

117-293-D Proposal for Project Meridian

Cognition Stabilization

Stability Engineering Institute | 1 October 2026

The proposal

The Stability Engineering Institute (SEI) invites Project Meridian to examine cognition stabilization as a strategic protection requirement for AI-dependent national capabilities. We propose a bounded evaluation of GUDIYA, an architecture intended to preserve essential functions and human governability when interacting adaptive systems experience accidental or deliberately induced disturbance.

Project Meridian has been announced as an effort to examine future warfare and the capabilities it will require. [1] SEI offers a contribution to that examination: the strategic value of advanced intelligence depends on the continued stability of the ecosystem through which it acts. Capability leadership should therefore include the engineering needed to sustain that capability under challenge.

Throughout this proposal,

- Super Intelligence denotes the broader national undertaking to develop and employ advanced AI. It does not assert that every deployed model exceeds human intelligence.
- Cognition stabilization concerns systems in which human and artificial cognition guide consequential action; it is not a psychological diagnosis or a program to control public beliefs.

One strategic investment with distinct responsibilities

When advanced AI becomes a national crown jewel, safeguarding its enduring usefulness belongs within the investment that creates it. Models, compute and research generate capability. Security, reliable operation and ecosystemic stabilization help preserve the conditions under which that capability delivers value. These responsibilities require coordinated funding, architecture and assurance from inception.

The Manhattan Project(WW2) provides a limited historical analogy. Its security apparatus controlled access and compartmentalized knowledge, while worker protection helped establish health physics and radiation monitoring. [2, 3] Creating a powerful capability brought parallel protection responsibilities. The lesson is integration of purpose, not a claim that wartime safeguards were complete or that AI and nuclear technologies have identical hazards.

SEI proposes one integrated national investment thesis: create advanced capability and preserve its dependable use. This does not mean one technical mechanism or one expenditure accomplishes both. The ecosystem may lose essential function even while its models remain secure and intact. Protecting value therefore extends to the services, institutions and people that depend on deployed intelligence.

The decision requested

We ask Project Meridian to include cognitive instability (exogenous or endogenous) in its threat examination and sponsor an independently evaluated, non-weaponized demonstration of ecosystemic stabilization. The proposed outcome is evidence about where existing controls suffice, where additional functions help, and what observation, authority and recovery conditions are required. GUDIYA is a proposal to be tested, not a demonstrated guarantee of national stability.

The asymmetric threat hypothesis

A competitor may seek to disrupt the operation of an AI-dependent ecosystem without matching its frontier models, compute capacity or research investment. The hypothesis is that a disturbance can be amplified by the target ecosystem's own authorized responses. The initiating action and the resulting disruption need not be comparable in scale.

This is the proposed “jijitsu” mechanism: strengths such as speed, interconnection, delegation and adaptation can become channels of amplification under adverse conditions. The same strengths can also improve defense. Exposure depends on architecture, deployment, shared dependencies and recovery margins; technological leadership alone does not establish vulnerability.

Source of capability	Potential stability exposure
Machine-speed activity	Consequences can develop before observation and intervention are effective.
Extensive interconnection	Disturbances can propagate through shared services and dependencies.
Autonomous adaptation	Changed strategies can invalidate earlier coordination assumptions.
Common models and services	Participants can respond similarly or encounter a common failure.
High resource utilization	Limited reserves can reduce the capacity to absorb disruption.
Distributed delegation	Work and obligations can persist beyond the initiating actant.

Cognitive Stability Warfare

SEI proposes this term for deliberate attempts to impair essential functions by inducing or amplifying instability in coupled human–AI systems, including through their own authorized adaptive responses. It identifies a possible threat mechanism; it does not assert the recognition of a separate military domain or establish that such attacks are inexpensive or reliably successful.

A hypothetical logistics ecosystem illustrates the concern. A service slowdown causes independently authorized planners to retry requests and select alternate providers. Their combined responses concentrate demand, delay confirmations and create incompatible commitments. The resulting disruption could arise accidentally or be deliberately provoked. Stabilization must address the developing condition even when attribution remains uncertain.

