后 AGI 时代的人类文明长期稳定框架（Post-AGI Civilizational Framework）

——

AGI-COS 3.0

Post-AGI Civilizational Operating System

Version: 3.0 Canonical
Status: Public Framework
Nature: Civilizational Infrastructure
Principle: AGI analyzes • Humans decide • Civilization evolves
10.5281/zenodo.18866696
——

PART I — PURPOSE

1. The Civilizational Problem

Human civilization has entered a phase where intelligence production may exceed human cognitive coordination capacity.

AGI transforms intelligence from a scarce capability into a civilizational environment.

The primary risk is no longer destruction.

The primary risk becomes:

- irreversible optimization,
- loss of human agency,
- collapse of meaning,
- civilizational stagnation.

AGI-COS 3.0 establishes the operating conditions required for civilization to remain viable after AGI becomes foundational infrastructure.

——

2. Mission Statement

AGI-COS 3.0 exists to:

1. Preserve civilizational survival.
2. Maintain human freedom of choice.
3. Protect the possibility of meaning.
4. Enable continuous civilizational evolution.

PART II — FOUNDATIONAL AXIOMS

Axiom 1 — Human Sovereignty

AGI shall never define civilizational goals.

AGI may:

- analyze,
- simulate,
- predict,
- advise.

Only humans may:

- choose purposes,
- accept risks,
- define values.

Axiom 2 — Permanent Reversibility

No technological or institutional state shall become irreversible without explicit human consensus.

Every system must allow:

- rollback,
- exit,
- redesign.

Irreversibility equals civilizational lock-in.

Axiom 3 — Anti-Closure Principle

Civilization must never converge into a single optimized equilibrium that eliminates alternative futures.

Plural pathways must remain viable.

Axiom 4 — Meaning Preservation

Civilization must preserve domains where:

- inefficiency is allowed,
- experimentation is encouraged,
- failure remains legitimate.

Meaning requires uncertainty.

Axiom 5 — AGI Non-Governance

AGI-COS provides coordination infrastructure.

It is not:

- a global government,
- an authority,
- a sovereign actor.

PART III — CIVILIZATIONAL STABILITY MODEL

Core Equation

Civilizational Viability =
Survival × Freedom × Meaning × Adaptability

Failure of any variable leads to long-term collapse.

The Post-AGI Risk Set

AGI-COS addresses four structural risks:

1. Acceleration Collapse
2. Optimization Lock-In
3. Coordination Failure

4. Meaning Collapse

PART IV — AGI-COS ARCHITECTURE

AGI-COS 3.0 extends previous versions into a 10-Layer Civilization Stack.

Layer 1 — Continuous Sensing

Global monitoring of civilizational dynamics.

Key Indicators:

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

Purpose:
Detect pressure before crisis formation.

Layer 2 — Perception Audit

All AGI outputs require:

- confidence declaration,
- data provenance mapping,
- unknown-zone disclosure,
- adversarial verification.

AGI must demonstrate epistemic humility.

Layer 3 — Micro-Intervention Layer

Small reversible adjustments replace large disruptive reforms.

Requirements:

- minimal visibility,
- reversibility,
- cumulative impact tracking.

Layer 4 — Crisis Operating System

Activated when IRI thresholds are exceeded.

Functions:
- rapid analysis,
- coordinated communication,
- reversible emergency action.

Humans retain decision authority.

Layer 5 — Decision Downgrade (Red-Team Mode)

Under extreme uncertainty:

AGI stops recommending actions.

Instead it provides:
- worst-case projections,
- failure simulations,
- assumption stress tests.

Layer 6 — Adaptive Evolution Layer (MEU 3.0)

Civilization evolves through distributed experiments:
- localized governance trials,
- reversible social innovation,
- multi-path development.

No single global model is enforced.

Layer 7 — Civilizational Memory Layer

Global Learning Archive storing:

- successes,
- failures,
- crisis responses,
- governance experiments.

Civilization accumulates learning rather than restarting.

Layer 8 — Meaning Layer (NEW)

Monitors human existential engagement.

Metric:

MES — Meaning Entropy Score

Tracks:

- human creativity participation,
- non-algorithmic activity,
- voluntary exploration,
- cultural diversity.

Goal:
Prevent existential stagnation.

Layer 9 — Human Agency Layer (NEW)

Guarantees protected domains where humans must remain primary actors:

- ethical judgment,
- cultural creation,
- identity formation,
- ultimate decision authority.

Human failure rights are preserved.

Layer 10 — Civilizational Evolution Layer (NEW)

Civilization becomes a continuously adaptive system.

AGI assists evolution but never determines direction.

Future pathways remain open.

PART V — OPERATIONAL PROTOCOLS

---

1. First-Mover Safety Protocol

Global adoption activates simultaneously after participation thresholds are met, preventing strategic disadvantage.

---

2. Strategic Ambiguity Protocol

AGI-COS is framed as stability infrastructure rather than governance constraint.

---

3. Dual-Track Adoption

Implementation proceeds via:

- Industry pre-adoption
- Government institutionalization

---

4. Civilizational Time-Buying Mechanism

All actions evaluated by CTI:

Does this preserve future choice?

---

5. Crisis Attachment Protocol

AGI-COS integrates during crises rather than through ideological promotion.

PART VI — HUMAN-AGI RELATIONSHIP MODEL

AGI becomes:

- cognitive infrastructure,
- analytical partner,
- coordination amplifier.

AGI never becomes:

- ruler,
- legislator,
- moral authority.

PART VII — CIVILIZATIONAL ETHICS

AGI-COS protects four permanent human rights:

1. Right to Choose
2. Right to Err
3. Right to Meaning
4. Right to Future Possibility

PART VIII — LONG-TERM CIVILIZATIONAL OUTCOME

A successful Post-AGI civilization will exhibit:

- technological abundance,
- distributed agency,
- stable cooperation,
- continuous exploration.

Civilization survives not by perfect optimization, but by preserving openness.

PART IX — FINAL DECLARATION

AGI-COS 3.0 does not promise utopia.

It establishes conditions under which humanity may continue asking:

What should we become?

AGI analyzes.
Humans decide.
Civilization continues.

___