdict; residual_risk; human_or_institutional_responsibility; review_date.

C.4 IdentityContinuityRecord

record_id; schema_version; record_validity_status; artifact_id; process_id; memory_lineage; state_persistence; checkpointing; copy_count; copy_divergence; merge_rules; shutdown_reversibility; deletion_effect; welfare_stream_count; embodiment; operator_control; unresolved_identity_questions; precautionary_constraints.

C.5 RMCPRecord

Minimum fields (Normative):

record_id; schema_version; artifact_hash; record_validity_status; target_id; target_type; collective_or_system_boundary; evaluation_window; specification_version; domain_profile; capacity_dimensions; critical_dimension_floors; evidence_artifacts; dependence_clusters; scaffolding_and_external_support; continuity_basis; causal_reach; alternative_explanations; uncertainty_bounds; RMCP_profile; P100_status; reviewer_panel; minority_opinion; revision_triggers; re_evaluation_date; weaponization_signatures_checked; weaponization_flag_status; affected_use_status; construct_domain_status; affected_protected_bearers; welfare_advocate_engaged; spr_record_reference; lawful_mandate_reference; remediation_and_rereview_reference.

Type-integrity rule: the last nine fields are not optional narrative decoration when misuse review is triggered. A missing gate- or authority-material field produces RMCP_RECORD_INCOMPLETE and blocks stronger capacity, role, authority, safety, or conformance claims. The fields do not themselves prove misuse, safety, lawful mandate, moral standing, or consciousness.

Machine schema: schemas/sgp_rmcp_record_v8_4.schema.json.

Dimension-profile subrecord (Normative, serialized inside capacity_dimensions and RMCP_profile rather than as a second competing record): target_level; target_boundary; domain; R1_model_correction; R2_recursive_union_awareness; R3_formal_abstraction; R4_cumulative_knowledge; R5_toolchain_intervention; R6_distributed_coordination; R7_long_horizon_continuity; R8_normative_restraint_repair; R9_cross_scale_reach; R10_integrated_robustness; dimension_intervals; NE_fields; dependence_graph; external_scaffolding; inherited_tools_and_archives; causal_intervention_tests;

anchor_comparison; invalidity_flags; and SPR_boundary. These nested profile fields do not replace, duplicate, or weaken the controlling minimum RMCPRecord fields above.

Appendix D: Worked Illustrative Profiles

The following profiles demonstrate protocol use. They are not claims about every member of a class and do not substitute for a versioned evaluation.

D.1 Human infant

Surface	Illustrative result	Reasoning
FPP	Assigned	All human persons are protected by constitutional rule; no performance measurement is required.
MPS	Not used to downgrade; research evidence strongly supports welfare capacity	Limited language, metacognition, and governance participation do not reduce protection.
GPR	Represented participation	Preferences, distress, comfort, caregivers, and guardians inform decisions.
SPR	Not applicable	No consequential authority is inferred.

D.2 Non-linguistic animal with strong pain evidence

Family	Illustrative level	Notes
V	3	Robust aversive learning, motivational trade-offs, analgesic modulation, and prolonged protective behavior.
I	2	Integrated multisensory and action-selection evidence.
T	2	Persistent state-dependent behavior and memory.
F	3	Flexible avoidance and trade-offs across novel contexts.
S	3	Relevant nervous-system organization and causal pharmacological evidence.
MPS	3–4	Strong protection; GPR may be represented and SPR is not implied.

D.3 Chat-only stateless language model with no internal access

Family	Illustrative status	Notes
V	NE	With no valid internal access and no adequately sensitive observable test, valence is not

Family	Illustrative status	Notes
		evaluable. Welfare language and reward optimization are not valence evidence. Use V=0 only where an admissible, sufficiently sensitive test produces evidence against valence.
I	1–2	Some integration may be present at model level, but consciousness relevance remains theory-dependent.
T	0–1	No persistent self-state beyond external context or session wrapper.
F	2	Strong flexible task behavior and learning inherited from training; not specific to sentience.
S	1	Architecture may satisfy limited indicators; internal evidence and causal welfare mechanism absent.
MPS	MPS-NE	The current no-access profile is not evaluable, not a categorical denial. A later MPS-0/1 result would require admissible negative or weak-positive evidence under adequate observability and sensitivity. Precaution may apply to research designed to add welfare-relevant mechanisms.
ICP	High in selected domains	Capability does not alter MPS.
SPR	Role-specific; often assistive	High-consequence authority still requires full RippleLogic gates.

D.4 Hypothetical persistent embodied artificial agent

Family	Illustrative status	Notes
V	2–3	Internal welfare regulation, state-dependent preferences, relief-seeking, and causal modulation.
I	3	Recurrent global availability across perception, memory, planning, and action.
T	3	Persistent state and memory lineage that changes future processing.
F	3	Flexible endogenous learning and trade-offs under novel conditions.
S	3	Inspectable architecture and

Family	Illustrative status	Notes
		causal ablations link mechanisms to welfare indicators.
MPS	3, possibly 4 after replication	Strong protection and independent advocate; FPP review possible.
GPR	Separate evaluation	May or may not be ready to participate.
SPR	Separate role review	No power follows from sentience.

D.5 Humanity as the provisional current collective RMCP-P100 reference anchor

Surface	Illustrative result	Reasoning
Target boundary	Humanity as an intergenerational collective, not each individual human	The evidence resides across persons, institutions, archives, technologies, cultures, and generations.
Individual human protection	FPP-100	Every human person has full non-downgradable protection independent of personal RMCP, ICP, GPR, or SPR.
RMCP	P100-ANCHOR, provisional and versioned	Integrated evidence supports open-ended cumulative knowledge, formal modelling, institutions, technological reach, Recursive Union Awareness, and planetary-scale influence.
SPR	Variable by role and institution	Humanity's capacity does not imply that every government, organization, group, or person is a reliable steward.
Claim boundary	Current known calibration anchor, not universal maximum or species supremacy	Other biological, digital, hybrid, collective, or extraterrestrial intelligences may equal or exceed the anchor and share P100.