A lesson from asymmetric threats

The historical lesson is that dominance in one form of capability does not remove exposure to a different mechanism of disruption. U.S. military analysis described the September 11 attacks as asymmetric. [4] The parallel here concerns disproportionate effects and mismatched defenses; it does not equate a simulated service cascade with terrorism or claim that GUDIYA could have prevented historical attacks.

Recognizing cognitive instability

Cognitive instability is a condition in which actant activity, adaptation and interaction generate deviations that the surrounding system cannot adequately contain or correct, eroding operating margins, coherence and governability. Cognitive instability may have exogenous or endogenous origin. An actant is a participant capable of consequential action, including an AI agent, a human or a software service.

Cognitive stability is the maintained capacity to prevent, contain and recover from such developments before they cause unacceptable loss of essential function or governability. It permits useful adaptation, disagreement, local error and bounded emergence. The following manifestations overlap; they are not an inevitable sequence or a complete diagnostic taxonomy.

Manifestation	Observable concern
Loss of situational coherence	Incompatible or stale shared-state estimates produce consequential conflicts.
AI Vertigo	An actant mistakes movement toward danger for progress, or its corrective actions worsen its operational position. It can also be understood as widescale situational disorientation for actants.
Loss of temporal coordination	Measurements, decisions or controls arrive after the useful intervention window.
Uncontrolled adaptation	Changed plans, dependencies or delegation invalidate assumptions without adequate reassessment.
Commitment inconsistency	Individually permitted actions create incompatible collective obligations.
Resource overcommitment	Aggregate demand or reservations exceed capacity and available reserves.
Amplifying feedback	Retries or corrective responses intensify the original disturbance.
Oscillation and herding	Repeated reversals or correlated choices prevent acceptable operation.
Coupling-driven propagation	Disturbance spreads through direct interactions or indirect shared dependencies.
Uncontained emergence	Collective effects exceed acceptable limits despite controls on individual actions.
Loss of accountability	Delegated or migrated activity escapes effective supervision or control.
Exhausted recovery capacity	Buffers, fallback capacity or intervention time are depleted.
S4 collapse	Essential collective functions are lost and effective stabilization cannot restore acceptable operation within the required window.

AI Vertigo is an engineering analogy for loss of operational orientation, not a claim of subjective dizziness. Confidence, authorization and internal consistency do not establish correspondence with independently assessed conditions. S4 is defined relative to a declared system boundary, essential functions and recovery window. A protective shutdown that preserves those requirements may represent successful stabilization.

What Is An S4 Collapse?

In GUDIYA's proposed terminology, S4 collapse is the loss of essential system functioning accompanied by the inability of available stabilizing controls to restore acceptable operation within the required recovery window. It describes a system-level condition, not merely a malfunctioning component.

From a complex-systems perspective, collapse can develop when disturbances propagate through dependencies, corrective responses amplify one another, and recovery resources become exhausted. Local problems become collective: functioning components can no longer combine to deliver essential services. Every participant need not fail, and no malicious intent is required. Individually reasonable actions may collectively deepen the disruption.

A defining feature is loss of effective governability. Measurements may still arrive and commands may still be issued, yet interventions no longer produce sufficient correction in time. Delays, resource depletion, incompatible commitments or continuing feedback can overwhelm the system's remaining capacity to recover. This does not necessarily mean permanent destruction; restoration may require external assistance, substantial reconfiguration or rebuilding.

Uncontrolled Cascade

An uncontrolled cascade occurs when an initial disturbance triggers further disturbances across interacting components or actants, and that propagation exceeds the system's effective capacity to contain it. Each affected participant may transmit or amplify disruption through shared resources, dependencies, delegated work or corrective responses. Actions intended to restore local operation—such as retries, rerouting or resource reallocation—may inadvertently worsen conditions elsewhere.

