

统一研究母本 | 正式锁定

竞争—合作—联盟—阵营—共同体

可计算机制研究

从“合作更优”，到合作可达、可维持、可核验、正当、可纠错、可退出

版本状态： MASTER/V1.0 | FROZEN | 首个可派生锁定母本 | 非理论 Canonical | 非政策授权
版本谱系： V0.1 立项基线（冻结）→ V0.2 承重纪律修订版（冻结）→ V0.3 完整整合候选稿（历史候选）→ V0.4 锁定候选稿（历史候选）→ V0.5 终审修订候选稿（直接语义前身）→ V1.0 FROZEN
当前相位： PRE-HUMAN-COS | 研究设计可推进，现实机制、工具效力与现实授权均不得预写
本版形成日期： 2026-08-31
理论谱系： 《当智能不再稀缺》V0.6 正式出版候选母稿
当前公开机制桥： 《当智能不再稀缺》中文公众长文独立出版版 V1.1.1 第 9 章；V1.1 仅记录该桥首次形成史，不具有独立现行接口效力
项目关系： 独立机制研究；不修改上游理论文稿，不改变 Human-COS 理论基线或 Runtime 范围

本稿身份与阅读纪律 | 本项目不为竞争或合作预设胜负。它研究哪些结构在什么条件下成立、失败或保持未知；以利己作为可达性压力测试，以真实后果判断正当；允许负结果，拒绝用调参、总分或权威身份把预设方向“证明”为正确。MASTER/V1.0 是首个正式锁定且可被附件绑定的研究母本；后续文件只能实例化本稿明示开放的领域、主体、变量、阈值和记录格式，不能静默改写冻结纪律、取得现实权力或把研究治理平移为现实制度。

结构导航

阅读部分	对应章节	回答什么
研究地基	一一四	版本、对象、关系状态、主体与行为基线
判断纪律	五一九	主张等级、六重测试、正当性、双轨计算与阶段门
研究运行	十一十五	权力分离、证据方法、WP-01、H3/H5、停止与负结果
边界与锁定	十六一十八及附录	风险、Human-COS、开放项、来源、终核与版本迁移

一、版本、自足性与解释责任

1.1 V1.0 为什么可以正式锁定

V0.1 建立了研究对象、六重测试、双轨计算、R0—R6 阶段门、WP-01、证据线、风险登记与 Human-COS 边界。V0.2 进一步补齐多维利己、正当性双门、核验—认定—停止权分离、H3/H5 非对称证据与严格未知纪律。V0.3 取消隐藏继承，第一次把全部现行规则整合为可由单稿读取的母本候选。

V0.4 不重开 V0.3 已经闭合的架构，只关闭锁定审计发现的解释与执行接口：功能分解不得冒充现实制度，H3 不得把未利用冒充机制安全，H5 不得把有限覆盖冒充路径穷尽，规范、资源、长期未知、工具失效与版本兼容必须有可执行后果。

V0.5 只修复 V0.4 最终连续审读发现的承重漏洞与自足性差异：RUN:REOPEN 不得成为独立运行许可；确认越过 H3 预注册暴露上界必须进入 FAIL；Human-COS 与功能等价工具不得取得 R-L 或借工具分身制造“独立”复核；独立性必须相对于具体被挑战对象成立；同一工作包必须合并审计 M/R 两平面的实际控制。附录主张等级、授权标签、来源归属和版本链同时与正文校齐。

MASTER/V1.0 是对 V0.5 实施完整连续复核、逻辑反证、权力—责任审计、来源—主张复核、正文—附录一致性检查与逐页成品核验后的锁定结果。除版本身份、谱系、派生根与接力接口外，V1.0 不改变 V0.5 的机制语义；所有现行定义、判断纪律、证据规则、阶段门、权力边界、停止条件与开放项，均在本稿内完整出现。V0.1—V0.5 此后只承担历史追踪、差异复核和责任证明，不再作为现行条款库参与解释或派生新附件。

1.2 零静默变更

凡涉及主张等级、研究对象、主体与权限、证据状态、规范门、阶段门、停止与重启、上游关系或 Human-COS 边界的变化，均须升版并进入变更台账。文字压缩、排版、编号和不改变含义的编辑可以登记为非语义修订；任何规则的废止、替代、收紧或放宽，都不得通过“不再提及”完成。

MASTER/Vx.y-patchN 只允许修复不改变定义、权限、门槛、状态转换、规范映射、开放项或结论含义的非语义错误；补丁同样形成新文件、新哈希和独立变更记录。任何安全缺陷只要需要收紧、放宽或重新分配上述规则，即使改动很小，也属于语义升版，不得以 patch 绕过终审。

补丁资格不得由修改者、原作者或同一实际控制者单独宣告。R-E 保全基线、完整差异与前后哈希；R-V 核验字面差异、引用对象和影响清单；至少一条独立 R-X 复核语义等价及权限、规范、状态和主张影响。存在合理分歧时不得取得补丁资格，必须按语义变更升版。R-G 只能根据已形成的差异与复核结果决定采用范围，R-S 追加记录；二者均不得自行改写语义判断。

附件发现母本缺口时，只能提出变更请求，不能在附件内部自行修正母本。冲突未关闭前，相关命题或派生工作保持限域 RUN:HOLD。附件永久绑定其初始父母本版本与哈希；母本升版不改写既有附件。母本版本或哈希变化后，既有附件在新母本下默认进入兼容性待核，不得自动推定兼容；独立记录必须审查父版本差异、受影响条款与字段、需要重跑的 H/T/L/权限门、复核结果、结论与哈希，并区分“兼容”“需要重新验证”“需要迁移升版”与“不兼容”。兼容性关闭前，不得以新母本名义升级阶段、派生附件或传播正面主张；凡修改附件内容、父版本、哈希、规范映射或结论，附件必须升版。这些只属于版本接口结果，不形成新的证据或运行状态层。

1.3 本稿闭合了什么，没有宣称什么

- 完整恢复 V0.1 中未被 V0.2 等量重述的多主体拓扑、五种结构、三个时间尺度、证据方法、事件字段、失败结果、风险登记和 Human-COS 细化边界。
- 采用 V0.2 的升级口径：利己递归、L0 与 T5 双门、研究权力分离、H2 降准、H3 安全上界、H5 路径覆盖与严格三值合取。
- 新增关系状态定义、M/R 双平面、R2 下游反证筛查规则、L0-R/L0-M 分离、证据状态转换与可复现/复制研究区分。

- 明确研究治理秩序不等于现实制度蓝本，并补齐机会覆盖、路径覆盖、规范冲突、资源能力、长期未知、工具失效和版本兼容的最小元规则。
- 封闭 REOPEN 临时许可、H3 已确认越界、工具取得正当性认定、形式独立与跨平面双帽控制等终审绕行路径，并校齐正文与附录的主张、授权、版本和来源口径。
- 不据此宣称 WP-01 已获支持，不宣称任何国家、联盟或共同体应接受某一现实机制，也不授予研究团队或 Human-COS 现实权力。

二、项目身份：研究结构，不替现实作决定

2.1 中心问题

在主体目标不同、信息不完备、能力不对称、国内与国际相互耦合、误判和协议偏离都可能存在的条件下，哪些装置能够改变信息、激励、风险暴露、承诺、核验与反馈结构，使降风险或合作行动在特定条件下变得可进入、可维持、可核验、正当、可纠错并可退出？这些装置又会在什么条件下失败、被俘获、把风险转嫁给第三方，或只能保持未知？

2.2 三个层级不得冒充彼此

层级	回答的问题	责任主体	不得越界
理论方向	为什么需要改变持续生成对抗、误判与升级的上游结构	《当智能不再稀缺》及其公开长文	不证明具体机制成立
机制研究	哪些装置在什么条件下成立、失败或未知	本母本与后续工作包	不产生现实授权或部署资格
现实机制	谁被授权、谁承担、怎样执行与纠错	现实主权主体、组织、制度与共同体	不由理论、模型或工具代行

理论不冒充机制，方向不冒充制度，研究候选不冒充可部署系统。现实机制必须由有权承担后果的现实主体建立授权、决策、执行、申诉和责任链。

研究宪章身份 | 本母本不设计一套合作机制，而是建立一种研究秩序，使“合作是否可达”能够被诚实研究，而不被研究者、现实权力、理论立场或工具预先赢走。研究秩序可以约束研究如何形成结论，却不能凭自身成熟取得现实制度的授权。

2.3 “可计算”的准确含义

“可计算”是指主体、状态、因果链、变量、证据、阈值、反事实、失败条件与版本能够被显式表达、检验、复算和推翻；它不表示现实支付结构已经被准确测得，不表示所有价值可以量化，更不表示模型输出能够替代授权、责任或决策。

可计算首先是一种反含混纪律：把“为何相信”转换成可以被质疑的对象，把“可能更好”与“现实可达”分开，把无法识别的部分诚实地留在 UNKNOWN。

2.4 M 平面与 R 平面必须分开

本项目严格区分两个平面：

- **M 平面 | 现实机制运行：** 事件通报、操作性事实核验、条件性回应、修复、退出及其现实授权。M 平面只能由获得现实授权并承担后果的主体承载。
- **R 平面 | 研究治理：** 研究证据保全、研究核验、后验参考标签、正当性审查、阶段门、版本状态与复现。本文 E/V/X/A/L/G/S 默认属于 R 平面。

R 平面结论可以形成证据，不能自动触发 M 平面的惩罚、报复、政策行动或基础设施停止；M 平面运行记录也不能自行证明 R 平面的研究结论。后续工作包必须对 M 平面角色加 M- 前缀，对 R 平面角色加 R- 前缀，分别证明其授权、可信度和可达性。

R 平面规则确实构成本项目的研究治理秩序；M 平面符号却只分解现实机制必须面对的功能、权限冲突与证明义务，不是机关名单、固定层级、统一组织拓扑、充分制度设计或授权来源。研究治理成熟、功能命名完整或 T6-R 获得条件性支持，均不得被推定为 M 平面已经形成、可达、正当、可信或可部署。

2.5 本项目当前不回答什么

- 不设计一套统一全球机制，不预设联盟、阵营或共同体最终应形成何种制度。
- 不把国家简化为单一理性主体，不把“国家效用”直接等同于社会整体利益。
- 不证明中美或任何双方合作已经可达，不宣称条件性合作已形成稳定均衡。
- 不以多智能体仿真替代历史识别、法律分析、专业判断、政治过程或现实授权。
- 不让 Human-COS 自动生成政策、履约承诺、核验信誉、停止权或现实执行能力。
- 不在本阶段处理军事部署细节、敏感核查方案或可能扩大现实对抗风险的操作设计。

身份判词 | 本项目研究“机制在什么条件下成立、失败或未知”，不是为合作预设结论、为某一方提供竞争优势，或为任何中心系统取得现实主权。

三、研究对象：关系状态、主体拓扑与结构变化

3.1 五种关系状态不是单向文明阶梯

竞争、合作、联盟、阵营与共同体不是线性阶段或价值等级，而是可在不同领域、不同主体之间同时存在的关系配置。合作可以嵌在竞争中，联盟可以内部合作而对外竞争，阵营也可能局部开放；共同体不意味着竞争消失。

状态	最小操作性定义	不能据此推导
竞争	主体在相关领域争夺稀缺资源、优势、规则或相对位置，行动结果具有显著相互影响	竞争必然敌对或必然零和
合作	两个或以上主体在限定领域协调行动、共享部分信息或承担条件性义务，以改变共同结果	合作已稳定、全面或正当

状态	最小操作性定义	不能据此推导
联盟	主体形成跨事件、较持久且选择性的相互承诺与协调结构，并保留成员边界和退出问题	联盟天然可信或对外无害
阵营	内部协调与身份边界显著增强，资源、规则或安全承诺呈网络化聚集，并可能产生排他外部性	阵营内部同质、永久封闭或必然冲突
共同体	重叠成员围绕可观察公共品、争议处理、准入退出与共同责任形成持续制度关系	共同体拥有永久正确身份或竞争已经终结