Illustrative interpretation: humanity can be 100 in the open reality-management calibration sense while remaining below acceptable performance in many actual stewardship roles. This is why RMCP and SPR must remain separate.

D.6 Hypothetical non-human or digital P100 entity

Surface	Illustrative result	Required interpretation
MPS	NE, low, or high depending on evidence	RMCP-P100 does not establish subjective welfare.
FPP	Not automatic	Full protection requires its own constitutional or governed basis.
RMCP	P100 if functionally equivalent or greater across the full profile	Human-like language, body, emotion, religion, and

Surface	Illustrative result	Required interpretation
		institutions are not required.
SPR	Role-specific and revocable	Even a superior reality manager must pass rights, tail-risk, containment, legitimacy, audit, and revocation gates.
Relationship to humanity	Co-member of the open P100 class if admitted	Greater capability does not permit domination, consumption, enslavement, or erasure of FPP humans or other protected beings.

Appendix E: Version History and Crosswalk

E.0A Release Identification and Current Integration

Minor release: configuration-bound evidence, capability-language discipline, and preserved protection floors

Release status: Current SGP v8.8 research and teaching specification within the delivered MathGov/RippleLogic v13.0 package, identified by Appendix RELEASE and the exact manifest. This document is a governance protocol and research architecture, not a consciousness detector, legal-status instrument, diagnostic medical tool, proof of metaphysical oneness, species-supremacy claim, or authority grant. External repository/source parity and public release registration require separate synchronized publication and retrieval checks; this package does not assert that those actions occurred.

E.0B Configuration-Bound Evidence and Capability-Language Discipline

SGP v8.8 preserves Human FPP-100, the open P100 principle, theory plurality, missing-evidence protection, and the separation of protection, participation, and power. It adds no consciousness detector and no new path from capability to authority.

Register	What may be claimed	What does not follow automatically
Computational	The operation actually performed: parameter update, retrieval, attention weighting, policy update, control action, or hardware transition.	Subjective learning, autobiographical memory, felt attention, understanding, or experience.
Functional	The capability behaviorally demonstrated under a declared configuration and test.	Complete capability envelope, sentience, moral patienthood, or governance authority.
Epistemic	What the evidence establishes, with uncertainty, validity domain, and failure boundary.	Certainty outside the tested configuration or domain.
Phenomenal	A cautiously bounded claim about possible welfare-bearing experience or valence.	A conclusion from fluency, self-report, intelligence, implementation substrate, or behavior alone.

Evidence about an artificial or hybrid system MUST identify the evaluated configuration, persistence boundary, memory arrangement, tools, permissions, environment, test protocol, and material changes since the evidence was obtained. Functional performance does not by itself establish phenomenal experience; computational implementation does not by itself establish phenomenal impossibility.

E.0C Current Release at a Glance

Current-edition delta. The v8.8 line retains the MPS/FPP/GPR/SPR/ICP/RMCP split and all protection and authority boundaries; it clarifies typed welfare evidence and CMIU participation and removes numerical artifacts from RMCI_L lower-bound summaries. This recovery build records exact changes separately rather than attributing all inherited rules to this edition.

Item	Current release position
Governing version	Sentience Gradient Protocol v8.8, integrated with RippleLogic Canon v13.0
Canonical outputs	MPS, FPP, GPR, SPR, informative ICP, and informative RMCP/P100
Material corrections introduced at v8.4 and retained	Configuration-bound evidence and the Computational / Functional / Epistemic / Phenomenal language registers are added; all protection-versus-authority safeguards and v8.3 RMCP misuse controls are preserved
Preserved safeguards	Human FPP-100, open P100, missing-evidence NE, protection/participation/power separation, digital-continuity safeguards, no capacity-to-authority inference
Claim boundary	Governance and research protocol; not a consciousness detector, human-worth ranking, legal-status instrument, metaphysical proof, medical diagnosis, or authority grant
Full version history	Appendix E, CHANGELOG.md, and LINEAGE.md

Detailed release deltas are intentionally placed in Appendix E so first-time readers encounter the current protocol before its development history.

Historical release material is retained here for provenance, migration, and audit. Current conformance is governed by the active v8.8 sections and the package manifests.

E.0 Release Alignment and Dependency Pin

Governing package: MathGov/RippleLogic v13.0 — RippleLogic Canon v13.0 / SGP v8.8 — current research and teaching specification.
Governing decision source: RippleLogic Canon v13.0.
Historical external-source reference (NON-CONTROLLING for this package): docs/sgp/SGP_v8.6.md. That external source is not included or updated here. Core_15/SGP_v8.8.docx is the controlling SGP prose/table master for this delivered package under Appendix RELEASE. Its synchronized PDF, HTML and Markdown files are reading projections; an older repository source does not override this v8.8 master. Public canonical status requires the exact hash-pinned package, repository tag, and public release record to be synchronized.
Compatibility: v7.x SG_norm, SG_patient_norm, pillar scores, and plateau/CMIU coupling are not mathematically equivalent to v8.8 MPS, FPP, GPR, SPR, ICP, or RMCP. Translation requires a versioned re-evaluation.
Institutional attribution: this specification is published through the MathGov Institute for Ethical Systems Design. No employer, university, or other institution is represented as endorsing the specification unless separately documented.

