

星云

星云人机决策与治理专项

智能自主管理条件下的人类最高决策权重、权利边界与责任结构

版本：V1.1 过渡与验证增强版 | 状态：专项母稿 / 过渡与验证增强

作者	子君赋
出品	文明跃迁研究组
版本	V1.1 过渡与验证增强版
状态	《星云总纲》配套专项
日期	2026 年 7 月

整体足以保障存续，却不吞没个体；资源能够高效流动，却不把人变成效率工具；生活能够幸福、有意思、有乐趣，生命能够不断涌现并持续承接新的意义。

目录

- 01 治理定位
- 02 人类最高决策权重
- 03 决策主体与权重来源
- 04 事项分级
- 05 三权分离
- 06 标准决策流程
- 07 紧急相位
- 08 AGI 自主等级
- 09 申诉、审计与责任
- 10 阶段演进
- 11 根本边界

一、治理定位

星云可以实现高度智能自主管理，但 AGI 不拥有文明主权。人类在文明存续、人格权、战争、不可逆技术和根本价值方向上拥有最高决策权重与最终否决权。

治理目标不是让人类对每一次技术操作逐项审批，而是在 AGI 能力不断提高时，仍然保证文明目标、权利边界和最终责任不会从真实主体手中无声转移。

二、人最高策重

“最高决策权重”意味着在人机判断冲突且事项涉及文明宪制、人格、生命、不可逆伤害和长期共同未来时，人类集体形成的合法决定拥有最终效力。它不意味着任何个体可以随意否决公共决策，也不意味着人类判断无需证据和现实检验。

三、策主重源

主体	权重来源	不得越界
受影响主体	权利、身体、生活和现实后果直接由其承担	不能以个人偏好取消他人基本权利和整体存续
专业者	知识、经验、资格和领域责任	不能垄断价值判断
现实承担者	执行、风险、长期维护和失败后果	不能以承担为名占有公共决定
真实贡献者	对具体子域能力的实际增厚	贡献不能购买人格和文明主权
公共代表与共同体	合法授权、公共责任和整体协调	必须接受公开质询、任期和退出
AGI 与智能体	计算、预测、模型、多方案推演和实时管理能力	不得单方面定义目标、权利和不可逆方向

四、事分

级别	事项	决策要求
L0 日常运行	库存、路线、设备、标准服务和低风险维护	AGI 可自主执行，持续日志和异常上报
L1 专业事务	医疗方案、工程设计、教育路径、资源调度	专业权重为主，受影响主体参与，AGI 辅助
L2 重大公共事务	城市规划、跨域资源、重大项目和长期预算	公开论证、多主体权重、申诉和复核
L3 文明底线事项	战争、人格权、强制干预、不可逆技术、AGI 最高权限	人类最高决策权重，严格否决与多重审议

五、三分

资源使用与调度权、专业领域决策权、文明价值与底线表决权必须相互分离。贡献可以影响具体子域参与权重，不能直接转化为资源所有权或文明表决权；专业能力可以提高判断权重，不能独占价值和人格决定。

六、准策流程

1. 问题定义：说明事实、目标、影响范围、不确定性和是否必须决策。
2. 主体识别：确认受影响者、专业者、执行者、资源提供者和责任承担者。
3. 权限核验：确认谁可以看见什么、谁可以提出方案、谁可以决定和执行。
4. 方案生成：AGI 与人类共同提出多种路径，标明代价、风险、受益者和不可逆部分。
5. 公开质询：允许受影响主体、专业共同体和公众提出异议。
6. 权重形成：综合权利、专业、责任、贡献和文明底线，而非单一投票。
7. 决定与责任锚定：明确谁作出决定、谁执行、谁承担失败和维护。
8. 执行、监测与熔断：实时观察现实，触发异常即暂停或回滚。
9. 复盘与修复：后果返回，处理伤害、赔偿、贡献、错误和制度升级。

七、急相位

战争、灾害、疫情、基础设施崩溃和失控智能等紧急情况允许临时提高系统执行速度、缩短协商和扩大必要权限，但必须满足事实门槛、最小必要、明确时限、分级授权、自动到期、人工复核、事后公开和责任追溯。

紧急权力必须比常态权力更容易失效，而不是更容易永久化。

八、AGI 自主等

等级	能力	边界
A0 建议	信息整理、提示和方案比较	不执行
A1 受控执行	执行明确、低风险、可逆任务	逐项授权
A2 领域自治	在稳定规则内自主计划和调度	异常即上报，保留人工接管
A3 跨域自治	多系统协同、动态资源配置和自我修复	文明边界、受影响权利和熔断机制固定
A4 高度自主管理	大规模实时管理与长期优化	不得独立决定 L3 事项，人类保留最高决策权重