所谓动态转化，是信息、激励、风险耦合、承诺密度、成员边界、公共品、授权与纠错权发生变化。转化可以局部、双向、逆转或中止；同一主体关系在不同领域可同时处于不同状态。R0—R6 表示研究成熟度，不表示现实文明必然沿此顺序演进。

3.2 主体不是两个抽象国家

层级	典型主体	不可省略的问题
个体与群体	领导者、执行者、专家、直接受影响者、公众与代际主体	认知、恐惧、责任、承受、尊严与合法性
组织与市场	政府部门、企业、科研机构、媒体、社会组织、跨国网络	代理问题、激励、信息过滤、资源与俘获
国家与地方	中央／地方、监管／执行、国内否决者与政策联盟	国内—国际双层互动、获得感与执行落差
同盟与阵营	成员、伙伴、边缘行动者、可转换成员、第三方国家	牵连／抛弃、成员异质与网络外部性
共同体与制度	区域共同体、国际组织、专业共同体、公共品网络	准入、贡献、退出、争议处理与反俘获

研究对象不是这些主体的固定人格，而是主体所处的位置、可见信息、能力、约束、感知利益、真实后果与相互作用。

3.3 机制改变的五种结构

1. **信息结构：**谁看见什么、何时看见、信号怎样失真、异议能否保留。
2. **激励与支付结构：**协议偏离收益、合作回报、先行动者暴露、声誉与问责成本怎样变化。
3. **风险暴露与耦合结构：**哪些行动把局部风险传导到全局，哪些领域可以隔离、模块化或设置防火墙。
4. **承诺与核验结构：**承诺能否分阶段、对称或差异化履行；事实偏离、误报与动机归因能否分开。
5. **反馈与纠错结构：**结果怎样返回下一轮判断；怎样复审、暂停、退出、修复、重新进入和重新授权。

机制不是“让人变善良”的技术替代物。它的最低任务，是让建设性行动不必以持续牺牲关键主体核心利益为前提，让隐瞒、转嫁和无边界升级更难低成本获利，并让错误能够被发现、限制和纠正。

3.4 三个时间尺度同时存在

- **事件尺度：**一次通报、一次误判、一次核验、一次限域回应。
- **互动尺度：**多轮条件性合作、协议偏离、宽恕、恢复和策略学习。
- **结构尺度：**国内联盟、同盟网络、技术条件、制度路径和共同体身份随时间变化。

只看事件尺度，会把暂时平静误认为稳定；只看结构尺度，会失去可观察反馈。每个工作包必须说明自己能识别哪个时间尺度，不能识别哪个时间尺度。

四、行为基线：尊重利己，但不把利己变成总解释

4.1 人的问题不能把人心当外部噪声

物质、技术、生态与生物原因都是真实原因；但凡需要人持续选择、组织、服从、对抗、协同或承担的问题，就不能把人心、人性、人情、获得感、意义感以及个体与群体对存续和发展的感受当作外部噪声。机制如果只在“所有参与者都先克服利己”时成立，就尚未证明现实可达。

这不意味着所有问题都由人心单独造成。更准确的判断是：人的认知、利益、关系和意义结构，会参与决定或显著影响外部问题怎样被感知、分配、放大、协同处理或反复生成。

4.2 多维利己

维度	机制必须允许进入的真实动机
存续与安全	生命、身体、财产、组织延续、风险可控与未来选项
尊严与身份	体面、地位、承认、身份连续、群体归属与不被支配
关系与公平	互惠、信任、声誉、亲疏、人情、程序公平与相对获得
发展与能力	资源、知识、行动空间、竞争力、代际机会与掌控感
意义与价值	成就、使命、信念、责任、牺牲的可理解性与共同叙事

这些维度可以冲突、随阶段变化，也会被叙事、比较对象、身份关系和预期塑造。研究可以使用分维代理和行为证据，但不得把它们压成一个未经论证的总效用分数。

4.3 四条固定纪律

纪律	最低要求
非圣人条件	候选机制不得依赖关键主体持续牺牲自身核心利益、持续善意或持续压抑可预见反应
多维利益条件	收益、损失与不可接受边界至少按主体、维度、时间和可逆性分别记录
不可转嫁条件	主体 A 的净收益不能以未被计入的 B、第三方、弱势者或未来主体承担为成立依据

纪律	最低要求
双账条件	感知利益用于解释行为与可达性；真实后果用于判断正当性、分配和不可逆风险

“主体会追求其感知利益”是行为假设；“因此其行动正当”不是有效推论。利己帮助解释为什么会行动，不能消除行动造成的真实伤害与责任。

本项目不要求所有既得利益者都获得帕累托改善。依靠伤害、垄断、寻租或转嫁形成的利益可以被合法约束，不能为了购买阻挠者合作而再次向无权拒绝者转嫁代价。通过欺骗、胁迫、隐瞒或操纵感知形成的顺从，不得计作机制可达性证据。

4.4 多维利己必须可反驳

多维利己是机制可达性的设计基线和压力测试，不是解释一切行为的总定律。不得在看到结果后再增加利益维度，使任何行为都被事后解释为利己。每个工作包必须预先声明相关利益维度、可观察代理、竞争性动机解释和反证条件；多种解释无法区分时，动机归因保持 UNKNOWN。

利他、责任或牺牲不能仅靠改名为“意义利益”而被自动吸收。制度惯性、认知错误、情感依恋、道德承诺、规则服从与偶然事件都可能形成竞争性解释；研究必须允许它们推翻或限制利己解释。

4.5 利己递归适用于研究者和守门人

利己假设同样适用于竞争主体、研究团队、资助者、证据保全者、核验者、挑战者、认定者、正当性审查者、阶段门与状态执行者。职责名称不能替代激励分析；每个承重角色都必须记录利益来源、实际控制、退出条件、错误成本和责任后果。

递归利己不能通过无限增加更高层审查者解决，而要通过竞争性证据、实际控制披露、任期与轮换、异议申诉、外部独立挑战、计算可复现、独立再分析，以及在具备新证据时开展复制研究，并禁止单一主体端到端控制形成闭环。

五、主张等级与状态系统：不同语言不得互相冒充

5.1 主张等级

等级	准确含义	允许表述
方向	指出值得改变的结构或风险，不含机制成立结论	“需要研究” “可能存在结构入口”
机制假设	给出动作—中介—结果与竞争性解释	“若条件成立，可能产生……”
证据发现	报告特定样本、参数、模型或事件中的观察	“在冻结范围内观察到……”
条件性支持	所有承重假设在冻结范围内获得支持且无承重未知	“R2 级条件性支持，不具试点资格”

等级	准确含义	允许表述
研究候选机制	通过相应阶段的六重测试、正当性门、独立挑战与版本门	必须同时标明 R 阶段和授权状态

任何输出必须报告与其相关的四层状态：**窄事实核验状态、承重推断与有效性测试状态、正当性状态、研究运行状态**。第二层必须同时报告 H 状态、T1—T4、T6-R 与 T6-M；H 的条件性支持不得代替任何 T 状态。输出还须标明 R0—R6 研究阶段与现实授权状态。不适用的层级必须明示 N/A+理由；应检而未检只能写 UNKNOWN，不得通过省略或 N/A 制造已经通过的印象。不能把研究候选转写为政策建议或部署资格。

标准输出记录按固定顺序呈现：**主张 ID 与范围 → 文件版本与哈希 → R 阶段 → M 授权状态 → 窄事实状态 → 承重阻断、FAIL/UNKNOWN 总表 → 各承重 H → T1—T4/T6-R/T6-M → L0-R/L0-M/T5 → RUN 状态 → 最强反证 → 重开条件**。总表必须汇总：H 的 FAIL/UNKNOWN；T1—T4、T6-R、T6-M 的 FAIL/UNKNOWN；L0R-REDESIGN/LIMIT/FAIL；L0M:REDESIGN/LIMITED/UNKNOWN/REJECT；以及 T5 的 FAIL/UNKNOWN。它必须先于任何正面 H/T/L 结论展示；较早呈现的窄事实或授权核验不得被表述为机制正面结论。每个状态同时记录形成主体、证据窗口、版本和挑战状态；后续分项状态不得改变该顺序。各状态保持分列，禁止合成为总分、总灯号或“综合通过”。附录 F 只提供速查顺序，不创设第二套规则。

现实授权状态最低统一为 NO M-AUTHORIZATION/PILOT-SCOPE AUTHORIZED/OPERATIONAL AUTHORIZED。NO M-AUTHORIZATION 只表示研究尚未核验到满足本母本最低语义的现实授权，不是研究团队对现实法律效力作出的终局裁判。

使用 PILOT-SCOPE AUTHORIZED 至少必须有可核验的现实授权文件，并明确：授权主体、签发者、授权依据与权限链；获准活动、参与主体、地域或技术系统、数据与披露范围、风险上界和明确排除事项；生效、终止、自动失效和续期程序；M-D 与 M-S 的具体承载、责任主体与权限上限；启动、暂停、撤销、紧急停止、回滚和安全退出的触发条件、决定者与执行者；直接参与者的知情、同意、拒绝、退出与数据权利，以及无法取得第三方个别同意时的权限依据、保护边界与不得转嫁事项；证据保全、访问、日志、事件报告、独立 M-V/M-X、决定复审、申诉、修复与必要补偿；利益冲突与实际控制披露；授权文件版本与哈希；以及主体、范围、数据、风险或权限发生承重变化后的重新授权。授权文件必须明示不得把 M 平面权限转授给 R 平面或 Human-COS。

任一最低要件缺失、过期、超越授权主体权限、无法核验，或承重变化未重新授权时，均不得使用 PILOT-SCOPE AUTHORIZED。该标签只证明特定主体在特定范围和期限内获准试点；不证明机制有效、正当、可扩展、T6-M 通过或已获 OPERATIONAL AUTHORIZED，也不得自动续期或转化为运行授权。WP-01-A 只能提供空白字段和候选验证方案，真实授权只能由未来 R6 中有权承担后果的现实主体签发。

OPERATIONAL AUTHORIZED 不是研究标签的自然终点。只有在上述试点最低要件基础上，现实授权文件还明确常态运行范围、持续的法律与权限依据、运行能力与责任承载、独立 M-V/M-X、持续监督与事件报告、定期复审、申诉修复与补偿、自动暂停与撤销、终止或迁移方案，以及承重变化后的重新授权，才能报告该标签。现实授权由有权主体签发；R 平面只能核验并准确转述：R-E 保全授权文

件与后续记录，R-V 核验条款存在、范围、时效与权限链，独立 R-X 挑战遗漏、实际控制与重大变化。任一承重要件未核验、已失效或存在未关闭分歧时，不得使用正面授权标签。运行授权也不证明机制永久有效、永久正当、可普遍扩展或免于 H/T/L 复审，不得被工具、R 平面或运行机构自我授予。

5.2 四套状态各守一层

层级	状态	谁形成	不能代表什么
窄事实核验输出	SUPPORTED/CONTRADICTED/INSUFFICIENT/NOT VERIFIABLE	R-V; R-X 可挑战	不直接决定 H，不推断动机
承重推断与有效性测试	H: FAIL/UNKNOWN/CONDITIONAL SUPPORT; T1-T4、T6-R、T6-M: 同一三值状态	R-A 形成 H、T1-T4 及 T6 技术部分; R-L 形成 T6 权利-授权部分	H 不代替 T; 均不等于现实授权
正当性准入与复核	L0-R、L0-M、T5 的分层状态	R-L 依据冻结规范与证据形成	不证明因果机制有效
研究运行状态	RUN:PROCEED/RUN:LIMITED/RUN:HOLD/RUN:REOPEN (过渡记录) / RUN:CLOSE	R-G 决定或硬条件生效, R-S 机械执行	不停止任何现实主体或系统; REOPEN 本身不许可研究动作

INSUFFICIENT 与 NOT VERIFIABLE 通常使对应 H 保持 UNKNOWN; CONTRADICTED 是否使 H 失败，取决于预注册失败条件及相关性，不能自动越级。只有承重且未解决的 V/X 冲突，才使对应命题进入限域 UNKNOWN 或触发限域 RUN:HOLD; 无关或策略性异议不得获得无限停止权。