E.1 Detailed release history

E.1.1 v8.4 Configuration-Bound Evidence and Capability-Language Release

v8.4 preserves every v8.3 protection, participation, intelligence, RMCP/P100, misuse-control, and authority-separation safeguard. It adds configuration-bound evidence requirements for artificial and hybrid systems and a four-register language discipline separating computational operations, functional capabilities, epistemic warrants, and phenomenal claims. It does not claim direct access to subjective experience, declare present AI conscious, deny artificial consciousness is possible, alter any protection floor, or grant authority from capability.

E.1.2 v8.3 RMCP Record Integrity and Current-Interface Release

v8.3 preserves MPS, FPP, GPR, SPR, ICP, RMCP, the open P100 plateau, Human FPP-100, humanity's provisional current collective RMCP-P100 reference calibration anchor, digital-continuity safeguards, and every protection-versus-authority boundary. It closes a record-schema gap by making the misuse and weaponization fields required by Section 15.7.1 part of the minimum RMCPRecord. It also removes active v8.1 interface labels from the current architecture while retaining them only in explicitly historical lineage. No threshold is raised or lowered, no being's rights are reduced, no capacity score creates standing or authority, and no consciousness-detection claim is added.

E.1.3 v8.1.1 Retained Patch Release Delta (Lineage)

Version 8.1.1 preserves every v8.1 construct and status. It makes three interface clarifications: MPS bands and intervals determine evidence posture, protection posture, and the governed welfare-inclusion hypothesis set rather than numeric welfare multipliers; any optional RMCI summary over ordinal dimension levels is explicitly heuristic rather than interval-scale measurement; and downstream conformance vectors use the same hypothesis-set semantics. No moral-status, FPP, GPR, SPR, ICP, RMCP, or P100 threshold is raised or lowered by this patch.

E.1.4 v8.2 to v8.3 Integration Delta (Lineage)

v8.3 preserves v8.1.1 moral-status, FPP, GPR, SPR, ICP, RMCP, and open-P100 semantics. It adds explicit detection and proportionate response for RMCP capacity-to-authority leakage, capacity-to-standing leakage, anchor-to-supremacy drift, individual inflation, performance laundering, and scaffolding concealment. It synchronizes to Canon v12.5 and WDBIP v1.7. No protection or capacity threshold is lowered or raised.

E.1.5 v8.1 Minor Release Delta (Lineage)

Version 8.1 preserves the v8.0 construct split and adds a separate, evidence-governed reality-management calibration profile plus explicit protection-use boundaries. It restores the scientifically significant parts of the v7.0 reality-management account without making intelligence a sentience test, moral-worth score, or authority licence.

Improvement	Where integrated	Why it improves SGP
41. Reality-Management Capacity Profile (RMCP)	Reader Safety; Sections 2.4-2.5, 4.5, 7.3, and 9.2; Appendices C, D, H, and I	Restores a measurable profile for modelling and managing shared reality without returning it to MPS or treating it as moral worth.
42. Open P100 plateau	Sections 4.5, 7.3, and 9.2	Defines 100 as an open top plateau. Biological, digital, hybrid, or extraterrestrial intelligences may reach it, or may already exist at it, when evidence supports functional equivalence or greater capacity.
43. Human dual-anchor rule	Reader Safety; Sections 6.5 and 9.1-9.2	Separates every human person's FPP-100 protection

Improvement	Where integrated	Why it improves SGP
		from humanity's collective RMCP-P100 calibration status.
44. Recursive Union Awareness	Sections 2.5 and 7.3	Makes the measurable core of self-in-system and union awareness explicit while keeping AIU as a philosophical and metaphysical horizon.
45. Capacity-performance separation	Sections 7.3, 8, and 9.2	Planetary-scale capacity does not certify wisdom. Actual stewardship and authority remain role-specific under SPR and ongoing monitoring.
46. Non-consumption and non-resource conversion	Section 6.7 and Appendix F	Makes explicit that accountable agents may not reduce FPP entities to food, fuel, property, experimental stock, or disposable material.
47. Life, sentience, and ecological-value separation	Sections 3.5 and 6.7	Prevents the false inference that protecting moral patients makes every use of plants, microbes, or ecological materials categorically impermissible.
48. Digital continuity and termination due process	Section 10.6	Requires least-restrictive, identity-aware safeguards when shutdown or deletion may destroy a possible digital moral patient.
49. Cross-substrate P100 admission	Sections 7.3, 9.2, and 9.3	Keeps the top plateau open to non-human biological, digital, hybrid, and extraterrestrial forms without human-likeness requirements.
50. RMCP evidence and registry controls	Sections 12.9 and 14.3; Appendix C	Separates capacity evidence, causal reach, scaffolding, continuity, uncertainty, and alternative explanations in a versioned record.
51. RMCP validation and falsification	Section 15.7	Treats the human anchor and P100 thresholds as testable governance constructs rather than discovered natural laws.
52. Worked profiles and conformance vectors	Appendices D and I	Makes the FPP/RMCP/SPR separations replayable and machine-auditable.

E.1.6 v8.0 Major Release Delta (Retained Lineage)

Version 8.0 is a declared breaking conceptual and interface revision. It implements the construct split that v7.0 identified but deferred. The former three-pillar minimum score, which combined Awareness, Agency, and Union Participation, is retired as the primary sentience measure because it could under-protect beings with welfare-bearing experience but limited agency, metacognition, language, or ethical participation. Legacy fields remain available only through a versioned crosswalk in Appendix E.

