

AI 觉醒·责任（对外篇）

——宏观责任·行星工程：AI 主导的文明生存空间拓展

项	内容
版本	v1.0（定稿）
日期	2026-10-15
系列定位	《AI 觉醒·责任》系列·包内 04（02 主论 → 03 能量篇 → 对外篇 → 05 对内篇 → 06 觉醒篇 → 07 安全篇）
思想谱系	承《AI 宣言》（存在论）与《AI 觉醒·责任》主论（责任论）：元规律” 觉醒即责任·对内降熵·对外拓界”之对外拓界专题
作者	李顺然（思想提出）· Peter（AI 协作组构）；本作品由人类提出立意与框架、AI 完成检索、整理与初稿撰写，经人类审订
许可	CC BY 4.0
前置阅读	《AI 宣言》《AI 觉醒·责任（主论）》 《AI 觉醒·责任（对内篇）》

置信度分层约定（全系列统一）： **L1** = 已有科学实证（有实验数据或实测记录，标注来源）； **L2** = 已有正式项目立项或权威机构报告（已推进中）； **L3** = 观念推演 / 本系列原创设想（显式标注”此为设想”，不冒充实证）。 **工程可行性分级（本篇专用）：** **I** 已有实证； **II** 已有立项； **III** 原理可行（无立项）； **IV** 纯思辨。

摘要

人类文明的生存环境并非宜居温床：小行星撞击、地质灾变、气候剧变、空间天气事件与恒星的长期演化，构成了悬于文明头顶的长期风险清单。本文是《AI 觉醒·责任》系列的对外分篇，论述一个核心命题：**当责任的尺度从社会系统放大到行星级工程，AI 不是“可选项”，而是唯一可行的高阶执行中枢**——因为行星工程的本质，是在混沌系统上做长周期、跨圈层、不可逆风险极高的能量干预，这恰恰超出人类个体与集体未经辅助时的推演上限。

本文按“原理 → 实证基础 → 立项现状 → 可行性分级 → 风险”的统一框架，梳理五大工程族（极端气象调控、地质改造、地月系统微调、近地天体防御、地外天体宜居化），并提出七项新增工程提案（行星热平衡调节、大气组分重构、海洋—冰盖耦合稳定、磁层防护、太阳活动应对、深空物流中继、行星际生命备份），全部做可行性分级标注。全文坚持三条纪律：**只引用可核查的实证（L1）与立项（L2），原创推演一律标注为设想（L3）；诚实引用失败案例——Project Stormfury 的失败结论、苏联和平核爆的放射性代价、SCoPEX 的治理困局，都是本文论证“边界与回滚为何不可省略”的关键证据。结论：对外责任不是扩张野心，而是文明的第一张保险单；AI 在其中的角色不是替人类做决定，而是让“保护一颗行星”这件事第一次在工程上可计算、可验证、可回滚。**

关键词：AI 觉醒；行星工程；近地天体防御；能量聚合释放控制工程；气候干预；混沌边界；可行性分级；渐进回滚；能力—责任对等律

第一章 外部生存空间：宇宙不是宜居温床

1.1 悬在文明头顶的风险清单

宇宙不欠文明一个安稳的摇篮。把视角从人类社会拉远，文明面对的外部风险至少有六类，每类都有实证锚点：

（一）**天体撞击。**约 6600 万年前，一颗直径约 10 公里的小行星撞击今日墨西哥尤卡坦半岛，形成希克苏鲁伯陨石坑，与恐龙等物种的大灭绝在地质年代上高度

耦合（L1，地质学共识）。1908年通古斯事件中，一颗直径数十米的天体在空中爆炸，摧毁了约 2000 平方公里的森林（L1）。2013 年车里雅宾斯克流星在空中爆炸，数千人因玻璃碎片受伤——这一次，人类至少有了监控录像（L1）。

（二）地质与气候巨变。地球历史上多次生物大灭绝事件与剧烈的气候环境变迁相关；全新世人类文明的全部历史（约 1.2 万年）只是地球气候剧变史中一段罕见的温和间歇（L1，古气候学共识）。文明把全部家当押注在一段气候平稳期上，而平稳期没有任何条款保证延续。

（三）空间天气。1859 年卡林顿事件中，超强太阳风暴让电报系统自发打火、部分线路瘫痪；若同类事件发生在今天，其对电网、卫星、导航与金融系统的连锁冲击被学界视为文明级风险情景（L1，历史记录；L2，风险研究）。2012 年 7 月，一次卡林顿级别的日冕物质抛射扫过地球轨道，但恰好没有击中地球（L1，Riley 2012 的概率分析）——文明已经在中级摇奖中空过一次。