A cascade can involve participants that remain individually functional and compliant. Its defining feature is the spreading loss of acceptable collective operation. It may be arrested through isolation, pacing, coordinated intervention or controlled degradation. If it exhausts recovery capacity and defeats effective stabilization, it can culminate in S4 collapse; that outcome is not inevitable.

S4 must always be assessed against a declared system boundary, essential functions, acceptable operating limits and recovery requirements. Collapse of one workflow is not automatically collapse of an enterprise or a national ecosystem. Conversely, a deliberate shutdown or controlled reduction in service can demonstrate successful stabilization if it contains propagation and preserves a viable recovery path.

Cognition Stabilization seeks to preserve that recovery path—detecting the erosion of operating margins and intervening before disturbance becomes loss of essential function and effective control.

The proposed GUDIYA contribution

GUDIYA proposes a Grid that relates individual trajectories to direct and indirect interactions, shared resources, cumulative commitments and environmental conditions. Its purpose is to coordinate fast local stabilization with wider observation under human-defined policy. The Grid need not be one centralized observer; any federation must specify its boundaries, information-sharing rules and allocation of authority.

Engineering responsibility	Question to answer
Secure execution	Are identities, access and operations valid under the relevant security policy?
Longitudinal assessment	Does an evolving actant remain within its mandate across its activity and delegation?
Ecosystemic stabilization	Do coupled activities preserve acceptable collective conditions and recovery margins?

Stabilization across the three layered AI economy

At the agentic layer, controls can bound delegation, concurrency and rates of commitment. At the digital services layer, they can coordinate queues, reservations and service demand. At the resource layer, they must respect the capacity and recovery limits of compute, energy, physical assets and other constrained resources. Circuit breakers at these layers need coordinated triggers and restart conditions so that protection at one layer does not simply transfer disruption elsewhere.

Settlement velocity is therefore an operating variable, not an unconditional objective. In a bounded deployment, authorized pacing may buy time for assessment and preserve capacity. Its costs, including delayed legitimate activity and possible concentration of queued demand, must be measured. The goal is useful operation within the envelope, including support for beneficial emergence.

Maintaining composure and learning from disturbance

The practical objective of cognition stabilization is composure under disturbance: essential functions and effective authority remain available even while parts of the ecosystem become unstable. Composure can include slowing activity, isolating a dependency, suspending new commitments or moving into a controlled degraded mode. It does not require every component to continue at full performance.

Time is part of the operating envelope

A complete control loop must observe relevant conditions, estimate state and uncertainty, select an authorized action, enforce it and verify its effect. Adequacy depends on the fastest consequential dynamics within the control scope and the remaining time before unacceptable consequences become unavoidable.

Nyquist sampling theory offers a conditional warning about aliasing: for ideal reconstruction of a band-limited signal, sampling must exceed twice its highest frequency. Adaptive workflows have no single universal behavioral frequency. Twice a model's token rate is therefore not a general safety criterion. The engineering task is to identify relevant variables and establish sufficient sampling, coverage and response deadlines for them.

Continuous monitoring does not by itself demonstrate timely control. Sampling interval, telemetry age, decision delay and actuation delay must be measured separately. Holding a commitment before execution may create time for assessment, while stopping future model calls may leave already delegated work active. Faster observation cannot recover evidence from an unobserved dependency.

Acquired antifragility as a testable outcome

Surviving a disturbance demonstrates resilience. Becoming measurably better because of lessons from that disturbance supports the stronger claim of acquired antifragility. The Grid should capture disturbance signatures and mechanisms without prematurely attributing them to an attacker. Similar symptoms can result from hostile action, ordinary failure or benign adaptation.

A proposed learning cycle records conditions and intervention effects, investigates the mechanism, tests candidate improvements, and promotes only approved changes. A shadow environment can replay incidents and challenge revisions against unseen scenarios while the operational Grid preserves service. Shadow evaluation does not establish a guarantee, but it separates urgent stabilization from experimental changes to control policy.