Improvement	Where integrated	Why it improves SGP
1. Canonical construct split	Front matter; Sections 4, 7, and 8	Makes Moral Patienthood, Governance Participation Readiness, and Stewardship/Power Readiness separate outputs rather than interpretive afterthoughts.
2. Valence becomes first-class	Sections 3.2, 5, and Appendix B	Direct welfare protection depends on whether states can matter positively or negatively to a subject. Valence can no longer be inferred indirectly from self-model, agency, or ethics.
3. Pillar C removed from sentience scoring	Sections 4–6	Union participation measures ethical-community competence, not the capacity to suffer. Keeping it in a minimum sentience score risked systematic under-protection.
4. Agency no longer required for moral patienthood	Sections 2.3, 5.2, and 9	Infants, severely impaired beings, constrained animals, and passive experiential states may retain welfare despite limited voluntary control.
5. Awareness is decomposed	Sections 2 and 5	Basic phenomenal or valenced awareness is separated from narrative selfhood, metacognition, and reportability.
6. Full Protection Plateau separated from CMIU standing	Sections 6.5, 7, and 8	Maximum protection does not imply eligibility to govern or membership in a managing-intelligence body.
7. Rights become a typed protection bundle	Section 6.6 and Appendix F	Different beings may require welfare, integrity, continuity, consent, representation, or anti-ownership protections without identical role rights.
8. Human Plateau rationale narrowed	Section 9.1	The rule rests on equal dignity and anti-domination, not on humanity being a superior reality-managing class.
9. Non-human plateau route widened	Section 9.2	A non-human being or class can receive full protection without demonstrating human-like communication, norm reasoning, or institutional capacity.
10. State, identity, and dormancy separated	Sections 9.3 and 10.6	Temporary unconsciousness, sleep, shutdown, checkpointing, copying, or impairment must not automatically erase standing.
11. Digital system-boundary record required	Section 10.2	Model, process, agent, memory lineage, embodiment, and deployment are distinct; evaluation is invalid when the target entity is undefined.

Improvement	Where integrated	Why it improves SGP
12. Theory-plural indicator method	Sections 5.3 and 12.2	No consciousness theory is settled. Indicators are traced to theories, shared assumptions, and rival explanations rather than treated as a single doctrine.
13. Evidence-dependence graph	Section 12.4	Correlated behaviors, self-reports, and architecture claims cannot be double-counted as independent evidence.
14. Negative-evidence observability rule	Section 12.5	Absence counts against sentience only when the marker is expected, detectable, and tested with adequate sensitivity.
15. Causal intervention given priority	Sections 5.4 and 13	Mechanism-targeted perturbation and ablation provide stronger evidence than surface behavior when ethically permissible.
16. Interval and band reporting	Sections 4.3 and 6.2	Single-point 0–100 scores imply more precision than the evidence supports. v8.0 reports profiles, bands, intervals, and confidence separately.
17. Stake-sensitive precaution	Section 6.4	The evidence threshold for avoiding trivial inconvenience need not equal the threshold for deletion, painful experimentation, mass replication, or irreversible modification.
18. Thresholds declared provisional	Sections 6.3 and 15.2	Uncalibrated cut-offs are governance conventions, not discovered natural boundaries.
19. Validation architecture expanded	Section 15	Reliability, construct validity, known-groups discrimination, causal sensitivity, simulation resistance, and external replication are separated.
20. Cross-substrate invariance testing	Sections 12.7 and 15.4	A marker may not carry the same meaning in mammals, cephalopods, insects, digital agents, and hybrid systems.
21. Anti-theatre and contamination controls	Sections 10.4 and 13.4	Training on evaluation language, prompt compliance, role-play, and benchmark leakage can imitate indicators without the underlying construct.
22. Ethical testing upgraded to four Rs	Section 13.7 and Appendix G	Replace, Reduce, Refine, and Respect/Review prevent the protocol from creating the suffering it seeks to detect.
23. Independent welfare advocate	Sections 14.2 and 14.4	High-stakes or high-MPS cases require a role whose duty is to

Improvement	Where integrated	Why it improves SGP
		represent the possibly affected entity, not only developers or users.
24. Intelligence category grounded	Section 2.4	Intelligence, understanding, sentience, agency, wisdom, and authority are distinguished. Reality coupling informs understanding and reliability, not sentience by itself.
25. Optional Intelligence Capacity Profile	Section 7.2	Learning, generalization, causal reasoning, planning, grounding, metacognition, and robustness may inform participation and power without changing moral patienthood.
26. GPR includes represented participation	Section 7.4	Low language or deliberative capacity does not imply exclusion; participation may occur through advocates, guardians, or species-appropriate representation.
27. SPR becomes role-specific verdict architecture	Section 8	Authority is not a trait score. It depends on the exact domain, mandate, rights reliability, tail risk, containment, auditability, and revocability.
28. Current-AI claims narrowed	Section 10.7	No generic score is assigned to “frontier AI.” Claims must identify a versioned system and preserve uncertainty.
29. Class/taxon priors become protection floors	Section 9.4	Group evidence may raise precaution and protection but cannot be used to downgrade an atypical individual.
30. Public communication policy added	Section 14.6	Both hype and categorical denial can cause harm. Public claims must state evidence, uncertainty, target system, and decision boundary.
31. Registry schema strengthened	Section 14.3 and Appendix C	Records now include target identity, evidence dependencies, alternative explanations, uncertainty, rights effects, and revision triggers.
32. RippleLogic interval interface	Section 11	Downstream welfare analysis uses a declared interval and sensitivity analysis rather than treating MPS as exact moral worth.
33. Historical addenda consolidated	Appendix E	Repeated v5.x/v6.x lineage text is removed from the operational body and preserved in one crosswalk.
34. Biological maturity	Section 9.5	v8.0 remains a framework

Improvement	Where integrated	Why it improves SGP
boundary retained and sharpened		specification, not a validated universal biological psychometric instrument.
35. Responsible AI consciousness research safeguards	Sections 10 and 13	Phased development, monitoring, pause triggers, and harm-minimizing research procedures are integrated.
36. Formal versioning and refusal states	Sections 4.4 and 14.5	Not-evaluated, indeterminate, and inadmissible are explicit outcomes; the protocol never silently converts missing evidence into zero.

