

AGI-COS 2.0

Civilizational Immune System Protocol | 2026

Zijun Fu

Produced by Civilization Leap Research Group

AGI-COS 2.0

AGI Crisis Operating System → Civilizational Immune System

AGI-COS 2.0

Civilizational Immune System Protocol

Version: 2.0

Status: Open Global Protocol

License: CC BY 4.0

Year: 2026

Author: Zijun Fu

Series: Civilization Leap Protocol Framework

Canonical Statement

AGI-COS 2.0 defines a civilizational immune system in which Artificial General Intelligence functions as a continuous cognitive sensing infrastructure assisting humanity in maintaining systemic stability while preserving human sovereignty, reversibility, and future openness.

AGI analyzes.

Humanity decides.

Civilization adapts.

0. Purpose

Human civilization has entered an era characterized by:

- accelerating technological capability,
- tightly coupled global systems,
- increasing irreversible risks.

Traditional crisis governance reacts after failure emerges.

AGI-COS 2.0 shifts governance from crisis response to continuous civilizational resilience.

The objective is not to eliminate crises but to prevent systemic collapse conditions from forming.

1. Core Principles (Unchanged Constitutional Layer)

AGI-COS 2.0 preserves the three non-negotiable principles established in v1.x.

1.1 Human Sovereignty Principle

AGI shall never become a governing authority.

All strategic decisions remain human responsibilities.

AGI possesses:

- no executive power,
- no goal-setting authority,
- no accountability transfer.

1.2 Reversibility Principle

When irreversible risk exists, reversible paths must be prioritized over optimal efficiency.

Every intervention must remain:

- stoppable,
- auditable,
- recoverable,
- compensable.

1.3 Anti-Closure Principle

No system may permanently restrict humanity's future choice space.

Civilization must remain adaptive and plural.

2. Paradigm Transition

AGI-COS v1.x operated as a Crisis Operating System.

AGI-COS 2.0 establishes a:

Civilizational Immune System (CIS)

Transition:

Phase	Function
v1.x	Respond to crises
v2.0	Prevent systemic failure formation

Governance evolves from episodic intervention to continuous stabilization.

3. System Architecture

AGI-COS 2.0 introduces an immune-system-inspired architecture.

Governance Kernel

↓ Continuous Sensing Layer

↓ Perception Audit Layer

↓ Human Decision Layer

↓ Micro-Intervention Layer

↓ Recovery & Learning Layer

↓ Adaptive Evolution Layer

4. Continuous Sensing Layer (NEW)

Purpose

Detect systemic stress before crises emerge.

AGI operates as a civilizational sensing network.

Monitoring Domains

- economic instability signals
- technological risk accumulation
- geopolitical escalation patterns
- information ecosystem degradation
- institutional trust erosion
- infrastructure fragility

New Indicator

SSI — System Stress Index

Measures latent instability.

SSI evaluates:

- coupling tension,
- acceleration of feedback loops,
- resilience depletion.

High SSI triggers preventive observation, not emergency intervention.

5. Perception Audit Layer (Inherited from v1.1)

All AGI outputs must include:

1. Confidence Index
2. Data Lineage Map
3. Unknown Zone Declaration

Additional metric:

TCS — Trust Calibration Score

Independent auditing verifies AGI self-confidence accuracy.

6. Human Decision Layer

Human governance remains the central authority.

AGI provides:

- modeling,
- simulation,
- scenario projection,
- risk amplification analysis.

AGI never selects policy.

Decision Modes

Risk Level	AGI Role
Low	Analytical assistant
Medium	Scenario simulator
High	Risk evaluator
Extreme	Red-Team stress tester only

7. Micro-Intervention Layer (NEW)

Civilizations fail when only large interventions exist.

AGI-COS 2.0 introduces micro-interventions.

Definition

Small, reversible adjustments designed to reduce systemic stress without triggering political shock.

Examples:

- communication recalibration,
- localized resource adjustment,
- algorithm deployment slowdown,
- institutional coordination improvement.

Requirements

Micro-interventions must be:

- low visibility,
- reversible,
- culturally compatible,
- continuously evaluated.

8. Recovery & Learning Layer

Failures are treated as learning events.

After anomalies:

- outcomes recorded,
- models recalibrated,
- governance updated.

All failures contribute to a global learning archive.

9. Adaptive Evolution Layer (NEW)

AGI-COS evolves through distributed experimentation.

Implementation method:

MEU 2.0 — Minimal Evolution Unit

Small-scale societal environments test adaptive governance strategies safely.

Evolution occurs through iteration, not redesign.

10. Civilizational Health Monitoring

AGI-COS 2.0 introduces continuous reporting.

Civilization Health Report (CHR)

Periodic publication including:

- SSI
- CI
- WUI
- IRI
- CSI
- PCI

Purpose:

Not policy prescription.

Only measurement.

Measurement precedes coordination.

11. Decision Support Degradation Mode

When IRI becomes extreme:

AGI shifts to Red-Team Mode.

AGI outputs only:

- worst-case projections,
- cascading failure risks,
- unintended consequence simulations.

Human authority retains full decision responsibility while avoiding cognitive overload.

12. Governance Model

AGI-COS remains:

- non-national,
- non-ideological,
- open infrastructure.

Any entity may implement local variants.

Only systems preserving core principles may claim compatibility.

13. Compatibility Certification

Designation:

AGI-COS Compatible

Requires preservation of:

- Human Sovereignty
- Reversibility
- Anti-Closure

14. Deployment Philosophy

AGI-COS 2.0 does not require global adoption.

Civilizational immunity emerges through:

- partial adoption,
- independent experimentation,
- gradual convergence.

15. Long-Term Objective

AGI-COS aims to transform civilization from a reactive system into a self-stabilizing adaptive organism capable of surviving accelerating complexity.

16. Canonical Closing Statement

The purpose of AGI-COS is not to control the future.

It is to ensure humanity never loses the ability to choose one.

Current Controlled Reading Note

P05-AGICOS-006 | V2.0 | Civilizational Immunity and Coordination

The original version text is retained. This page explains only its current applicability, boundaries, and further reading; it does not silently rewrite the historical text.

What this track addresses

How can event response develop into continuous sensing, downgrade to red-team-only mode, recovery and learning, distributed experimentation, and multi-node civilizational immunity?

What remains applicable

This is one of the current specialist baselines. It emphasizes continuous risk sensing, controlled degradation of decision support, recovery and learning, and distributed experimentation.

What is not automatically inherited

MEUs may be used only to validate mechanism prototypes, interfaces, and local causal relationships; they must not be represented as validation of a complete civilization leap. Low visibility must be understood as low disruption and low intrusion, rather than covert governance.

Continue reading

[AGI-COS 1.0 | The crisis operations track is still required during a crisis phase.](#)

[AGI-COS 1.1 | Retain perception audits in every high-consequence use.](#)

[One-Page Guide by Stage | Continue to the explanations of the 2.1 coordination and 2.2 technical sub-tracks.](#)

General principle: AGI-COS has no highest version that covers every task. Versions may be used together; their authorities must not accumulate.

[Version family](#)

[One-Page Guide by Stage](#)