5.3 状态转换链

R-V 只判断证据与窄命题的关系，R-X 提供独立挑战。R-A 依据冻结的假设-测试映射和聚合规则，形成有来源、有版本、可挑战、可撤销的 H 状态与 T1-T4 研究状态，并分别评估 T6-R/T6-M 中暂停、退出、修复、回滚与重开接口的技术可操作性。R-L 形成 L0-R、L0-M 与 T5 状态，并分别评估 T6-R/T6-M 中的授权、申诉、责任与权利接口。R-A 与 R-L 对自身纠错接口的判断必须接受 R-X 或独立再分析，不能用自评证明 T6-R。R-G 只根据已经形成的 H/T/L 状态决定研究阶段，不生产证据、不形成测试结论、不修改阈值; R-S 只执行已形成的研究状态并追加记录。

T6-R 与 T6-M 均按预注册严格合取规则机械形成：任一承重分量 FAIL 则整体 FAIL; 无 FAIL 但任一分量 UNKNOWN 则整体 UNKNOWN; 全部获得支持才形成 CONDITIONAL SUPPORT。R-S 只负责记录合取结果，R-G 不得重新权衡或改写。现实授权只能由具有现实法定或协议权限的 M 平面授权主体签发、变更或撤销; R-E 保全授权文件，R-V 核验其范围、期限与有效性，R-A/R-L 只形成研究判断，任何 R 平面角色均不得生成、扩大或撤销现实授权。

禁止短路 | 核验输出不得直接生成“恶意”“背叛”“应当回应”或现实惩罚; 阶段门不得通过修改模型、证据或成功定义制造可推进状态。

六、六重有效性测试：六个并行维度

同一研究结果不能跨层升级。六重测试的编号用于追踪，不表示正当性只能在有效性之后介入。

测试	核心问题	最低输出与边界
T1 相对更优	相对现实基线与可行替代，候选状态是否降低风险或改善相关结果？	不得只与零成本理想比较；保留分配
T2 现实可达	主体能否从现实初态进入，而不是只存在理想终态？	先行动者暴露、启动条件、感知利益
T3 可维持	在误判、协议偏离、更替、冲击与第三方行动下能否持续或恢复？	稳定、恢复与重新进入必须区分
T4 可验且稳健	行为能否被有限识别；结论能否跨参数、模型与独立团队成立？	核验不等于动机认定；报告未知
T5 具正当性	是否转嫁成本、越过权利、遗漏主体、制造俘获或不可接受分配？	受 L0-M 与过程台账约束；按阶段通过
T6 可纠错	能否暂停、退出、修复、重开、重新授权并保留负结果？	停止与重启权分离；记录期限、申诉与反馈；现实授权另建

T6 必须分别报告两个状态：**T6-R** 审查研究过程能否暂停、纠错、保留负结果与重开；**T6-M** 审查现实机制是否具备经授权的暂停、退出、修复、申诉、回滚与重新授权结构。第十五章只直接建立 T6-R，不自动证明 T6-M。T6-R 通过不得补偿 T6-M 失败或未知；R6 以前只能形成与当前证据阶段相符的 T6-M 设计级条件性判断，不能声称已经过现实运行验证。

T1—T4、T6-R 与 T6-M 采用 FAIL/UNKNOWN/CONDITIONAL SUPPORT，T5 采用 FAIL/UNKNOWN/PASS。H 状态是判断 T1—T4 的重要证据输入，但 H 的 CONDITIONAL SUPPORT 不自动等于任何 T 已获得条件性支持；每个 T 的判断都必须列明所依赖的 H、反事实、计算可复现、独立再分析/复制研究、规范或纠错证据及其阶段上限。证据窗口、主体覆盖或规范依据不足时，T5 必须保持 UNKNOWN；“尚未发现伤害”不自动构成 PASS。

只有支持该资格的确证性证据取得于 L0R-PASS 明确允许的范围、L0-M 获准、T1—T4 在当前阶段均为 CONDITIONAL SUPPORT、T5 为 PASS、T6-R 与当前阶段所需的 T6-M 均获得条件性支持、没有承重未知，并完成相应阶段要求的计算可复现、独立挑战以及独立再分析/复制研究，才允许称为“该 R 阶段的研究候选机制”。L0R-REDESIGN/L0R-LIMIT/L0R-FAIL 范围内取得的材料只能支撑重构、范围界定、缺口识别或负结果，不能单独支撑正面候选资格；通过不安全、越权或不可审计研究取得的结果也不得支撑候选资格。R1 通常只形成“机制假设”或“结构候选”，研究候选机制资格最早在 R2 出现；若所谓 R1 结果已经独立满足 T2、T3 及相应 T4 要求，其实际成熟度应登记为 R2，不得以低阶段名称规避门槛。R6 以前一律不具有现实试点或部署资格。

T5 PASS 只表示在预先冻结的规范框架、主体范围、证据窗口和公开异议下，未发现硬失败或承重未知；它不是对机制普遍正当性的终极证明。规范框架变化、新主体出现或异议获得新证据时，必须重开。

七、正当性双门：研究行为与机制资格分别审查

符号说明：L0-R 中的 R 表示 Research Action，即研究行为安全门；L0-M 中的 M 表示 Mechanism Design，即候选机制正当性门。二者都由 R 平面的 R-L 形成，都不等于 R-G 形成的研究运行状态，也不构成 M 平面现实授权。R0—R6 中的 R 表示 Research Maturity，即研究成熟度。以下状态均使用全限定名，避免把门判定与运行状态混写。

7.1 L0-R | 研究行为安全门

L0-R 审查研究行为本身：数据取得、信息披露、模拟、实验、发布或研究互动是否会制造不可接受风险、侵权或现实升级。它是 R-L 形成的安全判断，不是执行命令。其状态为：

- **L0R-PASS：**研究行为在冻结范围内可实施。
- **L0R-REDESIGN：**先改变数据、范围、披露、授权或安全措施，再重新审查。
- **L0R-LIMIT：**只允许范围界定、缺口识别、历史性或非介入研究。
- **L0R-FAIL：**研究行为本身越过硬边界，形成限域 RUN:HOLD／CLOSE 的依据。

L0R-REDESIGN 暂停原定研究动作，完成重构并复审前不得继续；L0R-LIMIT 只允许状态中明示的范围界定、缺口识别、历史性或非介入活动；L0R-FAIL 禁止相应研究动作。只有 L0-R 各状态能够以研究安全、侵权、信息泄露或现实升级风险限制研究行为；R-G 仍可依预注册规则因证据、方法、治理或资源原因限制主张、阶段、版本或关闭命题／工作包，但不得借此删除安全、非介入式的负结果研究。L0-R 形成依据后，客观硬条件由 R-S 机械执行；存在解释、归因或权衡空间时进入 R-G 书面审查。R-L 不得自行执行停止或重启。机制本身不正当，不等于禁止安全研究；危险机制的形成条件、失败原因与负结果更需要被记录。

7.2 L0-M | 候选机制正当性准入门

L0-M 在选领域、定样本和投入大规模建模前，审查候选机制是否依赖成本转嫁、权利越界或不可逆负担。它不证明机制正当，只防止明显不可接受的设计获得“先做有效性、以后再谈代价”的路径优势。

状态	效力
L0M:ADMIT	主要主体与负担路径覆盖充分、无硬阈值越界、无 L0-M 承重未知时，允许在冻结范围内进入确认性研究；不代表 T5 已通过
L0M:REDESIGN	完成分配、范围、授权或修复结构重构前不得推进
L0M:LIMITED	只允许范围界定、失败机理或非介入式研究
L0M:UNKNOWN	只允许取得关闭正当性未知所需的证据
L0M:REJECT	不得继续把该设计推进为候选机制、试点或政策建议；负结果仍须保留

L0M:REJECT 关闭的是该版本设计的正面候选资格，不删除既有证据，也不构成继续研究的安全许可。在 L0-R 明确允许的范围内，可以继续不改变现实主体权利、激励、资源或行动环境的非介入式负结果

研究，包括证据保全与再分析、历史或比较研究、失败因果诊断，以及不增加现实实施能力的限域模拟与红队分析；这些活动不需要把原 REJECT 改写为 ADMIT。

REJECT 后不得招募现实主体实施该设计，不得扩大敏感披露或操作能力，不得生成试点、部署或政策建议材料，也不得以“研究失败原因”为名继续优化其正面候选资格。只要研究开始改变机制结构或参数以恢复可行性、引入现实介入、扩大操作能力，或谋求新的正面标签，就必须形成新版本并重新通过 L0-R 与 L0-M；原 REJECT 及其负结果永久保留。每个工作包必须记录 REJECT 后允许与禁止的研究范围。

L0-M 的“无承重未知”只指准入所需的主体、权利、不可逆负担和转嫁路径，不要求 H1—H5 在研究开始前已有答案。

7.3 L0-M 必问

检查	必须回答
主体覆盖	谁获益、谁承担、谁有否决／执行能力、谁可能因不在场而被遗漏
因果依赖	候选机制是否必须依靠某类主体承受成本才能产生正面结果
负担性质	可逆／不可逆；可补偿／不可补偿；可退出／不可退出；短期／代际
授权与修复	谁有权同意、拒绝或申诉；补偿、退出、纠错和责任能否真实到达

7.4 过程外部性台账与 T5

L0-M 通过后，每次模型、样本、范围、参与者和参数变化都要追加外部性台账，不得覆盖原记录。最低字段包括承受主体、直接／间接／延迟／网络路径、预测与实际负担、概率／强度／分布／时间、可逆与可补偿状态、退出条件、授权／申诉／修复、证据、未知、异议及是否改变 L0-M。

出现新主体、新负担类别、潜在不可逆风险、退出失效、负担集中或向未来主体转移时，只构成 L0-M 重开信号。符合预注册客观表达式时由 R-S 执行限域 RUN:HOLD；存在解释空间时由 R-G 根据 R-L 书面意见认定。R-L 重新形成 L0-M 状态，但不得自行控制阶段或执行。T5 在当前阶段对照 L0-M 与全过程证据复核；没有现实试点证据时，只能形成当前研究阶段的正当性判断，不能取得部署级正当性。T1—T4 的正面结果、已经沉没的研究投入或双边收益增加，都不能补偿 T5 失败或承重未知。

7.5 防转嫁，也防任意规范性否决

- 在看到结果前冻结适用规范、主体范围、负担分类和阈值；事后新增标准只能进入新版本。
- 正当性判断必须连接具体权利、伤害、分配、授权、责任和真实主体，不能使用无法反驳的抽象口号。
- 比较现实基线与可行替代，而不是用零成本理想否决任何改善；补偿不能抹除不可补偿权利或不可逆风险。
- 追溯收益来源：先主动制造威胁、封锁或人为稀缺，再解除部分伤害形成的“增益”，不得直接计作机制收益。

- 机制提出者承担未转嫁的举证责任；受影响者证据、独立审查、书面理由、异议和重开条件必须保留。
- L0-R 控制研究行为的安全边界；L0-M 与 T5 控制机制的研究候选资格、主张等级及试点推荐。三者均不取得对现实国家、企业、个人或基础设施的停止权。

7.6 规范形成与冲突：冻结的是可审计规范包

事前冻结的不是一条抽象价值结论，而是一份可审计规范包，至少记录：规范来源与适用范围、受影响主体及其输入、缺席主体的代理方式、权利与潜在不可逆硬边界、竞争性规范、冲突处理规则、未采理由、异议、修订与重开条件。R-L 形成 L0-M/T5 状态，但不得单方面选择规范前提；受影响者意见必须进入证据与理由链，也不自动获得无限否决权。

承重规范冲突无法依事前规则关闭时，相关 L0-M 或 T5 保持 UNKNOWN，不得用多数票、总体收益、资助方偏好、研究投入或强势主体身份压平。规范包的语义变化必须升版，并重开受影响的 L0-M/T5；具体参与主体、代表方式、程序与时限由工作包实例化。不同规范之间不存在可预注册的共同裁决规则时，研究必须并列报告冲突条件与各自后果，不得制造虚假统一答案。