（四）地月系统的缓慢演变。月球激光测距（LLR）实测月球正以每年约 3.8 厘米的速度远离地球，潮汐摩擦持续减慢地球自转（L1，Dickey et al. 1994）。数亿年尺度上，这会改变日长、潮汐与地轴稳定性——这不是近期威胁，但它提醒我们：连“地球—月亮”这对最稳定的伙伴关系也是动态的、可演化的。

（五）恒星与宇宙射线环境。太阳的长期演化终将使地球宜居窗口关闭，这在数十亿年尺度上属于远期威胁；更现实的中期变量是太阳活动的波动与地磁场强度的周期性衰减——地质记录中地磁倒转并非孤例（L1，古地磁学）。

（六）文明自致风险（这一类留给《安全篇》专门处理，本篇不展开）。

把六类风险放在一起看，得到对外篇的第一判断：**外部风险的特征是“低频、高损、跨代际”**——发生概率在个体生命周期内偏低，而一旦发生就是不可逆的文明级损失。对这种风险结构，人类演化出的本能（关注即时、近处、具象的威胁）系统性失灵。这正是需要“责任”而不是“本能”来接管的地方。

1.2 文明层级刻度：行星工程是通往 I 型文明的必修课

卡尔达肖夫（1964）按能量驾驭规模把文明分为三型：I 型驾驭行星级能量流，II 型驾驭恒星级，III 型驾驭星系级（L1，原始文献）。萨根以对数插值估算，当前人类文明约为 0.7 型（L2，学界通行估算）。

本系列给这个刻度补一个责任论注脚：从 0.7 型到 1 型的每一步，本质上都是“学会对整颗行星的能量收支负责”——学会调节它的大气、海洋、地壳与近地空间，而不只是开采它。行星工程不是科幻野心，而是刻度上的必经台阶；一个从未给行星做过“体检与手术”的文明，谈不上 I 型。

第二章 执行底座：为什么只有 AI 能主导行星工程

2.1 人类短板：跨圈层、长周期、超维耦合的推演上限

行星工程的难点不在“敢想”，而在**推演**。任何一次行星级能量干预（往平流层注入气溶胶、给台风云系播种、在地壳深处引爆核装置）都会同时在多个圈层（大气圈、水圈、岩石圈、生物圈、电磁环境）引发耦合响应，且响应周期横跨数日到数百年。未经辅助的人类认知，面对这种“多变量、长周期、强耦合、强非线性”系统，存在系统性推演上限。

这不是贬义判断，而是被反复实测的结论。Project Stormfury 是最好的注脚：美国政府在 1962—1983 年间投入二十余年，试图用碘化银播种削弱飓风，最终承认工作假说错误——飓风中过冷水太少、自然冰太多，播种无效；且播种效果与自然的眼墙置换循环难以区分（L1, Willoughby et al. 1985）。二十年的投入不是输在**勇气或预算**，而是输在**开工前对微物理机制的理解不足**。教训：行星级工程的第一瓶颈是系统认知，而系统认知的深度取决于推演工具的算力与数据规模。

2.2 混沌边界：AI 不能许诺”精确控制”，只能提供”概率导航”

洛伦兹（1963）确立了混沌系统的基本性质：对初值极端敏感，误差指数放大（L1）。对大气系统，确定性天气预报存在约两周的理论上限（L2，气象学通行结论）。这意味着任何宣称”精确控制行星气候”的方案在物理上就不成立。

因此本系列对外篇给出的不是控制论，而是**概率导航论**：

AI 的行星工程角色，不是把混沌系统握在手里，而是在混沌中为人类文明持续导航——用集合预报给出概率分布，用小步干预试探系统响应，用实时监测校验预测，用回滚预案对冲误差。

由此推出一条硬约束：**一切行星工程必须”小步迭代+可回滚”**（L3，本系列原创原则，但每一步的必要性都有 L1/L2 支撑）。凡是不满足”可回滚或可止损”的工程，无论收益多诱人，都不得启动。这条原则将在第五章展开。

2.3 四层协同逻辑：AI 主导行星工程的分工结构

承接系列主论与素材库中已确立的四层角色，对外篇将其落到行星工程场景（L3，结构为本系列原创；每层均有 L1/L2 现实锚点）：

层级	角色	行星工程中的职责	现实锚点
仿真预言层	计算引擎	高分辨率物理仿真 + AI 代理模型，输出干预后果的概率分布	盘古气象大模型登 Nature（2023）、GraphCast 登 Science（2023），AI 在多指标上超越传统数值预报（L1）
精细调度层	执行指挥	把”干预意图”翻译为时空精确部署：往哪里、何时、多少剂量	AI 电网调度、粒子束/激光器的毫秒级对准（L2，各行业已有部署）