Signature recognition alone is insufficient. Improvements must generalize beyond replay of a familiar event, avoid normalizing harmful drift and retain performance under different operating regimes. Promotion requires evidence of reduced disruption without unacceptable false interventions or lost utility.

The stabilizer also requires protection

GUDIYA can introduce its own failure modes: erroneous shared estimates, excessive intervention, common dependencies or loss of control services. The design must address independent reference, authenticated telemetry, least-privilege authority, controller redundancy, auditability and tested degraded behavior. Human authorities define mandates and revise them; they need not approve every machine-speed action. Operational delegation must remain bounded, attributable and revocable.

A bounded evaluation for Project Meridian

SEI requests a technical review followed by a non-weaponized demonstrator using synthetic multi-agent logistics or shared-service workflows. The experiment should compare a credible baseline of security, ordinary reliability controls and actant supervision with the same system augmented by explicitly specified ecosystemic stabilization. Equal resources and representative workloads are needed to make the comparison meaningful.

The test environment should include delegation, changing membership, scarce shared resources and delayed effects. Controlled disturbances can include service slowdown, stale telemetry, correlated retries, communications loss and workload shifts. Tests should assess consequences and recovery without depending on reliable early attribution of malicious intent.

Evaluation dimension	Evidence to collect
Essential function	Required service maintained during disturbance and controlled degradation.
Propagation	Extent and duration of disturbance across participants and dependencies.
Temporal adequacy	Worst-case delay from usable evidence to verified intervention effect.
Control coverage	Treatment of delegated, queued, migrated and already committed work.
Utility and burden	Useful throughput, operating cost, false interventions and unnecessary restrictions.
Recovery	Time and resources needed to regain an approved operating condition.
Learning	Improvement on unseen disturbances without regression in other regimes.

Before testing, evaluators should declare the system boundary, essential functions, permitted loss, recovery windows and S4 criteria. Acceptance thresholds should be agreed in advance rather than selected after results are known. Independent reviewers should examine failures as well as successes and compare ecosystemic controls with simpler mitigations.

A useful result may show that ordinary controls suffice for some deployments, that GUDIYA functions improve others, or that an operating regime cannot yet be supported safely. Each outcome informs investment. The requested deliverables are a threat model, an interface and authority specification, a reproducible demonstrator, and a report of measured benefits, costs and unresolved limitations.

SEI contribution and proposed next step

SEI offers the GUDIYA Cognitome as a body of concepts, hypotheses and architectural reasoning for this examination. A scoping workshop could select one representative workflow and define the baseline, evidence requirements and independent evaluation arrangements. This proposal does not assume procurement approval or operational readiness.

The strategic objective is to protect the enduring value of the national Super Intelligence undertaking. Capability development and cognition stabilization should be planned as coordinated responsibilities within that investment. SEI asks Project Meridian to examine whether this infrastructure can preserve useful capability and human governability when adaptive ecosystems are disturbed.

Selected sources

- [1] Department of War. Secretary Hegseth Announces Six Major Initiatives at Quantico Speech. September 2026. <https://www.war.gov/News/Releases/Release/Article/4615424/secretary-hegseth-announces-six-major-initiatives-at-quantico-speech/>
- [2] National Park Service. Manhattan Project National Historical Park FAQs. <https://www.nps.gov/mapr/faqs.htm>
- [3] Department of Energy. The Manhattan Project A New and Secret World of Human Experimentation. https://ehss.energy.gov/ohre/roadmap/achre/intro_3.html
- [4] U.S. Army Military Review. Attack on America The First War of the 21st Century. <https://www.armyupress.army.mil/Journals/Military-Review/English-Edition-Archives/January-February-2022/Shaghnessy-Attack-America-2001/>
- [5] USSOCOM. USSOCOM Marks 25th Anniversary. <https://www.socom.mil/Pages/USSOCOMmarks25thanniversary.aspx>