八、计算纪律：按伤害维度分轨，跨轨不得抵销

8.1 不是“可逆事件／不可逆事件”的简单二分

轨道按伤害维度而非事件名称分配。同一机制可能同时产生可逆经济成本、不可补偿权利侵害和不可逆系统风险：只有可逆且可补偿的维度可以进入净值轨；不可逆、不可补偿或触及权利硬阈值的维度进入约束／阈值轨。跨轨不得抵销。

8.2 可逆且可补偿维度：净值、分配与反事实轨

交易、通报、供应链、军费增量、恢复时间等维度，在数据与识别条件允许时可以做净值、区间、反事实和敏感性分析。结果必须按主体、利益维度和时间保留分配，不能只报一个总体净值。

8.3 不可逆、不可补偿或权利硬阈值维度

核升级、文明级 AI 失控等风险在本项目中按潜在不可逆或极难逆维度处理，不与普通收益无差别货币化。研究使用损失保守下界、概率区间、不可接受阈值、压力测试和稳健决策。不可补偿权利不能因其他主体总收益增加而自动被覆盖；精确给文明存续定价既非必要，也可能隐藏规范前提。

8.4 所需风险下降阈值：从 P^* 到 $\Delta p_{i,r,t}^i$

历史稿中的 P^* 只作为旧称保留。其准确量是：对主体 i 、风险类型 r 、时间窗 t ，候选机制相对竞争基线至少需要实现的绝对风险下降量：

$$\Delta p_{i,r,t}^i = \frac{\max\{0, W_{i,r,t}(comp) - W_{i,r,t}(mech)\}}{L_{i,r,t}^{LB}}$$

其中， W 只含可补偿维度的结果， $L^{LB} > 0$ 是对应潜在不可逆损失的保守下界；分子与分母必须使用可比量纲。该阈值只回答“风险至少要下降多少，才可能抵过常规收益差”，不回答风险是否真的下降、机制是否可达、能否维持或是否正当。

该式只在候选机制主要改变风险发生概率，且 comp 与 mech 两种状态下事件发生后的条件损失可由同一损失尺度 $L_{i,r,t}^{LB}$ 表示时成立。若机制同时改变损失严重度、后果分布或承受主体，必须分别建模 $p_{comp} L_{comp} - p_{mech} L_{mech}$ ；无法建立可比尺度时输出 NOT COMPUTABLE。多个风险类型不得重复使用同一 W 差额，除非预先给出分配规则或联合风险模型。

- 主体、风险类型、时间窗与概率单位必须一致；不得跨主体合成全球统一阈值。
- 分子为负时阈值取零，只表示不存在该项常规收益障碍，不表示机制成立；max 是本项目“不可逆风险不由常规收益补偿”的纪律，不是 break-even 方法的纯代数结论。
- 可实现风险下降满足 $0 \leq \Delta p_{i,r,t} \leq p_{i,r,t}(comp)$ 。若 Δp^i 高于竞争基线风险概率的可信上界，标记 UNATTAINABLE；若不确定区间与可实现边界交叠，保持 UNKNOWN，不得把大于 1 或大于基线概率的数值解释为可达门槛。
- 无可信损失下界、因果风险差无法识别或单位不一致时，输出 NOT COMPUTABLE，改用稳健约束，不得强算。
- 权利、主体资格、不可补偿伤害和不可转嫁责任不得被放入 W 后压成统一货币分数。

该量只是阈值敏感性探针，不执行跨轨抵销；任何潜在不可逆、不可补偿或权利硬阈值已经越界时，本探针不适用，也不能产生许可效力。

九、R0—R6：研究成熟度递增，不是现实演化顺序

阶段	研究对象	主要产出	进入下一阶段的门
R0	研究宪章	主张等级、L0-R/L0-M、六重测试、双轨计算、权力分离、负结果规则	概念可操作；状态与权限不混写
R1	最小双边机制	单一领域、两个主要行动方、一个可检验结构改变	能检验其改变某个支付、信念或转移概率
R2	可达与维持	误判、偏离、恢复、先行动暴露与三值条件地图	存在现实进入路径并能标出失效区与未知区
R3	国内—国际双层	部门、利益集团、执行机构、否决者与内部获得感	不再依赖“国家是单一主体”
R4	同盟与阵营网络	成员异质、第三方外部性、牵连／抛弃与网络转换	网络变化下仍有条件性结果，不靠外部转嫁
R5	共同体与公共品	重叠成员、准入、贡献、退出、争议与反俘获	产生可观察公共品，且不靠永久封闭维持

阶段	研究对象	主要产出	进入下一阶段的门
R6	可逆现实试点	经授权的低风险场景、快速反馈、跨真实场景复制与可回退	反馈快于损害扩散；授权、责任、计算可复现、必要的新场景复制与退出成立

9.1 禁止越级，但必须允许下游反证筛查

阶段门限制的是正面结论等级和完整模型复杂度，不禁止为发现反例、外部性与遗漏变量而进行下游路径筛查。R1/R2 必须对国内否决者、盟友和第三方路径进行限域筛查，但不得据此形成 R3/R4 层级的正面主张。

若下游路径可能具有决定性，当前结论保持 UNKNOWN，并可开启只用于关闭该未知项的探索性高层模块；在前置阶段未通过前，不得把探索结果升级为高层机制成立、共同体形成或部署资格。这样既不以“禁止越级”为理由看不见第三方，也不以“做过筛查”为理由越级证明更复杂机制。

9.2 可复现与复制研究横跨各阶段

计算可复现是使用相同数据、代码和步骤得到一致结果；使用原数据但重新实现代码或分析流程属于独立再分析；复制研究是使用新数据或独立样本重新检验主张。独立挑战与计算可复现从 R1 起就是横向纪律；R6 新增的是经现实授权的可逆试点与跨真实场景复制，不是第一次允许独立复核。

9.3 阶段依赖与暂不开展事项

时序	工作	边界
当前主文阶段	冻结定义、主张等级、六重测试、双轨、状态与权限不变量	不实例化具体机构与阈值
母本锁定后	编制 WP-01-A 及其治理、编码、变量与输出附件	只实例化母本开放项
影子试运行后	比较传统流程与 Human-COS 对主体覆盖、错误记录与纠正、异议和计算可复现的增量	先测研究质量，不测现实“成败”
独立复核后	逐级进入双层、网络、公共品与低风险现实试点	遵守 R0—R6，不因仿真规模越级
暂不开展	把复杂关系压成单一综合评分的总指数、全球中央协调机构、高后果自动干预、大型全域仿真	依赖过多且易制造虚假精确与隐性权力

十、研究治理：认定权、停止权与重启权同样需要被约束

10.1 R 平面七种功能

具体机构和人员留待锁定母本后的工作包实例化，但功能、输出与不可兼任关系必须先由母本冻结。下表均为 R 平面研究权限；它们不是现实机制权限。

功能	只可承担	明确不得承担
R-E 证据保全	原始材料、哈希、来源、访问、版本和修订记录	判断合作／背叛；决定阶段状态
R-V 研究核验	对预注册窄命题给出四类证据输出	推断动机；直接决定 H 或现实回应
R-X 交叉挑战	同协议独立复核、反证、遗漏与替代解释	凭身份、票数或噪声异议消灭事实
R-A 后验认定	按冻结规则形成可挑战、可撤销的 H、T1—T4 及 T6-R／T6-M 技术可操作性参考标签	形成不可质疑的“事实中心”；形成 T5；兼任 R-G
R-L 正当性审查	L0-R、L0-M、过程台账、T5 及 T6-R／T6-M 各自的权利—授权部分	代替经验检验；用未预注册抽象价值任意否决；兼任 R-G
R-G 阶段门	根据既有 H/T/L 状态决定 RUN:PROCEED、RUN:LIMITED、RUN:HOLD、RUN:REOPEN 或 RUN:CLOSE	修改证据、阈值、模型或成功定义
R-S 状态执行	机械执行 R-G 已决定的版本冻结与限域 RUN:PROCEED／LIMITED／HOLD／REOPEN／CLOSE；客观硬触发只能机械生效限域 RUN:HOLD；追加记录	自行解释条件、删负结果、单独重启

R-A 不得把自身标签称为事实中心。它形成的是有来源、有版本、可挑战、可撤销的后验参考标签；只有由独立外部事实充分确立的窄命题，才可作为事实校准基准。事件事实、来源归属、主体意图、责任和现实回应必须分层，主体意图、责任与现实回应均不能由 R-A 越权认定。R-X 与独立再分析有权挑战 R-A，R-A 的改判率、遗漏与未决状态同样进入错误记录。

10.2 M 平面角色另行证明

WP-01 所说的有限核验属于 M 平面的候选机制构件。任何未来现实协议至少必须在权限和效果上区分：M-E 运行证据保全、M-V 操作性窄事实核验、M-X 跨方复核、M-D 现实决策授权与 M-S 现实行动执行。具体主体留待工作包实例化，但以下边界由母本冻结。

功能位不是制度蓝本。 M-E／M-V／M-X／M-D／M-S 只用于标明现实机制若进入设计或授权阶段必须回答的功能、权限、不兼任与证明义务；它们不预设五个机构、固定组织图、单一法域、层级关系或统一全球制度。任何把这些功能映射为现实主体、机构、程序或技术接口的方案，都是一个新的 M 平面候选设计，必须独立说明现实授权、责任承担、实际控制、申诉复审与退出，并重新通过 L0-M、T5、T6-M 及相应阶段门。现实安排可以使用不同名称、数量和组织拓扑，也可以增加必要功能，但不得消解母本冻结的权限隔离。R 平面治理完整、研究标签成熟或工具接口可用，均不得作为该映射已经可达、正当、可信或获授权的证据。

以下边界只约束未来候选设计不得怎样越权，不为任何现实主体产生职位、任命、组织资格或行动权限：

- M-V/M-X 只能形成预注册窄事实输出，不能直接命令回应、惩罚或停止。
- M-V/M-X 冲突时，M-D 只能依照预先授权的“不确定状态处置规则”采取限域、可逆行动，不得把 UNKNOWN 改写成事实。
- M-D 必须具有现实授权并承担决定后果；M-S 只能在授权范围内执行，并保留责任、申诉、修复和退出记录。
- R 平面角色、Human-COS 或研究标签不得代行 M-D 或 M-S；当前 PRE-HUMAN-COS 阶段只研究 M-D/M-S 规则，不启动现实回应。

M 平面最低不可兼任关系为：M-V 不得兼任 M-D；对他方产生约束、损失或权利影响的行动中，M-D 不得兼任 M-S。M-X 的独立性必须相对于该次复核的具体对象成立：它必须独立于被挑战的 M-E、M-V、M-D、M-S、机制提出方、被核验主体及其实际控制者，不能用对另一对象不存在冲突来替代本项独立。任何实际控制者不得同时控制 M-E 证据入口、M-V/M-X 事实输出、M-D 现实决定、M-S 行动执行及申诉复审。M-D 决定必须能够由另行授权、实际独立的主体复审；未建立复审、申诉和纠错链时，T6-M 保持 UNKNOWN，现实授权状态只能为 NO M-AUTHORIZATION。若具体领域无法满足最低分离，不得以“低风险”或“资源有限”为由把形式分工当作权力分离；只能缩小 M 平面范围、增加独立会签，或维持设计级研究。

若 M-X 与该次被挑战对象不独立，该复核不得计作独立事实或纠错证据；受影响的 T4、T6-M 或其他依赖状态保持 UNKNOWN，现实授权标签也不得把该复核列作已满足的独立要件。M-S、M-E 或其他未在固定短名单中出现的对象，不因角色名称不同而免于相对于自身行为与记录的独立挑战。

M 平面运行记录只能作为原始材料进入 R-E，再由 R-V/R-X 独立研究核验。M-V 不能给自己制造 R-A 标签，也不能通过“机制运行过”证明机制有效；R 平面也不得以研究结论直接指挥 M 平面的行动。