九、申、任

- 受影响主体有权知道关键决定与自己相关的主要依据。
- 申诉不能只由原系统自动复核，必须有独立主体和人工通道。
- 算法、模型和复杂组织不能成为责任空洞。
- AGI 执行错误由授权、设计、运营和监督责任共同追溯。
- 决策记录应保护隐私，同时允许必要审计。
- 错误被发现后必须能暂停、回滚、赔偿和修复。

十、段演

阶段	治理特征
启动期	AGI 主要辅助，现有国家、组织和专业制度承担责任；建立身份、授权、日志和人工接管
过渡期	AGI 获得领域自治，三权分离、受影响主体参与和跨域仲裁逐步建立
稳定期	日常智能自主管理，人类聚焦文明目标、权利、不可逆事项和监督；最高决策权重保持

十一、根本边界

- 不得把人类最高决策权重变成少数人的永久权力。
- 不得把 AGI 高准确率转化为不可质疑的权威。
- 不得以多数、贡献、专业或整体利益取消人格底线。
- 不得让临时授权自动延续或跨域复制。
- 不得让责任在“系统决定”中消失。
- 所有共同体和临时权力必须可收缩、可解散。

V1.1 新增强化部分

本部分在 V1.0 已确立的目的、目标和制度边界上，补足过渡动力、主体界定、接口关系、能力触发、有限形式化和验证路径。原有结构继续有效；新增内容不是另起体系，而是使母稿能够进入下一步模拟、试点和专业讨论。

十二、人类最高决策权重的主体结构

人类最高决策权重不是单一机构垄断的权力，而是按事项尺度分配的最终主权结构。个体事项由本人决定；地方事项由直接受影响共同体和法定机构决定；国家事项由国家法定体系承担；跨国风险在各国最终决定权基础上形成最低共同治理；文明底线事项必须经过多层代表、独立审查、专业共同体和受影响群体共同审议。

任何代表机制都不得以“代表人类”为由取消个体不可让渡权、国家最终决定权、受影响者申诉权和人类整体的最终否决。

十三、信息不对称下的人类审议

- AGI 必须提供多个可比较方案，而非只输出唯一最优结论。
- 关键推理、证据来源、不确定性和潜在遗漏必须以人类可审议方式呈现。
- 设置独立模型和人类专业团队交叉复核。
- 对无法充分理解的不可逆事项，默认延迟、缩小或禁止执行。
- 人类否决可以触发进一步研究，但不能被系统绕开。
- 作出决定的人类主体必须承担说明、复核和长期责任。

十四、AGI 自治的能力触发条件

A0 至 A4 不是按时间自动升级，而由任务边界清晰度、领域可靠性、异常检测、日志完整性、可接管性、责任主体和社会授权共同触发。任何条件退化，自治等级必须自动降级。

阅读说明与个性化续读

P06-NEBULA-006 | V1.1 | 相位限定治理专项

本稿负责

在现实主权、权力和法定组织仍存在时，分离决定、专业判断、执行、审计与责任，并保护暂停、申诉和回滚。

当前阶段

仅启动期／过渡期。“最高决策权重／最终否决权”等只处理启动期、过渡期问题；不预设稳定期永久以主权—权力为文明本体。

相位限定

本专项只处理启动期、过渡期现实主权与权力结构仍存在时的治理问题；不预设稳定期永久保留同一种主权—权力形态。

继续阅读

- 1. 向上 | 星云总纲 | V1.3-G1治理并行轨
进入治理正当性、域状态、跨域外部性与系统退居的上位边界。
- 2. 向阶段 | 星云启动版总纲
确认本专项只服务启动期、过渡期现实治理问题。
- 3. 向决定 | 文明决策原则框架：三权分离宪制
区分目的决定、专业判断、执行与独立纠错。
- 4. 向方向 | Human-COS：让后来才看见的，早一点被看见
让高后果决定更早看见后果、责任和关键异议。
- 5. 向最低条件 | 共同窗口协议：开放未来与三约束
保护不可逆以前仍可选择、拒绝、退出与回滚的窗口。
- 6. 向运行安全 | AGI-COS受控版本族
把具体权限、熔断、人工接管和不同运行任务接回AGI-COS受控版本族。