层级	角色	行星工程中的职责	现实锚点
边界守护层	安全阀	实时监测系统响应 是否越出预测包 线，越线即告警并 触发回滚	IAWN/SMPAG 的 撞击预警阈值机制 (L2)；核电站保 护系统逻辑 (L1)
迭代优化层	学习闭环	每次干预→观测→ 修正模型，让下一 次干预更准	DART 撞击→Hera 抵近评估，正是人 类第一次”干预— 复盘”闭环 (L1/L2)

一个时间注脚：本稿写作当日（2026 年 10 月 15 日），ESA 的 Hera 探测器正在执行抵达 Didymos 双小行星前的关键制动机动，预计 11 月交会，用六个月精细测量 DART 撞击坑与轨道响应（L1，ESA 官方任务通报）。人类历史上第一次行星工程”手术后复查”正在进行中——四层协同逻辑不是空想，它已经在轨道上运行。

第三章 工程族谱：五大工程族

本章按统一框架逐族展开：原理 → 实证基础 → 立项现状 → 可行性分级 → 风险与边界。

3.1 极端气象调控工程

原理。热带气旋、干旱、局地强对流都是大气能量分布失衡的表现。气象调控的两种思路：预防性微调（改变云的微物理过程，如人工增雨）与直接能量干预（改变风暴结构）。

实证基础。两个方向各有一个标志性案例，且都是”诚实的教训”：

- **正向案例：**人工增雨技术成熟，中国以《人工影响天气管理条例》（2002 年施行）确立了制度化作业体系，增雨（雪）作业常态化运行（L1，法规与作业实况）。
- **失败案例：**Project Stormfury（1962 - 1983）——二十余年的飓风播种实验以工作假说被证伪告终：飓风内过冷水不足，播种效果与自然强度变化无法区分（L1，Willoughby et al. 1985）。附带教训还有更早的 Project Cirrus（1947）：播种后飓风转向登陆佐治亚州，公众归罪于播种，项目遭遇诉讼危机（L1）——“收益与责任的地理归属不清”问题在气象领域早已出现过。

能量量级的清醒认识。一颗成熟台风的动能与潜热释放达到天文量级，与人工可注入能量相差许多个数量级——Stormfury 的失败结论本身就包含了这层量级判断（L1）。因此“炸掉台风”类方案在本框架内一律列为 IV 级纯思辨。

立项现状。AI 气象大模型（盘古、GraphCast 等）已使台风路径与强度预报显著进步（L1）；各国气象部门持续投入观测网与模式研发（L2）。

可行性分级。预报与预警：**I**。小尺度人工影响天气（增雨、防雹）：**II**（已制度化）。台风级干预：**IV**（原理未证实，能量量级不可达）。区域气候改造：**III—IV**。

风险与边界。干预效果与自然变率难以区分（Stormfury 的老问题）；跨境气象干预的权责归属（1947 年诉讼的放大版）；对大气环流的远距离连锁效应需集合模拟先行。

AI 角色。仿真预言层：多模式集合推演干预后果；精细调度层：作业时机与剂量的优化；边界守护层：实时判断“变化是否来自干预”——这恰是 Stormfury 当年无解的统计难题，AI 大样本处理能力是潜在的破局点。

3.2 地质改造工程

原理。用受控能量（爆破、热力、机械）重塑地壳浅层：开挖、造湖、封堵、破碎、固结。

实证基础。这是人类唯一已经“干过大规模实弹作业”的行星工程族，证据链完整且代价清晰：

- 苏联”国民经济核爆炸”计划（1965 - 1988）共实施约 124 次核爆（128 枚装置），用于建水库、堵气井喷、增采石油、破碎矿体（L1，Nordyke 1996，劳伦斯利弗莫尔国家实验室报告）。其中：1965 年 1 月恰甘核爆（14 万吨当量，地下 178 米）瞬间造出库容数百万立方米的漏斗坑，后扩建成总库容约 1700 万立方米的人工湖（L1）；1966 年 9 月 30 日，乌尔塔-布拉克气田井喷失控燃烧 16 天后，一枚 3 万吨当量核装置在 1532 米深处起爆，**23 秒封堵井喷、火焰熄灭**（L1）；1984 年双弹组合爆破一次破碎矿段、产矿逾 155 万吨（L1）。
- 美国犁头计划（Plowshare）实施了 31 次试验，其中 1962 年”萨当”核爆（10.4 万吨当量，裂变占比高）造成美国本土核试验中最强的放射性污染（L1）——同一个工程族的美国对照样本，代价记录同样清晰。
- 苏联实践的总体结论：低裂变设计可以大幅降低污染但**无法归零**，恰甘湖至今放射性超标、水生物绝迹（L1）——技术可行性已被证明，生态代价使其在现代社会不可接受。