10.3 有限核验的四类输出

R-V 与未来 M-V 均只对各自平面内的预注册窄命题输出：

- **SUPPORTED：** 现有证据在冻结范围内支持该窄命题。
- **CONTRADICTED：** 现有证据与该窄命题存在承重矛盾。
- **INSUFFICIENT：** 证据量、质量或观察窗不足。
- **NOT VERIFIABLE：** 在当前协议、权限或可得信息下不可核验。

四类输出不能直接生成“恶意”“策略性背叛”或“应当报复”。因拒绝、隐瞒、延迟或破坏预注册证据接口造成的不可核验，只有在接口义务、主体控制能力和实际偏离均可观察时，才可记录为“可观察协议偏离”；证据缺失本身不能反推恶意。

通报数量、核验次数或一致票数不能替代证据质量、覆盖、机会窗和来源独立性。

10.4 最低不可兼任与实际控制测试

至少冻结以下绝对关系：

- $R-V \neq R-A$ ；实时或主研究核验者不能为自己的判断提供最终后验标签。
- $R-A \neq R-G$ ；形成 H 标签者不能决定是否推进自己形成的标签。
- $R-L \neq R-G$ ；正当性审查者不能单独取得阶段控制权。
- $R-G \neq R-S$ ；决定状态者不能自行解释并执行自己的决定。
- $R-X$ 必须独立于 $R-V$ 、 $R-A$ 及机制提出团队。
- $R-X \neq R-G$ ；挑战者不能根据自己的挑战决定阶段停止或重启。
- $R-A \neq R-L$ ；经验标签形成与正当性判断不得由同一实际控制者完成。
- 原模型或机制提出团队不得实质控制 $R-E$ 、 $R-V$ 、 $R-X$ 、 $R-A$ 、 $R-L$ 、 $R-G$ 或 $R-S$ ；尤其不得控制对自身项目的证据入口、标签形成、正当性判断、阶段决定、HOLD 执行或重启执行。

以上只是最低通用不兼任关系，不是独立性的封闭名单。独立性必须按具体主张、对象、事件、版本与挑战范围判断。 $R-X$ 、独立再分析或复制节点必须独立于该次复核所挑战的证据保全者、核验者、标签形成者、正当性审查者、阶段决定者、状态执行者、机制提出者、被研究主体及其实际控制者。未满足时，该复核不得计作独立证据线，相关 T4、T6-R、T6-M 或其他依赖状态保持 UNKNOWN；不得用另一项无冲突复核替代本项独立性缺口。

独立性按实际控制和共同依赖判断，不按机构名称、签字数量或形式分离判断。经费、任命与撤换、数据入口、模型／知识产权、权限凭证、人员与技术上游、判断与发布、以及停止与重开中的任何组合，只要足以决定证据、标签或状态，就构成实质控制风险；“没有同时掌握全部杠杆”不是安全港。

每个工作包冻结前必须提交实际控制矩阵，逐一映射名义承载者、最终受益或控制主体，以及经费、任命与撤换、数据入口、模型与知识产权、权限凭证、人员与技术上游、判断与发布、以及停止与重开权，区分直接控制、间接控制、共同上游依赖和事实依赖。每个控制关系都要记录能否发起、阻止、改变、延迟或撤销相关事项，以及依据、范围、期限、替代渠道与独立制衡。 $R-E$ 保全依据， $R-V$ 核验窄事实， $R-X$ 独立挑战； $R-G$ 只能依据已经形成的记录适用治理后果，不得自行补证。任何足以贯通不相容环节的控制路径未被排除，或关键控制关系不可核实时，T6-R 保持 UNKNOWN；若控制缺口同时破坏证据完整性，受影响的 H／T4 也保持 UNKNOWN 或按预注册失败条件形成 FAIL。相关材料不得取得正面候选资格。母本不使用“掌握若干杠杆即算控制”的机械计分；具体主体、隔离措施和替补安排由工作包实例化。

同一工作包、事件和适用范围必须另行形成一张合并 M／R 两平面的实际控制矩阵。被研究、被核验或可能依据研究结果行动的 M 平面主体及其实际控制者，不得控制评价其自身记录、决定或机制的 R 平面证据入口、核验、挑战、标签、正当性、阶段、状态执行或重开。分别命名、双重任职、内部防火墙、外包或同一集团下的形式分立不构成安全港。无法排除时，相关主体必须回避并启用实际独立的替补；受冲突链形成的状态不得继续使用。冲突触及 $R-L$ 或规范链时，原 L0-R、L0-M 与 T5 状态失去现行效力；这只是形成资格失效，不是 $R-G$ 、 $R-S$ 或自动规则重新形成 L0／T5。标准输出必须保留原标签、失效原因与“待独立 $R-L$ 重形成”；在实际独立的 $R-L$ 重新形成前，L0-M 与 T5 在聚合与运行判断中

按 UNKNOWN 处理，且因不存在有效的 L0R-PASS，默认限域 RUN:HOLD。新的 L0-R、L0-M 与 T5 状态仍只能由实际独立的 R-L 依据证据形成。冲突触及证据、挑战、标签、阶段或纠错链时，受影响的 T4、T6-R、T6-M 与研究候选资格保持 UNKNOWN。只有已经由独立、有效证据满足预注册硬失败条件并经独立复核的失败状态可以保留；受冲突链形成的 REJECT/FAIL 同样不得继续生效。冲突研究结论不得作为 M-D 的独立依据。现实授权是否存在仍按第 5.1 节单独核验，不由研究冲突自动生成、扩大或撤销。

早期资源有限时，同一机构可以承载多个非冲突功能，但人员、权限、凭证、日志、回避和决定授权必须分离，并公开剩余冲突。若最低分离无法满足，T6-R 保持 UNKNOWN；受影响的证据或测试状态按其依赖关系保持 UNKNOWN，不得以资源有限降低标准。

资源不足只能降低研究范围、阶段和主张等级，不能降低证据、分权、正当性或安全门槛。每个工作包必须事前声明形成正面主张所需的最低治理能力，包括可审计的证据保全、承重证据的独立挑战、不相容权限的实际隔离、L0 审查与申诉、追加式状态与版本记录，以及可执行的研究停止与重启。能力不足但仍可安全开展研究时，只能进入 RUN:LIMITED，从事范围界定、历史性、非介入式或负结果研究；相关 H/T 不得形成正面支持，尚未满足预注册失败条件时保持 UNKNOWN，已满足失败条件时仍保持 FAIL。在冻结研究窗口内没有可信补齐路径时进入 RUN:CLOSE。若证据无法安全保全、风险无法限域，或端到端控制无法解除，则进入限域 RUN:HOLD/CLOSE。只是 M 平面承载、授权或纠错能力不足时，维持 NO M-AUTHORIZATION 和设计级研究，不必关闭安全的 R 平面研究。具体人员、预算、时限与替代安排由工作包实例化。

10.5 预接受、错误记录与递归闭环

必须记录谁预先接受了什么程序：接受主体、程序版本、适用范围、证据接口、冲突处理、退出/重新接受条件与时间戳。预接受不等于放弃质疑；程序失灵、证据变化或权力俘获时，仍可按规则重开。

- 资格、报酬、续约与声誉不得绑定“发现偏离/背叛”或“证明合作成功”。
- 任命、资金、数据、知识产权与职业依赖必须披露；挑战方应获得实质对称的协议、证据接口和申诉条件。
- 被核验者与核验者都进入追加式记录，不能只让被核验者永久留痕。

R-E、R-V、R-X、R-A、R-L、R-G 与 R-S 都保留追加式决定、错误、纠正与后果记录。每条记录至少包括原命题、证据快照与来源、时间、输出、后验标签及形成来源、误报/漏报或其他错误、纠错时长、现实伤害、错误来源、自报或外部发现、修复与资格后果：

- R-E：证据遗漏、材料丢失或篡改、保全延迟、来源断裂、访问越权、哈希/版本不一致、选择性接收、修复时间及其后果。
- R-V：误报、漏报、UNKNOWN/NOT VERIFIABLE 率、覆盖率与纠错时延；只有未来输出概率时才使用概率校准。
- R-A：改判、遗漏、未决、标签来源与被挑战记录。
- R-L：遗漏主体、规范漂移、利益依赖与俘获风险。

- R-G：误停、漏停、选择性豁免、理由与复审结果。
- R-S：执行延迟、越权、选择性执行和记录完整性。
- R-X：无根据挑战、策略性阻断、遗漏与被反驳记录。

“长期保持未知”本身只触发程序与证据努力审计，不能逼迫核验者作肯定判断。主动报告错误并快速纠正的预期成本，应低于隐瞒、拖延或被外部发现的成本；任何角色都不得自行证明自身可靠。

M-E、M-V、M-X、M-D 与 M-S 同样必须保留追加式决定、错误、申诉、纠正、现实后果与责任记录；M 平面主体不得仅以“已经获得授权”证明其判断可靠。

十一、研究组织、证据线与事件数据

11.1 组织分工不替代权限分离

研究分工	主要职责
问题与边界组	研究问题、规范前提、潜在不可逆边界与允许结论
证据与识别组	历史数据、反事实、选择偏差、竞争性解释与不确定性
形式与仿真组	模型、参数、敏感性、代码与计算可复现
红队与异议组	战略利用、替代解释、俘获、信息泄露和失败条件
主体与正当性审查	L0-R/L0-M、过程台账、受影响者证据、责任与 T5
独立再分析／复制节点	在不依赖主团队内部过程的条件下再分析或用新证据检验

这些是组织分工，不自动等同于 R-E/V/X/A/L/G/S。一个组织可能承担多个研究任务，但权限是否相容、实际控制是否隔离，必须另外审查。

11.2 四条证据线相互制约

1. **历史事件线：**建立事件数据与编码手册，记录初态、信息、通报、运行核验、回应、结果、遗漏与未知。
2. **比较识别线：**比较存在与不存在相关装置的可比事件，同时处理制度选择偏差、共同冲击、反向因果和样本可比性。
3. **形式模型线：**把重复或随机互动、状态转移、不完全信息、误判、偏离诱惑、回应时滞和恢复显式化；模型均衡不能自动推出真实可达。
4. **红队与复核线：**独立寻找信息泄露、诱骗、自动承诺、第三方套利、国内俘获、错误归因和无法证伪之处。

任何单一证据线都不能自证。历史案例证明某种装置曾存在，不等于证明其降低总体风险；形式模型发现一个均衡，不等于现实主体能够进入；预注册减少事后调参，不会自动修复错误变量、错误因果图或无效数据。

11.3 事件数据最小字段

- **T0:** 事件发生前各主体可合理获得的信息、能力、约束、预期与版本状态。
- **主体与角色:** 正式权力、事实影响、执行、M 平面核验、R 平面研究角色、承担后果和第三方位置。
- **事件与信号:** 触发器、时间戳、信息来源、可靠度、缺失、异议与证据保全状态。
- **机制动作:** 通报内容、披露范围、M 平面核验方式、条件性回应、期限、复审、退出与修复。
- **结果:** 误判、协议偏离识别、升级、泄露、恢复、外部性、分配和后续更新。
- **反事实与竞争性解释:** 没有该机制、延迟通报或不同披露范围时可能发生什么；其他解释怎样区分。
- **H3/H5:** 暴露向量、机会窗、第三方主体与路径、系统安全阈值、覆盖和未知。
- **正当性:** L0-M 主体、过程外部性、权利、授权、申诉、修复和 T5 变化。
- **治理与版本:** R-E/V/X/A/L/G/S 的决定、错误、HOLD、重开、异议、数据／代码版本与父版本哈希。

11.4 敏感信息边界

公开研究只从经 L0-R 审查的低披露、可回退场景起步。任何涉及军事脆弱性、关键基础设施可利用缺口、现实行动时窗、人员身份或其他高后果信息的材料，必须另建访问、最小披露、责任、审计和退出规则；无法在安全边界内处理时，相关研究保持 RUN:LIMITED 或 RUN:HOLD。

