PROJECT ENDURING PULSE

Testing and Validation Brief

Experiments, limitations and interpretation

Charter v1.2.2 | Website release 10 | 15 September 2026

https://enduringpulse.org

This is a public edition of a Project Enduring Pulse document. The Charter is reproduced unchanged. Explanatory documents include presentation and evidence clarifications dated 15 September 2026. Historical alternatives do not supersede the Charter; none of these documents is enacted law.

Research interpretation and source-report corrections: the recovered long-horizon code and aggregate outputs are available and checked. Conditional models do not independently validate their assigned institutional effects or forecast real-world probabilities. PULSE is fourth in six baseline comparisons, not second. Reduced-form counts are parameter samples, not simulated Markov histories. Human-directed, AI-assisted development and internal model critique are not external peer review. An earlier concise overview mixed a working-paper DOI with journal details; the current scholarly paper and reference register distinguish those editions. See https://enduringpulse.org/methods/.

Website prose and project documents: CC BY 4.0 to the extent applicable rights are held (https://creativecommons.org/licenses/by/4.0/). Third-party works retain their own rights. Project name, logos and brand assets are excluded from these general grants. No endorsement is implied. Publication terms and rights ownership remain subject to attorney review.

PROJECT ENDURING PULSE

Testing, Validation & Model-Claim Guide

What was tested, what held, what failed, and what can honestly be claimed

v1.2.2 Publication & Advancement Clarification | September 2026

Latest result: the long-horizon batch did not justify changing PULSE's major assumptions or five-doctrine architecture. The strongest new implementation warning is institutional complexity under weak state capacity.

1. Testing scale

Measure	Count	Meaning
Targeted model / stress-test families	48	Distinct micro/constitutional families through v1.2
Independently randomized targeted base scenarios/trials	43,415,600	Conservative recoverable count
Targeted comparative rule evaluations	>60.6 million	Many rules applied to the same base worlds for apples-to-apples comparison
Explicit multi-voter election worlds	31,960	Narrow election-world count
Long-horizon dynamic architecture trajectories	2,314,000	300-year scenario trajectories across baseline, sweeps, ablations, extremes and transition tests
Five-year long-horizon state transitions	138,840,000	Dynamic state updates; not independent civilizations
Reduced-form hazard/Markov trajectories	2,100,000	Structurally different cross-check
Baseline long-horizon archetypes	20	Demography, scarcity, integration, catastrophe, conflict, wealth, etc.
Compound extreme worlds	6	Multiple severe assumptions combined
Primary governance architectures compared	7	PULSE plus six materially different comparators
Value systems used for scoring	6	PULSE-aligned plus five non-PULSE moral priorities

2. Targeted model families

Presidential counting methods and cycle resolution

Clone, burial, no-show and turnout attacks

Cross-constituency legitimacy floors

Founding and mature Senate thresholds

Chief Justice confirmation

Permanent House districts and malapportionment triggers

Personhood-panel thresholds and correlated errors

Continuity guardian quorum/correlation

Cognitive-emergency forensic thresholds

Cognitive-change drift and reauthorization

Rogue-cohort pacing and split-provenance attacks

30% illicit-origin cap sensitivity

Anti-supremacy organization procedure

Emergency architectures

Redundancy/failure-domain models

Species restoration and archive resilience

Cataclysmic governance/trusteeship return

Extraordinary presidential certification

Functional Integration capability/network/hierarchy models

Long-lived wealth and persistent-return models

Least Reversible Harm hazardous-role allocation

Compound constitutional attacks

Interplanetary dual-legitimacy cases

3. Major validation-driven changes already made

Area	Testing consequence
President	Rank-sum/Borda abandoned after ~41.55% exact-clone winner-change behavior in the core overnight test; pairwise Condorcet architecture retained with Schulze/Beatpath working default.
Personhood panels	Near-unanimity rejected because false negatives became severe; 12/15 with multiple independent panels retained.
Senate	Ordinary law moved to 55% overall + 40% each delegation to reduce permanent species-veto politics while preserving cross-constituency support.
Chief Justice	60/45 retained after threshold comparison.
Rogue cohorts	Per-cohort/linked-source ceilings replaced by a 30% aggregate illicit-origin ceiling after split-provenance attacks exposed evasion.
Cognitive changes	Later reauthorization/continuing review added after cumulative-drift modeling.
House	Symmetric over- and underrepresentation alarms added after concentration stress.
Resilience	Independent failure domains elevated over raw site count.
Elections	Compulsory ballot return rejected; extraordinary impaired-election certification became the legitimacy backstop.
Functional Integration	Elevated to doctrine/protected architecture because parallel-society risk remained central.

4. Latest 300-year scenario program

The long-horizon program compared PULSE with six alternative architectures across twenty baseline world archetypes, six compound extremes and six value systems, with doctrine ablations, a PULSE-benefit skeptic sweep, capacity/homophily/private-concentration sweeps and a structurally different reduced-form hazard model. It is a conditional systems model, not an independent validation experiment: architecture-effectiveness parameters are explicit assumptions, and PULSE is assigned high values on capabilities its institutions are designed to deliver. The model asks what follows if those mechanisms work approximately as intended and where that conditional architecture breaks.