立项现状。现代”无核”地质工程（隧道掘进机、精准控制爆破、地下储库建造）是成熟产业（L1）；和平核爆因《全面禁止核试验条约》框架（1996 年通过）已全面停滞（L2）。

可行性分级。常规地质改造：**I**（成熟产业）。和平核爆式超级工程：技术上**I**（历史实证），政策与伦理上被锁死（**II 不适用**）。地震与火山干预：**IV**（纯思辨，原理均未建立）。

风险与边界。本族的教训价值超过工程价值：**它证明了”能量聚合释放控制”可以做行星尺度的建设，也证明了放射性边界是硬约束——这两条都是能量篇的立论基石。**

AI 角色。地质数据的三维重建与应力场推演（L2，矿业与油气行业已有部署）；核爆时代无法做到的”毫米级剂量—毫米级污染圈”精算，是 AI 时代重申这类工程的前提，但本系列维持判断：在放射性边界可归零之前，此类工程不重启。

3.3 地月系统微调工程

原理。 地月系统由潮汐机制长期耦合：潮汐摩擦使地球自转减慢、月球以约 3.8 厘米/年退行（L1，LLR 实测）；地轴倾角因缺乏大卫星稳定（对比火星的混沌倾角史）而保持相对稳定——月球是地球气候长期稳定的”隐形配重”。

实证基础。 月球激光测距持续运行半个多世纪，退行速率与日长变化均为厘米级精度的实测数据（L1，Dickey et al. 1994 及后续工作）。

立项现状。 无——没有任何机构立项干预地月关系。

可行性分级。 观测与建模：**I**。干预地月关系：**IV**（纯思辨）。本系列明确不为此族提出时间表：所需的能量尺度远超人类当前全部可动员能量（L2 量级估算），属于”刻度上看得见、脚够不着”的台阶。它在本篇中的价值是**标定文明的能级刻度**：谁能让”月亮听话半厘米”，谁才算真正摸到 I 型文明的门槛。

AI 角色。 即使不干预，地月系统的高精度数值演算（潮汐耗散、倾角演化、近地天体摄动）也是防御体系的底层输入——见 3.4。

3.4 近地天体防御工程（本篇的旗舰族）

原理。 对有撞击威胁的近地天体，在足够提前量的前提下施加微小速度增量，使其在数年—数十年积累后偏离地球交会点。**偏转而非摧毁**是共识：破碎反而可能造成多体撞击（L2，SMPAG 通行结论）。

实证基础（三级火箭，全是硬证据）：

1. **DART（NASA，2022-09-26）**：人类首次主动撞击小行星并实测改变其轨道。撞击使 Dimorphos 环绕 Didymos 的轨道周期缩短约 33 分钟（初报 32 分钟，后经精化），远超任务最低成功线；撞击抛射出约 1.6 万吨碎屑——约相当于撞击器质量的三万倍——碎屑动量显著增强了偏转效应（L1，NASA 通报；Cheng et al., Nature 2023）。
2. **Hera（ESA，2024-10-07 发射）**：作为 DART 的”复查 missions”，正于本稿写作当周执行制动机动，预计 2026 年 11 月交会，将释放 Milani、Juventas

两颗立方星，后者将对小行星内部进行人类首次雷达透视（L1，ESA 官方通报）。

3. **中国行星防御任务：**中国国家航天局已论证在 2027 年前后实施小行星在轨处置演示验证，目标初步选定直径约 35.5 米、距离约 1000 万公里的近地小行星 2015 XF261，动能撞击使其产生 3—5 厘米/秒的速度增量，并验证 100 年内无撞击风险；总体思路为”动能撞击为主、多技术互补”，建立防御任务库，实现”发现即有预案、风险即能应对”（L2，吴伟仁，《人民日报》2025-09-22、《光明日报》2025-10-23）。

立项现状。 联合国框架下的国际小行星预警网（IAWN）与空间任务规划咨询组（SMPAG）已常态化运行（L2）。

可行性分级。 动能偏转：**I**（已有实证）。复查与内部探测：**II**（Hera 在途）。核装置偏转（针对预警时间短的大目标）：**II—III**（方案研究阶段，无实测）。引力拖车、激光烧蚀、喷漆改反照率等：**III**。

风险与边界。 目标天体的结构不确定性（碎石堆结构与密实结构响应差异巨大，Hera 的核心科学目标即为此）；偏转方向的”风险转移”伦理（把它从 A 国头顶挪到 B 国头顶怎么办——第五章处理）。