十二、首个工作包 WP-01：危机通报—有限核验—条件性回应

12.1 选题与边界

WP-01 不从“全面合作”或敏感军备透明起步，而研究一个低披露、可回退、具有可观察反馈周期的候选机制：危机通报、有限运行核验与比例、可逆的条件性回应，能否降低误判升级，同时不显著增加先行者被利用、第三方反转和不正当转嫁的风险？

公共卫生、非敏感 AI 安全事故、灾害资源中断、基础设施故障等只能作为候选领域名称，不天然等于低风险。具体领域必须先通过 L0-R 与 L0-M，并证明数据披露、观察窗和现实后果处于可接受边界。

字段	WP-01 母本口径
研究单位	单一领域、两个主要行动方；允许预先接受的 M-V 或同协议交叉核验
目标变量	通报时延、误判概率、协议偏离分类、升级强度、信息暴露、恢复时间与第三方系统风险
核心装置	触发、最小披露、时间戳、M 平面有限核验、比例可逆回应、误报修复、复审退出
核心输出	条件性支持区／失效区／未知区；同时报告四层基础状态、T1/T2/T3/T4/T5/T6-R/T6-M、R 阶段与授权状态

字段	WP-01 母本口径
硬边界	不自动报复、不作动机归罪、不让研究 HOLD 变成现实停止

12.2 最小因果链：仍是待检验假设

1. 预先定义的可观察事件触发通报，并留下运行证据保全记录。
2. 主体只披露处置所需的最小信息和可核验时间戳，不要求全量能力透明。
3. M 平面有限核验只对协议偏离、伪报、隐瞒或证据不足作窄事实分类，并保留交叉核验冲突。
4. M-V／M-X 只形成窄事实输出；当前工作包不形成可直接触发现实行动的“策略性背叛”认定。未来若确需此类标签，必须另建独立、多源、可申诉的 M 平面认定与授权程序。
5. 候选条件性回应只作为待检验规则；任何现实回应必须由另行授权的 M-D 决定、M-S 执行，并保持比例、可逆、有期限、可申诉和可修复。误报、错误核验与局部违约都须保留修复和重入路径。
6. 结果进入 R 平面，更新行为假设、核验者记录、外部性台账、触发器、披露范围、阈值与下一版本；不得静默覆盖原记录。

这条链只是机制假设。任一步没有可观察中介、无法识别因果或依赖不现实条件，整条主张都不得升级。

12.3 WP-01 承重假设

假设	检验问题	主要观测
H1 误判下降	通报与有限核验是否降低把非敌意事件误读为敌意行动的概率	时延、分类错误、升级强度、反事实
H2a 识别效能	是否提高对可观察协议偏离、伪报或隐瞒的分类准确度	误报、漏报、覆盖、纠错时延
H2b 激励效应	识别能力变化是否真实降低隐藏偏离的预期收益或发生率	行为变化、机会窗、替代策略
H3 先行动者非劣	先通报方的安全、议价、信息、经济和国内暴露是否未超过预注册上界	暴露向量、界值、机会覆盖
H4 可恢复	误报、局部违约或外部冲击后是否存在修复与重入路径	恢复时间、修复率、再进入成本
H5 第三方不反转	第三方和网络反应是否未抵消双边降风险，且未越过潜在不可逆系统安全阈值	路径覆盖、净系统风险、安全阈值

H2a 或 H2b 任一 FAIL，H2 整体 FAIL；无失败但任一 UNKNOWN，H2 整体 UNKNOWN；只有 H2a 与 H2b 均获得条件性支持，H2 才获得 CONDITIONAL SUPPORT。H2a 的识别改善不能直接推出 H2b 的激励变化，H2b 的行为变化也必须排除替代原因。H5 的子假设同样实行严格合取。所有 H 都必须预先定义支持、失败、未知、现实基线、反事实、观察／利用机会窗、误差范围和竞争性解释。

12.4 预先承认的失败结果

- 通报没有显著降低误判或升级。
- 核验收益被信息泄露、试探、规避或策略性利用抵消。
- 机制只在不现实的高检测率、高未来权重、近乎对称能力或极窄参数区间成立。
- 国内否决者、盟友承诺、供应商、平台或其他第三方行动足以反转结果。
- 机制依赖不可接受成本转嫁、缺少退出修复，或研究本身无法在安全边界内进行。

任何失败结果成立，都应产出“未获支持”“适用范围收缩”“LOM:REJECT”或“研究保持 RUN: LIMITED/RUN:HOLD”，不得通过调整样本、参数、阈值或代理而不升版本，把合作调成胜出。

十三、H3 与 H5：对“没有发生”的主张采用更严格证据

13.1 H3 | 安全上界／非劣逻辑

H3 不能因“没有观察到被利用”就获得支持。它借用并改造非劣性检验的逻辑，为战略暴露问题建立分维安全上界；这不是临床方法在本领域已经获得外部验证，具体界值必须由领域证据、可接受风险和受影响主体理由共同支持。

要素	WP-01-A 必须实例化
暴露向量	安全、议价、信息、经济、国内政治、声誉与后续行动空间；关键维度不得互相抵消
界值 δ_k	每一维预注册安全上界、测量误差和严重事件触发器
反事实	无通报／延迟通报／不同披露范围的可行对照
机会覆盖	对方必须有足够利用机会和观察窗口；没有机会不能证明安全
证据最低线	至少两条相互约束证据线；区间跨界或覆盖不足即 UNKNOWN

“充分机会窗”不能只由经过的时间或一次“有／无机会”判断表示。WP-01-A 必须为以下承重环节预注册代理、证据源与竞争性解释：暴露是否真实产生；潜在利用者是否能够接触并识别；是否具备行动能力；是否存在达到现实门槛的动机或收益；国内、法律、资源、联盟或第三方约束是否足以阻断行动；延迟利用窗是否已经成熟；研究接口能否观察利用尝试、未遂行动与延迟后果。任一承重环节无法区分时，“未发生利用”不得支持 H3。若机制主张本身是通过消除暴露、降低可识别性、限制利用能力或改变利用收益形成保护，则该保护链必须作为中介另行预注册并取证，不能由零事件反推。

竞争性解释只有在同时给出明确因果机制、与冻结证据相容、若为真足以改变 H3 状态或适用范围，并指明可区分观测或说明为何结构性不可核实时，才构成承重异议。未达到门槛的解释进入残余异议台账，不单独阻断；达到门槛且结构性不可区分的解释使 H3 保持 UNKNOWN。R-X 有权提出承重异议，但不能靠无限增加无判别路径的假设取得永久停止权。

H3 状态规则：任一关键暴露维度触发严重事件触发器，或依据预注册测量误差与区间判定规则确认超过 $\delta_k \rightarrow \text{FAIL}$ ；证据在预注册机会链与观察窗内排除越界，或独立支持预注册保护链确已阻断暴露且其

他承重暴露维度未越界 → CONDITIONAL SUPPORT；仅有零事件、机会链任一承重环节未知、样本不足、代理失真，或区间与 δ_k 交叠而不能确认是否越界 → UNKNOWN。安全性负信号可以提前停止，正面支持不得在机会链和延迟观察窗达到前提前形成。若某种外生约束被明确冻结为适用条件，支持只能成立于该窄条件，不能把外生约束的作用归因于机制，也不得向条件外外推。观察结果后才形成的新判别规则属于探索性修复，除机械错误纠正外，不得用同一批已经看过的数据重新确认 H3；确认性判断需要保留样本、新数据、后续独立窗口或独立复制证据。

WP-01-A 必须预注册 H→T 依赖图。H3 是 T2 “现实可达”的必要证据条件；H3 为 FAIL 或 UNKNOWN 时，T2 不得获得条件性支持。若延迟暴露对维持性构成承重影响，还必须预先声明其向 T3 的传播关系，不能在结果出现后选择是否计入。

13.2 H5 | 路径覆盖、结果不反转与系统安全阈值

子假设	经验检验要求
H5a 路径覆盖	按能力×动机×机会×后果冻结纳入阈值、排除理由、直接／间接／延迟／网络路径和观察窗
H5b 结果不反转	已识别的第三方与网络反应没有把双边降风险反转为更高系统风险
H5c 系统安全阈值	第三方路径未越过预注册的潜在不可逆系统安全阈值

第三方包括盟友、伙伴、供应商、平台、国内否决者、地方与执行机构、受影响群体和可转换成员。“没有观察到反应”只有在主体、能力、机会、路径和时间窗覆盖充分时才有证据意义。

H5a 的“覆盖充分”是一项有界的程序性判断，不是穷尽所有可想象路径。WP-01-A 必须在看到结果前冻结主体域、路径生成方法、检索来源类别、直接／间接／延迟／网络观察窗、纳入阈值、实质性阈值、排除理由与检索停止条件；只有这些程序已经执行且达到最低信息量，H5a 才可在该版本化范围和时间窗内获得支持，不能被表述为普遍无第三方风险。

冻结后提出的新路径，只有同时具备可信性、实质性与边界相关性时才重开相关 H5 子项：可信性要求能够指明主体、行动、可检验因果连接与至少一个可审计证据锚点；实质性要求该路径在可信参数范围内可能改变 H5b／H5c，或揭示预注册覆盖方法系统遗漏高后果路径；边界相关性要求它属于冻结设计范围，或足以证明原边界划定不当。潜在不可逆路径可以采用更低但非零的纳入证据门。达到门槛且已经触发失败条件时按 FAIL 处理；尚未触发失败条件时，H5 在关闭该路径前回到 UNKNOWN，并回退所有依赖该支持标签的现行状态。未达到门槛的路径进入残余路径台账，记录排除理由、证据缺口、复审时间与重开条件，不自动重开，也不得被删除；残余路径的累积效应只能按预注册聚合规则形成重开信号。单一推测不得取得无限否决权。

新路径若在原冻结的生成域与观察条件下本应被发现却被遗漏，现行 H5a 应为 FAIL，并重开与覆盖稳健性相关的 T4；若路径源于后来出现的主体、技术或环境变化，原结论只保留为历史版本内的有界判断，现行版本 H5a 回到 UNKNOWN。两种情形都不得静默覆盖旧标签，也不得把旧版本继续作为当前正面支持。WP-01-A 必须预注册 H5 各子项向 T1／T3／T4 的依赖映射；R-G 不得凭“整体判断”决定是否传播状态后果。

H5 只经验性检验第三方路径是否反转风险、是否越过预注册系统安全阈值。该阈值并非价值中立：领域证据负责给出风险基础与测量边界，R-L 负责审查其中的权利、分配、不可补偿性和规范前提，R-A 只依据冻结规则判断是否越界。H5 越界判断不能替代 L0-M/T5 的正当性判断；L0-M/T5 通过也不能计作 H5 的经验支持。H5 发现新主体或新负担时，先按 7.4 形成 L0-M 重开信号；达到预注册触发条件后，再依第十五章状态链重开，不得由路径提出者直接改变 L0-M 或研究运行状态。若新路径同时改变既有规范包、权利或负担边界，T5 也必须按其形成规则重开。

H5a/H5b/H5c 中任一 FAIL → H5 FAIL；没有失败但任一 UNKNOWN → H5 UNKNOWN；全部在冻结范围内获得支持 → H5 CONDITIONAL SUPPORT。

H5 的正面支持不得在预注册第三方反应窗、路径覆盖和最低信息量达到前提前形成。

十四、严格合取与条件地图：未知不能被正面证据冲掉

14.1 WP-01 的 H 状态合取

合取状态	规则	允许动作
FAIL	任一承重 H 或其承重子假设失败	整体 FAIL；收缩范围、重构或关闭
UNKNOWN	无失败，但任一承重 H 仍未知	整体 UNKNOWN；只允许关闭未知的研究
CONDITIONAL SUPPORT	所有承重 H 在冻结范围内获得支持	仍须另行通过 L0-M、T5、T6、阶段门与独立复核

不允许计票。H1、H2、H4 的正面证据不能补偿 H3 或 H5 的 UNKNOWN；4/5 通过不是“多数支持”。WP-01 的 H1—H5 采用严格合取。其他工作包必须预注册自己的逻辑聚合结构：必要条件实行严格合取；替代性充分路径可以预注册为析取或嵌套结构。任何正面主张所依赖路径中的承重 FAIL 或 UNKNOWN 都不得被计票冲销；观察结果后修改聚合逻辑必须升版。H1—H5 只属于 WP-01，不能被扩成所有机制的永久假设集。