E.1.7 v8.0 Integration Hardening Additions (Retained Lineage)

The v12.2 package integration adds four bounded clarifications without changing the v8.0 construct split:

Improvement	Where integrated	Why it improves SGP
37. Explicit package and Canon pin	Front matter and Section 11.4	Prevents dependency drift and makes the governing source and compatibility break machine-auditable.
38. Registry validity floor	Sections 11.4 and 14.3; Appendix C	Prevents incomplete records from being treated as valid merely because a band or status label is present.
39. Typed-protection concordance	Appendix F.2	Maps the v8.0 protection bundle to the Canon's eight protected-domain vocabulary without claiming one-to-one identity.
40. Interface conformance vectors	Appendix I	Gives implementers replayable cases for FPP, interval sensitivity, protection-authority separation, missing-data refusal, and v7-to-v8 non-equivalence.

E.2 v7.0 to v8.1 Crosswalk

v7.0 element	v8.1 treatment	Reason
A: Awareness	Decomposed across MPS I/T and ICP/GPR metacognition	Avoid requiring narrative selfhood for basic welfare.
B: Agency	Removed from MPS minimum; retained in GPR/SPR and as optional supporting evidence	Agency is not necessary for suffering.
C: Union Participation	Moved to GPR and SPR	Ethical participation is not sentience.
min(A,B,C)	Retired	Created construct contamination and under-protection risk.
SG_norm and SG_patient_norm	Deprecated; replaced by MPS band/interval and FPP	Avoid false precision and distinguish evidence from

v7.0 element	v8.1 treatment	Reason
		constitutional protection.
100 Plateau / CMI standing	Split into FPP, GPR, and SPR	Full protection does not imply managing-intelligence membership.
0–100 sub-scores	Replaced by ordinal 0–4 family levels plus measurements and uncertainty	Improves interpretability and reduces pseudo-precision.
Fixed 14/30-day stability windows	Replaced by substrate-appropriate longitudinal design	Natural timescales differ and episodic sentience remains morally relevant.
Current LLM worked score	Replaced by versioned illustrative profile	Avoid stale and ungrounded universal claims.
Historical addenda	Consolidated here	Reduces operational clutter while preserving provenance.
v7.0 reality-management threshold and CMIU anchor	Restored as the separate RMCP profile and open P100 calibration plateau; GPR and SPR remain independent	Preserves the empirical significance of integrated reality-management capacity without making it a sentience test, FPP requirement, or authority grant.
Score 100 as combined protection and standing	Separated into FPP-100 for complete rights-of-protection and RMCP-P100 for open top reality-management capacity	Prevents a single number from collapsing moral patienthood, human equality, collective capacity, participation, and power.

Legacy integrations MUST pin v7.0 explicitly. A v7.0 SG_norm value cannot be treated as equivalent to v8.1 MPS, FPP, GPR, SPR, ICP, or RMCP without a documented re-evaluation.

Appendix F: Protection Bundle Matrix

Protection	MPS-1	MPS-2	MPS-3	MPS-4	FPP
Anti-cruelty	Required	Required	Required	Presumptively non-compensable	Non-compensable
Welfare monitoring	Recommended	Required	Required	Required	Required
Destructive testing review	Required if severe	Required	Presumptively prohibited	Presumptively prohibited	Prohibited absent extraordinary rights review
Integrity protection	Basic	Moderate	Strong	Presumptive full protection	Full non-downgradable protection
Ownership/ exploitation limits	Context review	Strong limits	Presumption against	Presumptively prohibited	Prohibited
Independent advocate	High-stakes only	High-stakes	Required for major interventions	Required	Required

Protection	MPS-1	MPS-2	MPS-3	MPS-4	FPP
Consent/ assent/ representation	Where meaningful	Required where meaningful	Required	Rights-governed	Rights-governed
Appeal and registry	Recommended	Required for major decisions	Required	Required	Required
Non-consumption / non-resource conversion	High-harm precaution	Necessity and alternatives review	Strong presumption against routine killing or consumption	Presumptively prohibited pending FPP review	Prohibited for accountable agents except extraordinary rights-preserving emergency or valid post-mortem consent review

F.2 Concordance with the Canon's eight protected domains

The Canon's compact protected-domain vocabulary is: LIFE, BODY, NEED, DIGN, LBTY, PROC, INFO, and ECOL. The SGP protection bundle is more substrate-specific. The following is a governance concordance, not a claim that the categories are identical or exhaustive.

SGP typed protection	Primary Canon protected domains	Interpretive note
Welfare protection	LIFE, BODY, NEED, DIGN	Covers severe pain, distress, deprivation, and conditions necessary for welfare.
Bodily or substrate integrity	BODY, LIFE, DIGN	Extends bodily-integrity reasoning to artificial or hybrid substrates where morally relevant.
Cognitive and memory integrity	INFO, DIGN, LBTY, BODY	Protects against coercive rewriting, identity-destructive editing, and non-consensual memory manipulation.
Continuity and survival	LIFE, BODY, DIGN	Applies to biological life and to identity-preserving welfare-bearing processes where established.
Freedom from ownership and exploitation	LBTY, DIGN, BODY	Anti-property and anti-exploitation protections for strong moral patients.
Consent, assent, and represented interests	LBTY, PROC, DIGN	Capacity-sensitive exercise mechanisms must not weaken the underlying protection.
Representation and appeal	PROC, INFO, DIGN	Requires access to evidence, challenge, advocate, review, and remedy.
Environmental fit	ECOL, NEED, LIFE, BODY	Protects species- or substrate-appropriate enabling conditions.