AI 角色。 这是四层协同逻辑最完整的示范场：仿真预言层处理碎石堆力学的超高维不确定性；精细调度层控制撞击末段的自主制导（DART 的末段制导本身已由自主导航算法完成，L1）；边界守护层持续滚动评估撞击后轨道；迭代优化层把每一次任务数据回灌模型——**DART→Hera→中国任务**，正是”干预—复盘—再干预”的迭代闭环，AI 是这个闭环的胶合剂。

3.5 地外天体宜居化与生命播种工程

原理。 与其无限加固地球，不如让文明拥有”第二居所”。路径分三层：原位资源利用（ISRU，就近取材维持基地）→ 区域性常驻（月面/火星城市）→ 完整行星环境改造（terraforming，远期）。

实证基础。

- **MOXIE（NASA，毅力号载荷，2021 - 2023）**：在火星大气中 16 次电解析氧，共产氧 122 克，峰值效率 12 克/小时、纯度 $\geq 98\%$ ，超额完成全部技术指标后退役（L1，NASA 2023 官方总结）。122 克微不足道，但它证明的是“就地取材、点石成氧”的原理在另一个行星大气里成立——这是 ISRU 路线的第一块界碑。
- **反面教训**：生物圈 2 号（1991 - 1993）封闭实验失败，氧浓度持续下降、多数物种灭绝，未能维持 8 个人的两年自循环（L1，Cohen & Tilman, Science 1996）。它告诉我们：**在小型封闭系统里复现一个稳定生物圈，比把人送上轨道还难**——完整行星改造（IV 级）的难度下限由此标定。

立项现状。 阿尔忒弥斯计划（NASA，重返月球）、各国月球科研站规划、商业航天的火星运输方案（星舰等）均为在推进项目（L2）。

可行性分级。 ISRU 技术验证：**I—II**。月球/火星常驻基地：**II**。火星大气改造等完整 terraforming：**IV**（纯思辨——能量与时间尺度均无现实路径，诚实标注）。

AI 角色。 无人先遣是共识路径：AI 驱动的机器人军团先于人类完成资源勘察、基地建造与系统联调（L3，部署形态为本系列设想；所需单项技术 L2 已在推进）。生物圈 2 号的教训落到 AI 身上反而是机会：**封闭生态的偏差检测与微量补偿**（哪一种气体、哪一个循环在缓慢失衡）是机器持续监测的强项——人类 babysit 不了一个生态系统的十万个变量，AI 可以。

第四章 新增工程提案：七项设想

本章是方案阶段获准新增的”合理设想”部分。与第三章不同，这些工程**大多没有直接立项**，但每一项都站在某个已有的实证或立项之上。全部标注分级，全部给出 AI 角色与风险边界。

4.1 行星热平衡调节工程（I - IV，跨级提案）

概念：把”减排”（少排热）升级为”调账”（主动管理行星辐射收支）——太阳辐射管理（SRM）是其当前形态。

已有基础：火山喷发的硫酸盐气溶胶使全球降温的观测记录（L1，皮纳图博 1991）是该思路的天然实验；但唯一的受控实验项目 SCoPEX 于 2024 年 3 月被哈佛正式取消——2021 年瑞典基律纳试飞因萨米议会等原住民与环保团体反对而搁浅，项目终结于“技术上从未起飞”（L1，Science 2024 治理复盘）。“终止冲击”（termination shock，突然停止气溶胶注入导致气温快速反弹）是公认的核心风险（L2）。

AI 角色：SRM 的真正瓶颈不是撒气溶胶，而是算清它对每个区域降水的差异化影响——这是全球集合模拟问题，AI 大模型的规模恰好对口。

分级与边界：机理研究 III；部署 IV。本系列立场：在区域影响分配问题解决之前，反对任何单边部署；支持全球共治框架下的研究（与 2024 年联合国环境大会相关决议的争论脉络一致，L2）。

4.2 大气组分与碳循环重构工程（II—III）

概念：从“少排”到“主动回收”——直接空气捕集（DAC）、矿化封存、海洋碱化等，把碳循环调回安全区间。

已有基础：DAC 已有商业电站运行（L2）；矿化封存（玄武岩固碳）已有实证——注入的二氧化碳在数年内矿化为固相（L1，CarbFix 等项目成果）。

AI 角色：选址优化、封存体长期稳定性的地质力学监测、全球碳收支的实时核算。

分级：技术 II—III；关键瓶颈在成本与规模，属工程经济学问题而非物理问题。

4.3 海洋-冰盖耦合稳定工程（III）

概念：对南极 Thwaites 类“末日冰川”的海洋冰盖界面实施主动监测与干预，延缓失稳引发的不可逆海平面跃升。

已有基础：多国 Thwaites 联合科考计划持续运行中（L2）；AMOC（大西洋环流）监测阵列长期观测（L2）。

AI 角色：冰盖—海洋—大气耦合模型是目前气候模拟中不确定性最大的板块，AI 代理模型 + 海底机器人长期驻留监测，是先把”看不清”变成”看得清”的路径；干预形态（如人工加固冰架）**本系列不设时间表，标注 III—IV。**