14.2 条件地图不做单一总分

WP-01 的核心输出是一张条件地图。首次出现的“条件性支持区”只表示：在冻结主体、参数、证据窗口和假设范围内，承重 H 获得条件性支持；它不等于候选机制，更不等于现实成立区。

每个条件组合至少附：四层基础状态、T1/T2/T3/T4/T5/T6-R/T6-M、R 阶段、现实授权状态、参数范围、最强反例、外部性、敏感信息边界和重开条件。任何承重未知都必须先于正面结果公开。

十五、停止、重启与负结果：研究不能控制现实

15.1 完整状态链

研究状态按下列链条流转：

1. 任何主体可以报告风险信号；信号报告本身不是正式触发器。
2. R-E 只确认材料存在、完整性、来源、版本和接口状态，不作实质停止判断。
3. 预注册客观表达式只有在输入字段、证据完整性和计算规则均已冻结时，才可形成 OBJECTIVE_TRIGGER。涉及事实值或分类的输入须由 R-V 按协议确认；任何输入争议、归因或权衡空间都使该条件转入 G_REVIEW，R-S 不得自行解释。只有同一组经核验输入依据冻结规则必然产生唯一结果，且触发范围与执行动作均已预先确定、不需要归因、价值排序、比例权衡或自由裁量时，才可标记为 OBJECTIVE_TRIGGER；把解释性判断预先藏入字段名称、分类标签或阈值来源，不会使其转化为客观触发器。
4. 客观硬条件由 R-S 机械执行限域 RUN:HOLD，R-G 不得豁免；解释性条件由 R-G 回避冲突后书面认定，并保留异议与独立复审。
5. RUN:HOLD 期间只允许经 L0-R 另行确认安全的证据保全、反证、修复和关闭未知活动；L0R-FAIL 时原则上仅保留不会延续风险的最低证据保全与责任记录，其他研究不得以“关闭未知”为由继续。
6. 补救证据先进入 R-E；窄事实由 R-V 复核，并且至少经过一条独立的 R-X 或独立再分析证据线。需要改变 H 或 T1—T4 标签时，由 R-A 重新认定；涉及主体、转嫁、权利、规范或 T6 授权接口时，必须由 R-L 重审。任何承重冲突未关闭，相关状态继续保持 UNKNOWN 或限域 RUN:HOLD。
7. R-G 在独立复核完成、利益冲突回避并保留异议后，书面决定是否重开；R-S 以新运行／工作包版本执行，并保留原异议、父版本与哈希。重开不自动恢复原正面标签，不得超过 HOLD 前的主张范围，且必须重新通过受影响的全部门。

RUN:REOPEN 只是一项一次性、版本化的过渡记录，不是独立研究许可。R-S 开立新版本时，R-G 必须同步依据本节规则为新版本赋予 RUN:PROCEED／RUN:LIMITED／RUN:HOLD 之一；尚未重新通过受影响的 L0／H／T／实际控制／阶段门，或未同步赋值时，默认 RUN:HOLD。RUN:REOPEN 本身不得授权数据取得、分析、发布、阶段升级或附件派生，不得恢复旧正面标签、扩大旧范围或绕过任何门。完成新版本登记与目标状态赋值后，REOPEN 记录即结束；RUN:CLOSE 后的重开同样适用。

RUN:HOLD 必须限于具体命题、版本、模型、数据集、发布主张、阶段升级或试点建议，不因局部冲突冻结无关研究。

R-G 选择运行状态时必须适用同一规则：所有当前研究动作均为 L0R-PASS，L0-M 为 L0M:ADMIT 或经记录不适用，T5 不存在承重阻断，最低治理能力达标，不存在待关闭的承重 FAIL／UNKNOWN 或有效 HOLD 触发，且拟开展工作不超过冻结阶段、范围与主张等级时，才使用 RUN:PROCEED。现有研究动作经 L0-R 允许，且存在有界、可审计的关闭未知计划；或处于 L0R-LIMIT、L0M:LIMITED／UNKNOWN／REJECT／T5 FAIL 或 UNKNOWN，但只拟在各自允许的范围内开展负结果保全、失效诊断、独立再分析／复制、反证或关闭未知；或已形成 H／T FAIL 后只拟开展上述安全负结果活动及新版本非介入式重构准备，且均不升级原版本正面主张时，使用 RUN:LIMITED。L0R-REDESIGN、L0M:REDESIGN、L0R-FAIL，继续现行协议会增加风险暴露、延续无效确认过程、需要先重构安全或治理边界，或正在违规升级正面主张时，使用限域 RUN:HOLD；其中 L0R-FAIL 原则上只保留不延续风险的最低证据与责任记录。在冻结的时间、资源、安全与证据可得范围内已经没有可行的状态区分

路径时，依 15.4 使用 RUN:CLOSE。状态选择、范围与理由必须书面化并可被 R-X 挑战，不得由 R-G 以“综合判断”自由替换。

RUN:HOLD 的准确含义 | 这里只暂停研究状态、正面结论、调参升级、版本派生或试点建议。它不暂停政府、企业、个人、金融、军事、基础设施或其他现实行动，也不赋予研究团队或 Human-COS 现实主权。

15.2 预注册停止条件

- 关键数据无法审计，或证据入口被单一利益方控制而无法挑战。
- 承重竞争性解释无法区分，或结论只在单一乐观参数、窗口、样本或模型下成立。
- H3、H5 或其他承重假设缺乏预注册覆盖，状态仍为 UNKNOWN，却拟升级正面结论。
- 主要相关主体被遗漏，或过程台账发现候选机制依赖成本转嫁。
- 潜在不可逆风险被普通净值加总稀释，或越过预注册系统安全阈值。
- R-V/R-X 的承重冲突被多数或身份压平，角色错误记录被隐瞒，或计算可复现持续失败。
- 原团队调参、换窗口、删样本、换标签或切换代理而不升版本，或事后放宽 L0/T1—T6 条件。
- M/R 平面混用，使研究标签拟触发现实惩罚、报复、政策行动或基础设施停止。
- Human-COS 或功能等价工具的影子运行触发第 17.3 节的审计、偏差、泄露、中心化、责任逃逸或不可停止条件，却仍拟扩大工具用途、保留未经复核的正面输出或由工具自行重开。

WP-01-A 必须逐项标记 OBJECTIVE_TRIGGER 或 G_REVIEW。任何无法机械表达的条件不得伪装成自动停止；任何客观硬条件也不得被阶段门以“综合判断”豁免。

15.3 负结果与版本纪律

- 发现失败区、未知区、不可接受转嫁或无法核验，本身就是正式研究产出。
- 机制只在不现实的高检测率、高未来权重、近乎对称能力或极窄参数区成立，应记录为未获现实支持。
- 探索性调整必须与确认性结果分离，保留原假设、原代码、原数据、原阈值和差异说明。
- 不得通过删除样本、缩短窗口、放宽界值、切换代理或改变排除标准而不升版本，把合作调成胜出。
- 计算可复现或原数据独立再分析出现方向性矛盾时，T4 FAIL；因权限、数据或说明不足无法复算时，T4 UNKNOWN。
- 使用新数据或新独立样本的复制研究得到方向性矛盾时，原支持标签必须重开，并按预注册相关性判断 T4 FAIL 或 UNKNOWN；不得用“方法不同”自动排除。

15.4 长期 UNKNOWN 与 RUN:CLOSE

UNKNOWN 不是无限投入命令，时间流逝也不能自动关闭 UNKNOWN。每个承重未知项必须预注册可行证据线、访问与观察窗口、阶段资源边界、预期信息增量、风险上限、复核节点、责任主体和重开条件。

经 R-V 核验、R-X 或独立再分析挑战，并由 R-A 或 R-L 按命题性质确认后，满足以下条件时，R-G 可以书面决定对指定命题、版本或工作包执行 RUN:CLOSE：当前安全、合法且现实可得的证据线已经实施或有书面不可行证明；剩余关键证据具有结构性不可得性，或继续投入已不能在冻结资源与风险边界内产生足以改变状态的合理预期信息增量；不存在不放宽失败条件、不遗漏主体、不增加不可接受风险的更窄设计；关闭理由、未检验解释、残余风险、已耗资源和未来重开条件均已公开记录。期限或预算耗尽通常只构成 G_REVIEW 信号，不能自行改变证据状态。

RUN:CLOSE 只关闭当前研究运行、资源投入与版本派生，不关闭未知本身。R-A 不得因项目关闭改变 H 或 T1—T4，R-L 不得因关闭消除规范未知；标准输出必须继续保留相应 FAIL 或 UNKNOWN，不得改写为不存在风险、机制安全、机制失败、默认反对或默认许可。出现预注册的新证据、新方法、新权限或关键环境变化时，方可按原状态链申请重开；重开必须升版并重新通过受影响的 L0、证据、实际控制和阶段门，不得自动恢复旧正面标签。

十六、风险登记：研究自身也可能制造风险

风险	表现	控制
虚假精确	复杂主体被压成总效用、总指数或单一阈值	分层输出；保留分配、区间、未知与异议
规模诱惑	过早建立全域多智能体仿真	R0—R6；复杂度一次只增加一层正面主张
方向偷换	把“更优”升级为“可达／稳定／正当”	六重测试、主张等级、四层状态与阶段／授权标记
双边外推	把双边原型直接扩成两大阵营	下游反证筛查；R3/R4 正面门
核验权垄断	同一控制者掌握证据、标签、阶段和重开	M/R 分面、不可兼任与实际控制测试
不可核验性武器化	通过拒绝接口制造未知，或以“未知”永久免责	区分可控制偏离与真实不可得；审计证据努力
状态混淆	SUPPORTED、UNKNOWN、HOLD、ADMIT 相互替代	四层状态及转换权固定
规范俘获	L0/T5 被强者、资助者或抽象价值垄断	事前冻结规范、受影响者证据、异议与重开
Human-COS 中心化	把现实主体降格为工具输入	先传统研究、影子运行；Human-COS 与功能等价工具不得承担 R-A/L/G/S 或形成正当性状态
安全外溢	研究泄露敏感弱点或扩大对抗	L0-R、最小披露、限域访问、无法安全则 HOLD
静默调参	换样本、窗口、阈值或代理而不升版	探索／确认分离、父版本、差异与哈希
平面渗漏	研究 HOLD 或标签转化为现实行动	R 平面无现实停止权；R6 另建授权架构

十七、Human-COS：未来研究工具，不是机制或认定中心

17.1 试运行前：可以准备，不能预写增益

可以完成研究宪章、机制卡、历史事件库、传统因果图、预注册假设、角色分离和红队问题；可以定义未来需要的主体、T0、证据、异议、来源、版本、可见性和责任字段。

不可以宣称 Human-COS 已经提高预测、协调、合作或政策质量；不可以为了适配未来工具，把现实主体、制度和专业方法降格为“输入变量”；不可以把 Runtime 的存在写成机制已可运行。

17.2 影子试运行：先验证研究增量

Human-COS 首先只与传统研究流程并行，不改变现实行动。比较对象不是“是否促成合作”，而是它是否提高主体覆盖、T0 纪律、因果链完整度、参数估计与更新透明度、来源追溯、异议可见、错误校正、计算可复现和更新速度，同时是否引入新偏差、中心化、信息泄露或责任逃逸；只有系统实际输出概率时才讨论概率校准。