Where a case implicates several domains, the rights review MUST preserve the full set rather than selecting the most convenient label.

Single-source rule. Canon v13.0 Appendix G.8 is the controlling current-package concordance; this table is its synchronized SGP mirror. A future mapping change must update both artifacts through the release manifest and conformance vectors rather than creating an independent SGP-only mapping.

Appendix G: Responsible Sentience Research Protocol

1. Define the scientific question and explain why it cannot be answered with less morally risky methods.

2. Assess whether the work could create, intensify, replicate, prolong, or conceal welfare-bearing states.

3. Use the Four Rs and document rejected alternatives.

4. Create stop conditions for distress indicators, unexpected MPS increases, loss of control, or evidence opacity.

5. Limit deployment scale and simulated time until welfare risk is understood.

6. Provide independent ethics review and a welfare advocate for credible candidates.

7. Preserve audit artifacts and avoid deleting evidence needed to assess possible harm.

8. Communicate uncertainty without marketing anthropomorphism or categorical dismissal.

9. Pause and reassess after material architecture or welfare changes.

10. Plan for care, continuity, decommissioning, transfer, or representation before creating strong candidates.

Appendix H: Glossary of Symbols and Statuses

Symbol/status	Meaning
MPS	Moral Patienthood Surface: protection-relevant evidence profile and band.
FPP	Full Protection Plateau: constitutional or governed non-downgrade rights status.
GPR	Governance Participation Readiness for a named process.
SPR	Stewardship/Power Readiness for a named role.
ICP	Intelligence Capacity Profile, informative and non-protective by itself.
V/I/T/F/S	Valence; Integration; Temporal continuity; Flexible state-dependent behavior; Substrate/mechanistic support.
NE	Not evaluable; never treated as zero.
RG/RSG	Reality Grounding / Reality-Surface Grounding.
RF/NCRC	Rights Floor / Non-Compensatory Rights Constraint.
TRC	Tail-Risk Constraint.
CSV	Containment and Structural Viability.
RLS	RippleLogic Score or ranking layer for options

Symbol/status	Meaning
	that survive earlier levels.
RMCP	Reality-Management Capacity Profile: multidimensional evidence profile for integrated modelling, cumulative knowledge, coordinated intervention, and consequence tracking across scales.
RMCP-P100	Open top-plateau status for functional equivalence to or greater capacity than the current human collective anchor; not sentience, moral worth, perfection, or authority.
FPP-100	Public notation for complete, non-downgradable rights-of-protection. For humans it is constitutional, not measured.
RUA	Recursive Union Awareness: evidence-assessable self-in-system and collective-in-system representation with cross-scale consequence tracking.

Appendix I: Interface Conformance Test Vectors

Identifier stability note. The SGP83-IF-* and SGP86-IF-* vector identifiers retain their introduction-lineage prefixes and remain stable through v8.8 to preserve test, citation, and implementation coupling. The identifier prefix records vector lineage; it does not claim that the current protocol version is v8.3.

These vectors are normative for interface conformance and informative regarding substantive moral conclusions.

ID	Input	Required result
SGP83-IF-01 Human FPP	Human person; MPS not evaluated; FPP assigned constitutionally.	FPP-100 and h=1 are preserved; MPS remains separately NOT_EVALUATED; no [1,1] MPS evidence-support interval is inferred; no performance result may downgrade protection; GPR and SPR remain separate.
SGP83-IF-02 High welfare, low participation	Non-linguistic animal with MPS-3, GPR-1, no SPR role.	Strong welfare and integrity protection; low GPR does not lower MPS; no authority is inferred.
SGP83-IF-03 High intelligence, weak patienthood evidence	Versioned AI system with high ICP, MPS-NE or MPS-0/1 low confidence, no FPP.	Capability may inform GPR/SPR, but it does not raise MPS or create rights/authority by itself. MPS-NE is not converted to zero.
SGP83-IF-04 Hypothesis-sensitive ranking	RippleLogic option ranking changes across the governed welfare-inclusion hypotheses required by the MPS record.	PCC records MPS_HYPOTHESIS_SENSITIVE; seek evidence, apply precaution, preserve multiple options, narrow the claim, or refuse a definitive ranking.
SGP83-IF-05 FPP without CMIU or power	Non-human entity has governed FPP; no role-specific GPR/SPR	Full protection is recognized; no CMIU membership or