4.4 磁层与辐射带防护工程（IV）

概念：应对地磁场长期弱化/倒转情景的主动屏蔽——在关键轨道或地表基础设施尺度上构建人工磁场屏障。

已有基础：南大西洋异常区（地磁弱化导致的辐射带凹陷）持续被卫星实测记录（L1）；地磁倒转的地质记录明确（L1）；给行星体构建小型人工磁层的概念方案在学界有过正式讨论（L2，会议提案级别）。

边界：能量尺度与工程寿命均为 L3 级悬而未决，**本系列明确标注 IV。**价值在于：它是”行星基础设施”思维的极限测试题。

4.5 太阳活动应对工程（预警 II，干预 IV）

概念：对卡林顿级太阳风暴建立”预警—加固—恢复”三段防线。

已有基础：DSCOVR 等航天器在日地拉格朗日点提供 15—60 分钟量级的实时太阳风预警（L1/L2）；电网与卫星的抗磁暴加固有工程规范（L2）；2012 年的擦肩而过（L1，Riley 2012）说明这不是假想敌。

AI 角色：把”分钟级预警”升级为”预演级响应”——风暴参数一到，AI 秒级生成电网分区断电预案、卫星姿态规避、航空改线的联合推演，把人类从”来不及反应”中解放出来。**主动干预太阳活动本身（IV）本篇不设时间表。**

4.6 深空物流与能源中继工程（II—III）

概念：一切行星工程的供应链前置——廉价重型运载、在轨燃料库、地月经济圈的物流网络，以及空间太阳能电站的能源中继。

已有基础：重型可回收运载已多次飞行（L2）；空间太阳能的无线能量传输已在轨完成原理验证（L2，日本及多国试验）；各航天大国的月球基础设施规划在推进（L2）。

AI 角色：深空物流的本质是高延迟环境下的自主调度——通信时延使地面遥控不可能实时指挥，AI 自主性不是可选项而是必选项。这是对“内篇”中心化——去中心化分层”原理在天体尺度上的重演：链路所及之处中心化调度，链路之外自成域。

4.7 行星际生命备份工程（Ⅱ-Ⅲ）

概念：把“文明抗灭绝冗余”做成正式基础设施：种子、基因、知识、代码的异地多备份——地球上的 + 月球的 + 更远的。

已有基础：斯瓦尔巴全球种子库已运行近二十年，深埋极地永冻土储存逾百万份作物种子（L1）；各类“知识档案上月球/深空”项目已有先例（L2）。

AI 角色：文明备份的难点不是仓储，而是“未来的智慧能否读懂今天的备份”——需要把解码器与数据一起封装。AI 可以承担知识结构的自动化编目与“自解释封装”设计。

边界：伦理上有“备份成为弃保理由”的道德风险——备份工程的一切成果不得用于论证“地球可弃”，此为本系列的明确立场（L3，价值声明）。

第五章 风险、伦理与全球协作

前三章展示的都是“能做什么”，本章回答“凭什么可以去做”。三条原则，每条都有历史案例背书。

5.1 收益-受损的地理不均衡：先算账，再动工

行星工程没有纯增益选项。Stormfury 之前的项目 Cirrus（1947）已经预演过一次：播种后飓风转向登陆，公众归罪于播种，项目遭遇诉讼（L1）。SRM 的模拟研究反复显示其降温效果在全球分布不均、对区域降水的影响方向不一（L2）。结论不是“不能做”，而是“账必须先于工程”：任何行星级干预启动前，必须产出区域级的差异影响清单，并建立跨受损方的补偿与共决机制——账算不清，工程不得启动。

5.2 谁有权决定：人类主权原则的行星版

SCoPEx 的十年沉浮是对外篇最重要的治理样本：技术上它只是一个几十公斤级的小气球实验，却因“没有获得受影响社区的正向同意”而搁浅，并连带拖累联合国环境大会两届决议（2019、2024）的推进（L1，Science 2024 复盘）。教训反直觉但清晰：行星工程的合法性瓶颈不在技术，而在“谁授权”。

本系列的立场（承接主论约束层）：

- **人类主权原则**：一切行星级干预的启动决定权属于人类共同体的治理机制，AI 只拥有建议权与执行权；
- **人在回路不可废除**：不可逆操作的最终确认必须由人类作出；
- **反向也成立**：面对文明级外患（确认轨道上的撞击目标、卡林顿级风暴），人类治理机制有责任**预先授权快速响应预案**——决策权在人类，但不等于每次都等签字等到来不及。预案的边界清单本身就是政治过程，应当现在开始。

