

AGI-COS 受控版本族

先选现实问题、阶段、相位和任务，再选版本；不要按最大版本号选稿

核心判定 | AGI-COS 没有一个覆盖全部文稿的“最高/最新版”。当前结构是：跨版本不变量 + 专业/研究轨 + 伴随限制/表达 + 现实任务派生。轨道可以组合调用，权限不能因组合而叠加。ACTIVE 也不等于自动启用。

你正在解决什么？

现实问题/阶段	优先入口	主要用途	不要自动继承
所有高后果判断	T0 + 1.1	跨版本护栏；来源、置信度、未知、低置信度理由；先把模型输出降为可审计建议。	任何风险分/模型结论直接变决定或权限。
危机／高不可逆	加 1.0 + 1.2	Detect→Diagnose→Window→Impact→Path→Execute→Recover→Learn；对抗数据、认知负荷、二阶反馈。	48 小时、影响百万人等历史固定阈值；心理说服/行为塑形。
当前/近程跨主体协调	2.0 + 2.1 + 1.1	持续风险感知、恢复学习、分布式试验、非主权协调、共享风险认知、可逆响应。	固定年份、席位、参与门槛；MEU=完整文明验证；隐秘微干预。
真实部署/接口原型	2.2 + 1.1 + 当前治理	身份、授权、审计、日落、验证节点、故障安全、隐私等候选接口。	<100ms、默认 72 小时、特定链/IPFS/Dead Man’ s Switch 等历史实现通用化。
稳定/后 AGI 长期设计	3.0 + 2.0 + LC-1.0	主体性、失败权、开放未来、文明记忆、系统退居；稳定后减少显性介入。	精确文明效用公式、意义分数、固定全球激活。
韧性/失败文化专题	3.1 研究轨	失败记录、Shadow Audit、低技术/断网韧性。	公共 Entropy Log、SII/NCS 治理评分、强制训练/资格。
星际现实条件出现后	4.0 / 4.1 研究轨	分布式协调、协议兼容、记忆、漂移与远距修复。	把多行星假设倒灌当前；AGI 伦理覆盖主权；按熵/贡献调治理权。
超长期哲学压力测试	Ω / Ω+	非终局、意义再生、未知开放、守护者递归与反僵化。	单一存续目的、Challenge Injection、制造危机、强制 Goal Obsolescence。

每次调用前必须回答 7 问

1 现实问题是什么？危机/常态风险/协调/部署/长期/星际/超长期	2 文明阶段、环境相位、任务过程态是什么？不要用版本号代替	3 真正需要哪种能力？感知/韧性/危机/免疫/协调/接口/长期主体性	4 谁有权决定？谁专业判断？谁执行？谁独立纠错？
5 打开什么数据/资源/域/权限？如何退出、回归、修复、终止？	6 这次引用的是专业本体、历史参数、政策表达还是研究情景？	7 是否触碰禁回退条款？若触碰，是新证据，还是只是实现方便？	最短合同：版本号≠权威；能力≠权限；输出≠决定；高后果必须可逆、可退出、可重授权

跨版本禁回退底线 | 能力不生成主权；指标不生成权力；高后果必须可暂停/撤回/退出/修复/终止/重授权；不得操纵主体；最小必要数据；危机权限到期撤回；三目标分别显影且不可兑换；系统成熟后应能退居；4.x/Ω 不倒灌当前现实。

当前受控入口：G-03《AGI-COS Controlled Version Map》V1.0 + 多版本连续性审计 V1.1 + G-04 限制与禁回退宪章。

个性化续读指引

不是“相关阅读”列表，而是从这篇文稿继续追问下一步问题。

你已经知道AGI-COS不能按“最大版本号”调用。下一步要把机器侧安全轨同时接回人的责任、最低开放未来边界与今天为何需要这种治理。

向责任

Human-COS 启动期：不可逆前判断与责任窗口

高后果判断同时需要人的责任与方向侧。

civitas.top/documents/p05-humancos-005/

向边界

共同窗口协议：开放未来与三约束

读取所有高能力系统都不应破坏的最低开放未来边界。

civitas.top/documents/p05-window-005/

向现实因果

AGI时代的人类文明选择

回到为什么这种版本治理在今天变得必要。

civitas.top/documents/p03-agichoice-001/

质疑、反例与修订入口

新的事实、反例、测试失败、新能力、新攻击面或阶段条件变化，都应进入相应文稿族再审；续读不能被用来维持既有结论。

完整文库 → civitas.top/library/

公众总入口 → civitas.top/documents/p01-entry-006/