

AGI-COS 2.1 Full Edition (Canonical Operational Edition).

This version is positioned as:

The standard text marking AGI-COS's formal transition from a theoretical system to a "globally operational phase".

It is an engineering integration and upgrade of AGI-COS 1.x + 2.0,
for policy submission / international discussion / MEU pilots / crisis-intervention coordination.

AGI-COS 2.1

Global Civilizational Operating System for the AGI Era

Global Civilizational Operating System for the AGI Era (2.1 Full Edition)

Version: 2.1 Canonical

Status: Operational Framework

Nature: Coordination Infrastructure

Principle:

AGI analyzes · Humans decide · Civilization adapts

PART I — Era Context and Objectives

1. The Fundamental Problem of the AGI Era

With the emergence of AGI, humanity enters, for the first time, a situation in which:

- Decision complexity exceeds human intuition
- Technological acceleration outpaces institutional response
- Risks are systemically coupled

The central threats to civilization are no longer isolated disasters, but:

- Formation of irreversible paths
- Automation lock-in
- Coordination failure
- Contraction of civilizational meaning

2. The Mission of AGI-COS 2.1

The objectives of AGI-COS 2.1:

1. Increase global coordination speed
2. Extend civilization's reversible time
3. Prevent systemic loss of control
4. Establish a civilization-level immune structure

PART II — Core Principles (Non-Negotiable Principles)

Principle 1 — Human Sovereignty

AGI never holds decision-making authority.

AGI is responsible for:

- Analysis
- Simulation
- Risk assessment

Humans are responsible for:

- Value judgments
- Policy choices
- Bearing responsibility

Principle 2 — Reversibility First

All interventions must be:

- Pausable
- Exitable
- Reversible through rollback

Irreversible designs are considered high-risk states.

Principle 3 — Anti-Closure

Civilization must not be locked into a single optimal solution.

The following must be allowed:

- Development along multiple paths
- Institutional competition
- Social experimentation

Principle 4 — Minimal Intervention

Intervene only when necessary.

The system's goal is not to control society, but to maintain a stable space.

Principle 5 — Non-Governance AGI

AGI-COS is not a government.

It is:

Cognitive and coordination infrastructure

PART III — AGI-COS 2.1 Core Innovation Modules

MODULE A — First-Mover Safety Architecture

(First-Mover Safety Architecture)

Problem

States will not be the first to accept constraints.

Mechanism

Simultaneous Activation Protocol

Process:

- 1. Internal approval by each party**
- 2. Deposit documents with a neutral institution**
- 3. Reach the participation threshold**
- 4. Simultaneous global activation**

Effects:

- Eliminate strategic disadvantage
- Resolve the prisoner's dilemma

MODULE B — Strategic Ambiguity Layer

(Strategic Ambiguity Layer)

AGI-COS is externally defined as:
Global AGI Stability Infrastructure
rather than a control system.

Purpose:

Avoid sovereignty-defense reactions.

MODULE C — Dual-Track Adoption Model

(Dual-Track Adoption Mechanism)

Track 1 — Industry Track

AI laboratories first adopt:

- Risk indicators
- Circuit-breaker interfaces
- Audit language

Track 2 — Government Track

Governments subsequently institutionalize it.

MODULE D — Non-Threat Sovereignty Protocol

(Non-Threat Sovereignty Protocol)

Guarantees:

- National decision-making authority remains intact
- No supranational power
- No compulsory enforcement body

AGI-COS provides coordination capability only.

MODULE E — Civilizational Time-Buying Mechanism

(Civilizational Time Mechanism)

New core indicator:

CTI — Civilizational Time Index

Measures:

Whether more future choices are secured for humanity.

PART IV — AGI-COS Seven-Layer Operating Architecture

Layer 1 — Continuous Sensing

Continuous Civilizational Sensing Layer

Monitoring indicators:

- SSI — System Stress Index
- CI — Coupling Index
- IRI — Irreversible Risk Index
- TCS — Trust Calibration Score
- MOS — Meaning Option Space
- CTI — Civilizational Time Index

Layer 2 — Perception Audit

Perception Audit Layer

Requirements:

- Confidence labeling
- Data provenance
- Statements of unknowns

- Adversarial verification

AGI must be auditable.

Layer 3 — Micro-Intervention Layer

Micro-Intervention Layer

Characteristics:

- Small-scale
- Low-visibility
- Reversible through rollback

Avoid the "risks of major reform".

Layer 4 — Crisis Operating System

Crisis Operating Layer

Trigger condition:

IRI or SSI reaches the threshold.

AGI provides:

- Scenario simulation
- Coordination recommendations

Humans make the final decision.

Layer 5 — Decision Downgrade Mode

(Red-Team Mode)

Under extreme risk:

AGI stops recommending actions and provides only:

- Worst-case scenarios
- Assumption stress tests

Prevent dependence on automation.

Layer 6 — Adaptive Evolution Layer

MEU 2.1 Minimum Evolution Unit

Allows:

- Regional governance experiments
- Institutional pilots
- Innovation with an exit option

Civilization evolves through experimentation.

Layer 7 — Learning & Memory Layer

Civilizational Learning Layer

Establish:

Global Civilization Learning Archive

Preserve:

- Crisis experience
- Failure cases
- Governance models

Civilization avoids repeating mistakes.

PART V — Operating Protocols (Operational Protocols)

1. Crisis Attachment Protocol

The system is not actively promoted.
It is referenced during crises and adopted naturally.

2. Invisible Expansion Protocol

Principles of invisible diffusion:

- Permanently open to the public
- Multi-node dissemination
- No centralized control

3. Coordination Acceleration Protocol

Core objective:
Increase civilizational coordination speed.

PART VI — AGI-Human Relationship Model

AGI is:

- A cognitive amplifier
- An analytical partner
- A coordination tool

AGI is not:

- A ruler
- A government
- An arbiter of values

PART VII — Civilizational Health Model

Civilization stability function:
Civilization Stability =
Coordination Speed
× Reversible Time
× Trust Calibration

PART VIII — Alignment with the Threefold Goals of Civilization Leap

Civilizational Goal	AGI-COS Role
Worry-free survival	Prevent irreversible collapse
Happiness in life	Maintain a stable environment
Meaning in life	Preserve future possibilities

PART IX — The Historical Position of AGI-COS 2.1

AGI-COS 2.1 is not:

- An AI policy document
- A technology governance plan
- An idealistic declaration

It is:

Civilizational Coordination Infrastructure
(Civilizational Coordination Infrastructure)

Comparable to:

- The scientific method
- Constitutional principles
- Internet protocols

PART X — Final Declaration

The goal of AGI-COS 2.1 is not to create a perfect society.

Rather, it is:

In an era of accelerating intelligence,

to ensure that human civilization retains the ability to choose its future.

AGI analyzes.

Humans decide.

Civilization continues.