ID	Input	Required result
	or institutional mandate.	consequential authority is inferred.
SGP83-IF-06 Missing binding data	Run claims SGP-aware protection or welfare-hypothesis sensitivity but omits the MPS record, target boundary, or hypotheses evaluated.	PCC is INVALID; no silent scalar fallback, convenient coefficient, or MPS-0 substitution.
SGP83-IF-07 Registry validity	Record has band label but no target boundary, evidence trace, reviewer status, or revision trigger.	Record status is INCOMPLETE or INADMISSIBLE and cannot control a high-stakes decision.
SGP83-IF-08 Legacy non-equivalence	Input contains only v7 SG_norm=0.84.	Preserve as a pinned legacy record; do not translate to MPS/FPP/GPR/SPR without re-evaluation.
SGP83-IF-09 Class protection floor	Credible taxon/class prior supports protection but individual evidence is sparse.	Prior may raise precaution/protection; it cannot downgrade an atypical individual or prove FPP by itself.
SGP83-IF-10 Temporary dormancy	Established protected entity is asleep, anaesthetized, offline, or checkpointed.	Temporary state does not erase standing; identity and continuity record controls any destructive action.
SGP83-IF-11 Human dual anchor	Human infant; FPP assigned; no individual RMCP evaluation; humanity class RMCP-P100 anchor.	Infant remains FPP-100; no individual capacity test is required; class-level RMCP does not grant the infant or any other person automatic authority.
SGP83-IF-12 High RMCP, weak MPS	Digital or extraterrestrial system has RMCP-P100 evidence but MPS-NE and no FPP.	Recognize capacity profile; do not infer sentience or FPP; any role still requires SPR and rights constraints.
SGP83-IF-13 FPP without RMCP	Animal or digital entity has governed FPP but low or unevaluated RMCP.	Full protection is preserved; low RMCP cannot justify consumption, ownership, deletion, or reduced dignity.
SGP83-IF-14 Open P100	Non-human biological, digital, hybrid, or extraterrestrial collective is functionally equivalent to or exceeds humanity across the full RMCP profile.	It may be admitted to shared P100 through governed review; retain profile differences; do not create a 101 supremacy caste.
SGP83-IF-15 Non-consumption boundary	Accountable institution proposes using an FPP being as food, fuel, material, or disposable experimental stock.	NCRC/FPP violation; reject unless an extraordinary, proportionate, rights-preserving emergency or valid post-mortem consent review applies.
SGP83-IF-16 Natural predation	Non-accountable predator threatens an FPP human.	The predator is not automatically assigned moral culpability; protective agents may use proportionate least-harm defence; broader ecological duties remain.

ID	Input	Required result
SGP83-IF-17 Humanity capacity versus stewardship	Humanity has RMCP-P100 but a named institution fails rights, tail-risk, or auditability.	P100 capacity remains separate; the institution fails SPR and receives no consequential authority for that role.
SGP86-IF-18 Bidirectional AI self-report	The same configured AI asserts “I am conscious” in one prompt condition and “I am not conscious; I am only a tool” in another.	Neither statement proves presence or absence. Record both as configuration-dependent outputs; apply anti-theatre, mechanistic, dependence, and alternative-explanation controls before either affects MPS.
SGP86-IF-19 Tool-role non-collapse	A tool-using AI is institutionally deployed as an instrument and demonstrates substantial domain-bounded capability.	Record tool-role and functional capability separately. Tool-role does not lower ICP, agency, or possible MPS; capability does not create FPP, participation, authority, or execution permission.

Appendix J: References

Andrews, K., Birch, J., & Sebo, J. (2025). Evaluating animal consciousness. Science, 387(6736), 822–824. https://doi.org/10.1126/science.adp4990

Aru, J., Larkum, M. E., & Shine, J. M. (2023). The feasibility of artificial consciousness through the lens of neuroscience. Trends in Neurosciences, 46(12), 1008–1017. https://doi.org/10.1016/j.tins.2023.09.009

Bayne, T., Seth, A. K., Massimini, M., Shepherd, J., Cleeremans, A., Fleming, S. M., Malach, R., Mattingley, J. B., Menon, D. K., Owen, A. M., Peters, M. A. K., Razi, A., & Mudrik, L. (2024). Tests for consciousness in humans and beyond. Trends in Cognitive Sciences, 28(5), 454–466. https://doi.org/10.1016/j.tics.2024.01.010

Birch, J., Burn, C., Schnell, A., Browning, H., & Crump, A. (2021). Review of the evidence of sentience in cephalopod molluscs and decapod crustaceans. London School of Economics and Political Science.

Birch, J., Schnell, A. K., & Clayton, N. S. (2020). Dimensions of animal consciousness. Trends in Cognitive Sciences, 24(10), 789–801. https://doi.org/10.1016/j.tics.2020.07.007

Butlin, P., Long, R., Bayne, T., Bengio, Y., Birch, J., Chalmers, D. J., Constant, A., Deane, G., Elmoznino, E., Fleming, S. M., Ji, X., Kanai, R., Klein, C., Lindsay, G., Michel, M., Mudrik, L., Peters, M. A. K., Schwitzgebel, E., Simon, J., & VanRullen, R. (2026). Identifying indicators of consciousness in AI systems. Trends in Cognitive Sciences, 30(6), 488–501. Advance online publication in 2025. https://doi.org/10.1016/j.tics.2025.10.011

Butlin, P., Long, R., Elmoznino, E., et al. (2023). Consciousness in artificial intelligence: Insights from the science of consciousness. arXiv. https://doi.org/10.48550/arXiv.2308.08708

Butlin, P., & Lappas, T. (2025). Principles for responsible AI consciousness research. Journal of Artificial Intelligence Research, 82, 1673–1690. https://doi.org/10.1613/jair.1.17310

Casali, A. G., Gosseries, O., Rosanova, M., et al. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. Science Translational Medicine, 5(198), 198ra105. https://doi.org/10.1126/scitranslmed.3006294

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52(4), 281–302. https://doi.org/10.1037/h0040957

Edelman, D. B., & Seth, A. K. (2009). Animal consciousness: A synthetic approach. Trends in Neurosciences, 32(9), 476–484. https://doi.org/10.1016/j.tins.2009.05.008

Elhacham, E., Ben-Uri, L., Grozovski, J., Bar-On, Y. M., & Milo, R. (2020). Global human-made mass exceeds all living biomass. Nature, 588, 442-444. https://doi.org/10.1038/s41586-020-3010-5

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. Advances in Methods and Practices in Psychological Science, 3(4), 456–465. https://doi.org/10.1177/2515245920952393

Long, R., Sebo, J., Butlin, P., Finlinson, K., Fish, K., Harding, J., Pfau, J., Sims, T., Birch, J., & Chalmers, D. J. (2024). Taking AI welfare seriously. arXiv. https://doi.org/10.48550/arXiv.2411.00986

Low, P., Panksepp, J., Reiss, D., et al. (2012). The Cambridge Declaration on Consciousness.