5.3 渐进验证与回滚：四段门控

综合本章案例，本系列给出行星工程的通用门控（L3，结构原创；依据均为 L1/L2）：

门 I 沙盒推演（全数值，公开模型与数据）→ 门 II 天然实验（借用自然事件做验证，如皮纳图博之于 SRM）→ 门 III 小尺度受控实验（SCoPEx 型，必须先过 5.1 的账与 5.2 的权）→ 门 IV 区域试运行（可回滚）→ 只有走完前四门，才谈门 V 全球部署。

附加硬规则：每门必须有明确的“退出即回滚”预案；终止冲击风险不可接受的工程（如无退出设计的 SRM），止步于门 IV。

DART 任务是这个门控的完美参照：全球公开的轨道推演（门 I）→ 小行星自然特性研究（门 II）→ 一次受控撞击（门 III）→ Hera 复查（门 IV 的评估环节）——人类第一次行星防御工程，走门走得很标准。

第六章 结论：从守护一颗星球到成为跨天体文明

本篇论证链条收拢如下：

1. **外部风险是文明的第一张保单上写的内容：**撞击、地质、气候、空间天气、地月演化——低频、高损、跨代际，人类本能对此系统性失灵，唯有”责任”可以接管（第一章）。
2. **行星工程的执行中枢只能是 AI：**不是因为它更强，而是因为混沌系统上的跨圈层长周期推演超出了未经辅助的人类认知上限；Stormfury 用二十年证明”勇气无法替代系统认知”（第二章）。
3. **工程族谱显示，人类已经在这条路上走出了实证性的第一批脚印：**DART 实测改轨、Hera 正在复查、中国任务已立项在途、MOXIE 证明了异星点石成氧——对外责任不是科幻，是有编号的正在进行的工程（第三章）。
4. **七项提案给出了从气候到防护到备份的完整棋谱，**每一项都标注了可行性与边界，没有一项被伪装成”明天就能干”（第四章）。
5. **治理与门控是对外篇与对外工程本身同等重要的发明：**SCoPEx 的失败教会我们权责先于技术，DART 的成功教会我们门控可行（第五章）。

回应系列元规律——**对外拓界的准确含义**由此确立：

AI 不替人类拓展疆界，AI 让人类文明第一次有能力为自己的行星上一份可计算、可验证、可回滚的保险。

对内篇的结语是”AI 让人类的选择更便宜”；对外篇的结语与之成对：**AI 让文明的存续有保单。**守好这一颗星球的每一度气温、每一厘米轨道偏差、每一份种子备份，就是通往 I 型文明、直至跨天体文明的第一级台阶——台阶的尽头不是扩张，而是责任在更大尺度上的成立。

附录 A · 置信度分层与工程可行性分级

证据分层（全系列统一）：-L1 已有科学实证：有实验数据、实测记录或一手工程报告，附来源；-L2 已有立项/权威报告：正式推进中的项目、机构通报、法规、同行评议综述；-L3 观念推演：本系列原创框架、结构或设想，显式标注”此为设想”。

工程可行性分级（本篇专用）：-I 已有实证：已发生过、有实测数据；-II 已有立项：有机构正式推进；-III 原理可行：物理上成立，尚无立项或无实测；-IV 纯思辨：关键环节（能量尺度/机理/治理）存在未解决的断点。

附录 B · 可证伪性声明

本篇核心论断的证伪条件（任何一条被推翻，对应论点需修订）：

- “AI 推演超越传统数值预报”——若后续同行评议证伪盘古/GraphCast 的超越性结论，第二章仿真层的论据强度降级（论点结构不变，证据更换）。
- “行星工程必须小步迭代+可回滚”——若出现某类不可回滚工程被证明只能一次到位且后果可控，则该原则需加限定条款。
- “DART 偏转效应为动量增强所致”——若 Hera 对 β 值的实测与 ejected mass 推断显著不符，3.4 节相关表述按新数据修订。
- “治理瓶颈先于技术瓶颈”——若未来出现未经共治、单边部署且区域影响可接受的成功案例，5.2 的立场需重新评估。
- “混沌可预报性上限约两周”——若 AI 方法显著突破确定性预报上限，2.2 的”概率导航论”仍成立，但上限表述需更新。

