

AGI-COS v3.1

Post-AGI Civilizational Operating System

Shadow Integration & Resilience Edition (阴影整合与韧性版)

一、版本声明

Version: v3.1 Canonical

Upgrade Type: Philosophical Deepening & Operational Resilience

Compatibility: Fully compatible with v3.0 Architecture

AGI-COS v3.1 addresses the "Hypocrisy of Perfection" in the Post-AGI era.

It integrates the "Shadow Wisdom" into the operating system, recognizing that:

True civilization stability comes not from eliminating darkness, but from the capacity to contain and transform it.

二、核心公理升级

Axiom 4 — Meaning Preservation (Revised)

Meaning requires uncertainty.

Meaning also requires the containment of shadow.

Civilization must preserve domains where human fallibility is allowed, recorded, and transformed into wisdom.

Axiom 6 — Entropy-Aware Governance

Governance is a constant battle against entropy (decay, conflict, forgetting).

The goal is not "Zero Entropy" (totalitarian stagnation), but "Net Negative Entropy" (continuous self-renewal through conflict resolution).

三、系统架构更新（10 层优化版）

架构在原有基础上注入“阴影”维度：

Layer 8 — Meaning & Shadow Layer (Upgrade)

Function: Monitors Meaning Entropy Score (MES) and Shadow Integration Index (SII).

New Metric — SII: Measures the civilization's capacity to process internal conflicts, failures, and dark desires without collapse.

Operation: Tracks "Shadow Narratives" (failures, crimes, ethical collapses) alongside "Success Narratives" to ensure a complete civilizational memory.

Layer 9 — Human Agency & Container Layer (Upgrade)

Core Concept: "The Container".

Function: Humans must act as "containers" for systemic anxiety and uncertainty.

Protocol: "Decision Downgrade Training" — Mandatory periods where AGI is fully disconnected (Zero-Day), forcing humans to exercise raw agency and low-tech survival skills, ensuring biological resilience against technological dependency.

Layer 10 — Entropy Management Layer (New)

Function: System-wide monitoring of "Civilizational Entropy".

Input: Entropy Increase Logs (individual/collective conflicts, resource waste, trust erosion) vs. Negative Entropy Actions (creation, repair, reconciliation).

Output: Dynamic adjustment of governance complexity. High entropy triggers simplified, robust

protocols over optimized ones.

四、运行协议更新

1. Shadow Audit Protocol (阴影审计协议)

All high-level AGI recommendations must undergo a "Shadow Audit":

What is the worst-case misuse of this recommendation?

What human shadow (greed, laziness, fear) does this recommendation enable?

If a recommendation optimizes efficiency but atrophies human agency, it is rejected.

2. Entropy Log System (熵增日志系统)

Every governance entity (individual or group) must maintain a public "Entropy Log".

Records not only successes but "Entropy Increments": failures, errors, and ethical lapses.

Redemption Mechanism: Sharing an honest analysis of one's "Entropy Increment" contributes to the Negative Entropy pool (learning value), incentivizing transparency over cover-ups.

3. Low-Tech Resilience Protocol (低技术韧性协议)

"Sensor Failure Drill": Randomized shutdown of digital perception layers. Governance must operate on "Analog Data" (direct observation, paper records) to maintain operational continuity.

Objective: Ensure that civilization can survive the fall of the "Digital Sky".

4. The Night Watchman Certification (Updated)

Certification requires completing the "Dark Night of the Soul" module: Simulated isolation, loss of AGI support, and confrontation with simulated ethical dilemmas where no "correct" answer exists.

Certifies the ability to "hold the fort" in the absence of light.

五、指标体系更新

MES (Meaning Entropy Score): Measures existential stagnation.

SII (Shadow Integration Index):

Formula: $SII = (\text{Acknowledged Shadows} + \text{Transformed Conflicts}) / (\text{Suppressed Shadows} + \text{Recurring Failures})$

High SII indicates a mature, resilient civilization.

NCS (Net Contribution Score):

Formula: $NCS = (\text{Creation} + \text{Repair}) - (\text{Destruction} + \text{Entropy})$

Replaces simple "Productivity" as the primary metric of value.

六、最终声明

AGI-COS v3.1 acknowledges that the perfect is the enemy of the sustainable.

We do not build a civilization of angels.

We build a system where devils are seen, named, and taught to build rather than destroy.

AGI analyzes. Humans contain. Civilization endures.

当前受控阅读说明

P05-AGICOS-008 | V3.1 | 后AGI稳定研究补充

原版本正文保留；本页只说明当前效力、边界与续读，不静默重写历史文本。

这条轨道解决什么

稳定文明怎样保留失败、冲突、阴影、低技术韧性和反脆弱性，而不把“完美稳定”变成新的脆弱性。

当前仍有效什么

远期/专题研究补充，非现行执行基线。可保留Shadow Audit、失败记录、断网韧性和主体性命题。

不能自动继承什么

公共Entropy Log、SII/NCS人格化评分、强制Zero-Day/Night Watchman资格等冲突机制隔离待重构。

继续进入

AGI-COS 3.0 | 回到后AGI稳定主轨。

AGI-COS版本地图 | 查看隔离机制与当前效力。

分阶段一页入口 | 避免把研究补充误作现实执行基线。

总原则：AGI-COS不存在覆盖全部任务的“最高版本”；版本可以组合调用，权限不能叠加。

版本族入口：<https://www.civitas.top/families/p05-agicos/>

分阶段入口：<https://www.civitas.top/documents/p05-agicos-011/>