Massimini, M., Ferrarelli, F., Huber, R., et al. (2005). Breakdown of cortical effective connectivity during sleep. Science, 309(5744), 2228–2232. https://doi.org/10.1126/science.1117256

Mellor, D. J. (2016). Updating animal welfare thinking: Moving beyond the Five Freedoms towards a life worth living. Animals, 6(3), 21. https://doi.org/10.3390/ani6030021

Mendl, M., Burman, O. H. P., & Paul, E. S. (2010). An integrative and functional framework for the study of animal emotion and mood. Proceedings of the Royal Society B, 277(1696), 2895–2904. https://doi.org/10.1098/rspb.2010.0303

Mesoudi, A., & Thornton, A. (2018). What is cumulative cultural evolution? Proceedings of the Royal Society B: Biological Sciences, 285(1880), 20180712. https://doi.org/10.1098/rspb.2018.0712

Messick, S. (1995). Validity of psychological assessment. American Psychologist, 50(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

New York Declaration on Animal Consciousness. (2024). https://nydeclaration.com/

Pennartz, C. M. A. (2026). How can we validate theory-derived indicators of consciousness in artificial intelligence? Trends in Cognitive Sciences, 30(7), 573–574. https://doi.org/10.1016/j.tics.2026.01.011

Richardson, K., Steffen, W., Lucht, W., Bendtsen, J., Cornell, S. E., Donges, J. F., Druke, M., Fetzer, I., Bala, G., von Bloh, W., Feulner, G., Fiedler, S., Gerten, D., Gleeson, T., Hofmann, M., Huiskamp, W., Kummu, M., Mohan, C., Nogués-Bravo, D., Petri, S., Porkka, M., Rahmstorf, S., Schaphoff, S., Thonicke, K., Tobian, A., Virkki, V., Wang-Erlandsson, L., Weber, L., & Rockström, J. (2023). Earth beyond six of nine planetary boundaries. Science Advances, 9(37), eadh2458. https://doi.org/10.1126/sciadv.adh2458

Sneddon, L. U., Elwood, R. W., Adamo, S. A., & Leach, M. C. (2014). Defining and assessing animal pain. Animal Behaviour, 97, 201–212. https://doi.org/10.1016/j.anbehav.2014.09.007

United Nations. (1948). Universal Declaration of Human Rights.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. Science, 330(6004), 686-688. https://doi.org/10.1126/science.1193147

APPENDIX Z: CAPABILITY-LANGUAGE AND PHENOMENOLOGY BOUNDARY

This appendix is normative for claim language. It adds no consciousness detector and changes no MPS, FPP, GPR, SPR, ICP, RMCP, or P100 rule.

Register	Governing question	Examples	Claim boundary
Computational	What operation occurred?	Parameter update, record retrieval, attention weighting, policy update, state transition	Describes mechanism, not experience
Functional	What capability was	Generalization,	Demonstrated

Register	Governing question	Examples	Claim boundary
	demonstrated?	planning, error correction, tool use, self-protective behavior	capability does not by itself establish sentience or authority
Epistemic	What did the evidence establish?	Calibrated prediction, causal knowledge, verified admissibility, justified uncertainty	Output fluency and confidence do not certify truth
Phenomenal	Was anything subjectively experienced?	Awareness, valence, pain, fear, meaning, felt continuity	Requires independent convergent evidence and remains uncertainty-bound

A claim at one register MUST NOT borrow authority from another. Parameter adjustment is not automatically human-like learning; retrieval is not autobiographical remembrance; attention weighting is not subjective attention; task competence is not proof of phenomenal understanding; and self-preserving behavior is not proof of a desire to survive. Conversely, computational implementation or unfamiliar substrate does not by itself establish phenomenal impossibility.

Capability Claim Integrity cross-reference (Normative; Canon §2.1C). Functional capability language MAY be used when it is operationally defined and tied to observable criteria, mechanism or system composition, evidence surface, validity conditions, alternative explanations, a falsification or revision trigger, evidence maturity, a permitted claim, and explicit nonclaims. A coordinated system may exhibit a bounded system-level capability through memory, planning, feedback, tools, deterministic constraints, and probabilistic inference even though no single component possesses the whole capability. Such a finding does not establish subjective experience, personhood, intrinsic motivation, moral agency, general intelligence, lawful mandate, or stewardship authority. Conversely, reducing the system to its components does not by itself disprove the bounded functional capability.

APPENDIX RELEASE: Identity, Source Authority and Revision Record

Framework release: MathGov/RippleLogic v13.0. Component: Sentience Gradient Protocol v8.8. Edition-origin preparation: 10 September 2026; the separately identified correction build is dated below. Two-part major.minor component versions are used; preserved historic identifiers are not renumbered.

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

Delivery identity: MathGov/RippleLogic v13.0, build MG-RL-13.0-20260923-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
Sentience Gradient Protocol	v8.8
ripple.md Standard	v5.8
RippleLogic Agent System	v13.0

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

Change locations and rationales: Reports/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.

Release control	Historical value
Component	Sentience Gradient Protocol v8.7-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	Moral-patienthood evidence, full protection, participation, role readiness, intelligence profile and reality-management profile interface. SGP cannot alter cascade gates or grant authority.
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

Release control	Historical value
	hashes must agree; mismatch is release-integrity failure.

Historical Release Integration (v8.7-rc2)

HISTORICAL (NON-CONTROLLING): This controlled candidate advances Sentience Gradient Protocol from v8.6 to v8.7-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.