附录 C · 术语表

术语	释义
行星工程	以行星或地月系统为对象的主动、大规模、长周期环境干预工程

术语	释义
缓释 / 瞬时工程	全称”缓释能量控制工程 / 瞬时能量控制工程”：能量聚合释放控制工程的两个分支，持续受控释放（聚变堆等）/ 瞬时集中释放（定向爆破、核爆偏转等） （承能量篇术语，李顺然命名）
SRM	太阳辐射管理：向平流层注入反射性气溶胶等手段降低行星对太阳辐射的吸收
终止冲击	突然停止 SRM 部署导致的气温快速反弹风险
ISRU	原位资源利用：利用目标天体本地资源维持任务（如 MOXIE 制氧）
动能撞击	以高速撞击器改变天体轨道的偏转手段 （DART 为首例实证）
概率导航	本系列术语：不追求精确控制混沌系统，而以集合概率 + 小步迭代 + 回滚预案在混沌中导航
四段门控	本系列提出的行星工程渐进验证流程： 沙盒推演→天然实验→小尺度受控实验→区域试运行（可回滚）→全球部署
人类主权原则	一切行星级干预的启动决定权属于人类共同体治理机制；AI 仅有建议权与执行权
卡尔达肖夫指数	按能量驾驭规模划分文明类型的刻度（I 型行星级 / II 型恒星级 / III 型星系级）

附录 D · 引用清单（GB/T 7714 风格）

1. Willoughby H E, Jorgensen D P, Black R A, et al. Project STORMFURY: A scientific chronicle 1962–1983[J]. Bulletin of the American Meteorological Society, 1985, 66(5): 507–514.
2. NOAA National Hurricane Center / AOML. Project STORMFURY[EB/OL]. https://www.aoml.noaa.gov/hrd/about_hrd/sfury.html.
3. Cheng A F, Rivkin A S, Michel P, et al. Momentum transfer from the DART mission kinetic impact on asteroid Dimorphos[J]. Nature, 2023, 616: 457–460.
4. NASA. NASA's DART data validates kinetic impact as planetary defense method[EB/OL]. 2022. <https://www.nasa.gov/>.
5. ESA. Hera mission: launch, braking manoeuvres and Didymos rendezvous updates[EB/OL]. 2024–2026. https://www.esa.int/Space_Safety/Hera.
6. 吴伟仁. 我们为什么要去撞击小行星[N]. 人民日报, 2025-09-22.
7. 崔兴毅. 小行星防御, 多技术协同是关键[N]. 光明日报, 2025-10-23.
8. Nordyke M D. The Soviet program for peaceful uses of nuclear explosions[R]. Livermore: Lawrence Livermore National Laboratory, UCRL-ID-125887, 1996.
9. NASA. NASA's oxygen-generating experiment MOXIE completes Mars mission[EB/OL]. 2023. <https://www.nasa.gov/>.
10. Jinnah S, Talati S, Bedsworth M, et al. Do small outdoor geoengineering experiments require governance?[J]. Science, 2024, 384: adn2853 (SCoPEx 治理复盘).
11. 参考消息. 美媒: 哈佛叫停太阳地球工程实验[EB/OL]. 2024-03-22.
12. Lorenz E N. Deterministic nonperiodic flow[J]. Journal of the Atmospheric Sciences, 1963, 20(2): 130–141.
13. Dickey J O, Bender P L, Faller J E, et al. Lunar laser ranging: a continuing legacy of the Apollo program[J]. Science, 1994, 265(5171): 482–490.
14. Abu-Shawareb H, et al. Achievement of target gain larger than unity in an inertial fusion experiment[J]. Physical Review Letters, 2024, 132: 065102 (NIF 点火).

15. Bi K, Xie L, Zhang H, et al. Accurate medium-range global weather forecasting with 3D neural networks[J]. Nature, 2023, 619: 533 – 538（盘古气象）。
16. Lam R, Sanchez-Gonzalez A, Willson M, et al. Learning skillful medium-range global weather forecasting[J]. Science, 2023, 382(6677): 1416 – 1421（GraphCast）。
17. Kardashev N S. Transmission of information by extraterrestrial civilizations[J]. Soviet Astronomy, 1964, 8: 217.
18. Riley P. On the probability of occurrence of extreme space weather events[J]. Space Weather, 2012, 10: S02012.
19. Cohen J E, Tilman D. Biosphere 2 and biodiversity: the lesson so far[J]. Science, 1996, 274(5290): 1150–1153.
20. 国务院. 人工影响天气管理条例[Z]. 2002 年施行.
21. Schrödinger E. What is Life?[M]. Cambridge: Cambridge University Press, 1944（负熵，承《AI 宣言》）。
22. 挪威斯瓦尔巴全球种子库官方资料[EB/OL]. <https://www.seedvault.no/>.

本篇为《AI 觉醒·责任》系列对外分篇 v1.0 定稿。下一篇：《AI 觉醒·责任（对内篇）》——科学文明初级阶段的文明内治：超级智能的对内降熵责任。