Result	Model output	Meaning
Scenario × value ranking	PULSE first in 114/120 baseline comparisons; fourth in 6/120. The six losses are the same weak-institutions scenario scored under six value-weight schemes. [Editorial model context] These results assume PULSE's institutions work well; the model doesn't independently establish that. In the skeptic test, moving selected PULSE institutional effects roughly 15–20% closer to the model's neutral setting removed its average lead over the strongest liberal-democratic comparator. This does not mean reducing every PULSE assumption by that amount. It is not an empirical confidence interval, a measured safety margin or evidence that PULSE is 15–20% better.	All six losses were the same deliberately weak-institutions world viewed through six value systems.
Raw catastrophic-failure avoidance	PULSE 62.7%; liberal comparator 63.1%	The liberal comparator slightly leads on constitutional survival and the no-catastrophic-failure composite. These are distinct measures; neither simply means species survival.
Free + integrated outcome	PULSE 56.8%; liberal comparator 28.5%	PULSE’s modeled advantage is preserving liberty/integration/anti-domination together.
Functional Integration ablation	~57% to ~17% free/integrated outcome	Functional Integration is structural in the model, but the ablation combines capability, continuity, and cohesion effects; it should not be presented as proof of social engineering.
Skeptic sweep	Aggregate advantage disappeared after moving selected PULSE institutional-effect coefficients ~15–20% toward neutral 0.5	PULSE mechanisms cannot fail badly and still automatically win.
Extreme hostility	All architectures degrade badly under sufficiently intense cross-species hatred/separation	Constitutional design can create useful opportunities, aligned incentives, and shared competence; it cannot and should not manufacture private goodwill through coercion.
Weak institutions	Simpler liberal architecture can outperform mature PULSE	Mature PULSE is institutionally demanding; transition/staging deserves separate work.

5. Safe public claims

Targeted mechanism testing covers 48 model/stress-test families; counts belong in methods/provenance, not as a headline proof of correctness.

Targeted simulations used more than 43 million independently randomized base scenarios/trials across the archived program. This is a workload/provenance fact, not a measure of evidentiary strength by itself.

A separate 300-year scenario program explored conditional civilization outcomes across competing architectures, followed by a reduced-form cross-check. Its proper use is dependency and failure-mode analysis, not empirical forecasting.

In the baseline long-horizon parameterization, PULSE ranked first in 114 of 120 scenario × value-system comparisons; all six losses were the same weak-institutions world. This should never be presented without the parameterization caveat: architecture-effectiveness values are model assumptions, and moving selected PULSE institutional-effect coefficients 15–20% closer to the model's neutral value of 0.5 eliminated its aggregate advantage over the strongest liberal-democratic comparator.

[Editorial model context] These results assume PULSE's institutions work well; the model doesn't independently establish that. In the skeptic test, moving selected PULSE institutional effects roughly 15–20% closer to the model's neutral setting removed its average lead over the strongest liberal-democratic comparator. This does not mean reducing every PULSE assumption by that amount. It is not an empirical confidence interval, a measured safety margin or evidence that PULSE is 15–20% better.

The testing did not prove PULSE will work. Targeted tests identify narrower mechanism properties; long-horizon models identify conditional dependencies, tradeoffs and failure modes.

6. Avoid misleading claims

Do not say "200 million civilizations were simulated." Five-year transitions, comparative evaluations, base worlds and trajectories are different units.

Do not call model outputs empirical forecasts or real-world probabilities.

Do not claim human intergroup-contact research proves human-synth social cohesion will work; it is only an empirical analogue for a secondary benefit. The primary capability and continuity rationales do not depend on that social effect.

Do not say PULSE is proven to be the best framework on the market.

Do not imply current AI has been determined conscious or entitled to PULSE personhood.

7. What held in the final batch

Five foundational doctrines remain intact.

House/Senate/shared-executive architecture remains intact.

Equal personhood + material relevance remains intact.

Functional Integration remains justified, but the testing should be read in the correct order: the doctrine primarily combines complementary capabilities and distributes civilizational competence; the modeled social-cohesion benefit is an important secondary mechanism.

Mutual Preservation and failure-domain resilience remain justified.

The major new warning concerns implementation capacity, not a contradiction in the mature architecture.

8. Evidence hierarchy

Category A — Targeted mechanism tests: narrow rule-level questions with explicit assumptions. Use these to compare election rules, thresholds, quorum structures, demographic caps, failure-domain designs and similar mechanisms.

Category B — Conditional civilization models: integrated 300-year scenario models in which institutional-effect parameters are hypotheses. Use these to ask what follows if a package works approximately as designed, where it loses, and which dependencies matter most.

Category C — External empirical analogues: human evidence on intergroup contact, wealth concentration, constitutional design, infrastructure resilience, demography and related subjects. These constrain directions and plausibility; they do not directly predict human-synthetic society.

Simulation count is never a substitute for assumption quality. Public-facing pages should lead with questions tested, failures found, design changes and limitations—not total trial volume.

9. Development method and provenance

Project Enduring Pulse was developed through human-directed systems design with AI-assisted drafting, literature synthesis, adversarial review, consistency checking and simulation/code generation. The steward made the substantive normative and architectural decisions and repeatedly accepted, rejected or revised model suggestions. Internal frontier-model critique is useful red teaming, but it is not external peer review.