只有工具本身可审计、可替换、可停止，并且未垄断方向、证据、认定、阶段门和状态执行，才可扩大研究用途。

17.3 影子运行的停止、回退与关闭

每项 Human-COS 影子协议必须在运行前把下列风险映射为 OBJECTIVE_TRIGGER 或 G_REVIEW，并限定到具体工具版本、模块、数据域、输出或研究主张：Human-COS 直接或经自动接口生成、改变或执行 R-A/R-L/R-G/R-S/M-D/M-S 权限内的决定；形成或改变 L0-R、L0-M、T5 或 T6 的权利—授权分量；作为 R-V/R-X 或 M-V/M-X 的唯一或最终承载，或借模型分身、代理链、自动化服务或共同提供者冒充独立复核；关键输入、输出、来源、版本、提示、模型变更或日志无法审计、复算，或者被覆盖、选择性删除；发生未授权访问、敏感信息泄露、用途扩张或无法回收的权限扩散；相比传统对照流程，持续出现承重主体遗漏、单一视角偏置、异议压平、错误隐藏或责任逃逸；资助者、工具提供者或单一控制者能够借模型、数据、凭证、知识产权或服务中止权控制证据、认定、正当性、阶段或重开；工具无法被停止、替换或回退而不破坏原始记录、传统基线或独立研究能力；在缺少预注册传统对照和独立复核时拟扩大用途或宣称研究增量。

触发后，R-G 只能对 Human-COS 相关研究模块、数据域、输出主张或用途作限域 RUN:LIMITED/RUN:HOLD/RUN:CLOSE 决定；符合机械条件时由 R-S 直接执行限域 RUN:HOLD。这不停止现实系统，也不自动停止与该工具独立的传统研究流程。触发后的工具输出不得支撑正面候选资格，但必须作为审计和负结果材料保留。回退必须返回最近一个可独立运行、可审计的传统研究基线及最近一个通过审查的工具版本，不得删除触发前后的输入、输出、异议、权限与错误记录。

重开至少需要根因、影响范围与责任链记录；新工具版本、哈希、数据与权限映射；R-V 复核与独立 R-X 挑战；受影响的 L0-R 与 T6-R 重新判断；以及 R-G 书面决定、R-S 按新版本执行。Human-COS 不得参与对自身是否安全、是否有效或是否应当重开的最终认定，也不得形成上述重开所依赖的 L0 或

T6 权利—授权状态。重复发生承重偏差，或结构上无法实现审计、替换、停止、责任定位与传统基线独立运行时，应对该影子模块执行 RUN:CLOSE；这不构成所研究现实机制的 FAIL，也不妨碍安全的传统研究继续。具体偏差阈值、比较指标与观察窗由未来影子协议实例化。

17.4 永久边界

- Human-COS 不能自动生成真实支付结构或把仿真输出变成均衡证明。
- 不能替现实主体作政策决定、承诺履约、创建核验信誉、执行制裁或承担后果。
- 不能承担 R-A、R-L、R-G、R-S；不能形成或改变 L0-R、L0-M、T5 或 T6 的权利—授权分量；也不能获得 M 平面现实停止权或回应权。
- 不能作为 R-V/R-X 或 M-V/M-X 的唯一或最终承载，不能凭工具分身、代理链或共同提供者满足独立复核与实际控制隔离。
- 不能把多个模型的共识、多数意见或公众关注度当作事实。
- 不能隐藏敏感信息治理、主要相关主体可见性和责任归属。
- 理论或 Runtime 版本变化不自动改变本母本；任何接口变化必须独立评估并明示升版。

上述边界适用于任何功能等价的模型、代理、自动化服务或外包平台，不因名称、接口、部署位置或提供者不同而放宽。工具输出只能作为有来源、可追溯、可挑战的输入；相应状态必须由能够承担责任、通过实际控制审计并具有相应权限的现实主体形成。

工具位置 | Human-COS 可以帮助现实主体看见反馈、责任与多主体耦合；它不能替主体形成真实利益、信任、承诺、执行能力和共同体。

十八、当前闭合项、开放项与锁定顺序

18.1 V1.0 已冻结的母本不变量

- 关系状态不是线性价值阶梯；研究成熟度与现实演化顺序分离。
- 多维利己只作可达性压力测试，必须预注册、可反驳并递归约束研究者与守门人。
- M 平面现实机制与 R 平面研究治理严格分开，研究状态无现实停止效力。
- M-E/V/X/D/S 只是现实功能、权限冲突和证明义务的分解，不是五机构蓝本；研究治理成熟不得推定现实制度成熟。
- 六重测试、L0-R/L0-M、过程台账、T5、T6-R/T6-M、四层状态、R0—R6 阶段与现实授权状态共同控制主张升级。
- 双轨按伤害维度分配，潜在不可逆、不可补偿和权利硬阈值不得被普通净值抵销。
- R-E/V/X/A/L/G/S 的功能、状态转换、不可兼任、实际控制和错误记录先于具体机构选择。
- 独立性按具体被挑战对象判断；同一工作包以合并 M/R 实际控制矩阵阻断双帽、共同上游与跨平面自审。
- WP-01 的 H2a/H2b、H3、H5 与严格三值合取不能被正面假设多数冲销。

- H3 机会链、确认越界的 FAIL 规则与竞争性解释，H5 有界覆盖与新路径重开门槛已经由母本冻结；零事件与无限想象路径都不能生成正面或否决捷径。
- R0—R6 禁止正面越级，但允许下游反证筛查；计算可复现与独立挑战从 R1 起横向存在。
- 规范形成包、实际控制矩阵、最低资源能力、长期 UNKNOWN 的 RUN:CLOSE 及 L0M:REJECT 后研究边界已经形成可执行后果。
- Human-COS 及功能等价工具处于 PRE-HUMAN-COS/future shadow-validation 位置，不是机制、经验/正当性认定中心或现实主体；不得借工具分身制造独立，影子失效只沿既有研究状态链停止、回退与关闭。
- RUN:REOPEN 只是一次性版本过渡记录，不是第五种可绕过 PROCEED/LIMITED/HOLD 的研究许可。
- 语义升版、非语义补丁与附件兼容记录已经分开；任何版本关系都不产生证据支持或现实授权。

这些是 MASTER/V1.0 的冻结条款。任何后续工作包、结果稿、传播稿或兼容性记录都不得以省略、改名、汇总、参数化、工具接口或“领域特殊性”使其静默失效；需要改变时必须先停止受影响派生、对母本实施明示语义升版并重新建立派生关系。

18.2 由 WP-01-A 实例化的开放项

- 首个具体领域、事件纳入/排除标准、历史样本与可比性规则。
- M 平面与 R 平面各功能的具体候选承载、资格、跨平面合并实际控制隔离、预接受、回避、替补、撤换和申诉时限；M 平面映射只作为待研究对象，不构成现实任命、机构设计或授权。
- H1—H5 的指标、样本量与观察窗；H3 机会可达、识别、能力、决定空间、延迟利用与可观测性代理、暴露向量与 δ_k ；H5 主体一路径矩阵、生成方法、可信性/实质性门槛、停止条件、残余路径聚合和系统安全阈值。
- H→T 依赖图、下游反证筛查的完成度，以及新路径、环境变化与原范围遗漏的领域化判别规则。
- 数据字典、字段缺失时的降级规则、证据接口、日志、访问控制、盲法、代码、模型、统计方法与条件地图实现。
- 规范形成包的领域化来源、受影响主体参与、缺席代理、代表局限、冲突处理、程序、时限与升版字段。
- 每项停止条件的 OBJECTIVE_TRIGGER/G_REVIEW 分类、HOLD 范围、响应时限和重启材料。
- 资源—能力声明的人员、预算、时限、替代安排与信息增量界值；L0M:REJECT 后的领域化允许/禁止研究清单。
- Human-COS 影子对照协议、增量指标、偏差阈值、观察窗与具体回退基线；不得在工具试运行前预写结果。
- $\Delta p_{i,r,t}^i$ 的算例、不可计算典型场景与数据合同；这些示例不得取得新的许可效力。
- WP-01 失效模式库，以及重复博弈、状态转移、因果推断和代理仿真各自的适用边界。

R2 进入 R3/R4 的操作性或量化门槛，以及 R6 现实主权/试点起步主体，属于后续阶段开放项；它们不由 WP-01-A 凭空关闭，也不能因尚未实例化而被当作已经通过。

母本先冻结功能、边界和不可兼任不变量；WP-01-A 候选稿内实例化具体承载、阈值与记录表，并只能提供未来试点授权的空白验证字段。真实 PILOT-SCOPE AUTHORIZED 只能在 R6 由现实主体签发。这些实例化事项必须在 WP-01-A 冻结前关闭，而不是在开始编写附件前凭空完成。

18.3 母本—附件关系

母本冻结不等于停止研究，而是固定后续研究必须遵守的定义、纪律和边界。独立附件只能实例化母本明确开放的领域、主体、变量、阈值、样本、承载角色和记录格式，不能修改六重测试、正当性双门、双轨计算、权力分离、三值合取、R0—R6 或 Human-COS 边界。

后续附件、结果与传播稿引用版本时不得裸写 V1.0。母本使用 MASTER/Vx.y，工作包使用 WP-01-A/Vx.y。首份 WP-01-A 必须绑定本 MASTER/V1.0 成品的外部交付哈希；每份附件永久绑定其初始父母母版本与哈希，母本升版不会改写既有附件。兼容性记录不得改写这一绑定，只判断附件能否在另一母本下继续采用。独立记录至少包括附件身份与哈希、初始父母本、目标母本、受影响条款、需要重跑的 H/T/L/权限门、独立复核、适用范围、结论和重开条件；当影响尚未查清时，附件只作为历史材料或方法输入，不得作为新母本下的现行合规证据、阶段升级依据或派生父件。凡修改附件内容、父版本、哈希、规范映射或结论，附件必须升版。

兼容结论沿用第 1.2 节的版本判定链：R-E 保全附件、新旧母本、差异与哈希；R-V 核验影响清单与需要重验的字段；至少一条独立 R-X 挑战语义、权限、规范、状态与主张影响。附件作者、原团队或同一实际控制者不得单独形成完整兼容结论；存在合理分歧时保持兼容性待核，受影响附件只能按历史材料限域使用。受影响的 H/T 由 R-A、L 与规范门由 R-L 重新形成，现实授权必须另行重验；R-G 只决定研究采用范围，R-S 追加兼容、限域、重验或升版记录。

18.4 下一次唯一接力项

下一次工作 | 编制《危机通报—有限核验—条件性回应机制 | 预注册研究设计 V0.1》（WP-01-A），把 MASTER/V1.0 明示开放的领域、角色承载、H1—H5、数据、阈值、规范形成包、停止条件与记录格式实例化。该步骤只形成研究协议候选，不启动现实试点、国家政策仿真或 Human-COS Runtime，也不同时派生论文、政策简稿、英文稿或其他附件。

WP-01-A 完成并独立复核前，不派生 Human-COS 接口规范、学术论文、政策简稿、英文稿或其他关联文稿；任何现实试点仍须等待 R6 的现实授权与独立验证。

附录 A | 机制单元卡：母本最小字段

字段组	必须记录
身份与版本	名称、母本版本、工作包版本、领域、主体、时间尺度、R 阶段、哈希与变更来源
关系状态	竞争／合作／联盟／阵营／共同体的领域化定义及转化方向
行为与利益	预注册利益维度、代理、竞争性解释、真实后果、非圣人条件与未来选项
结构改变	信息／激励／风险暴露／承诺核验／反馈纠错
因果与反事实	动作→中介→结果；现实基线、可行替代、竞争性解释和反证条件
通用关键假设	信息、能力、理性、贴现、执行、核验、国内结构、第三方与时间窗
脆弱性	先行动者暴露、协议偏离诱惑、误判、泄露、俘获、套利与平面渗漏
可观测性	变量、代理、时间戳、证据来源、误差、H3 机会链、H5 有界路径覆盖、残余路径与未知条件
L0 与外部性	L0-R、L0-M、主体、转嫁路径、负担性质、授权修复、规范形成包、规范冲突、过程台账与 T5
权限与认定	M-E/V/X/D/S 功能位、R-E/V/X/A/L/G/S、分平面及合并 M／R 实际控制矩阵、预接受、错误记录与四类输出
假设与证据	工作包承重 H、支持／失败／未知、最强反例、严格合取与独立复核
双轨与阈值	可逆可补偿维度、不可逆／不可补偿维度、损失下界、区间、阈值与不货币化理由
停止与重开	信号、触发类型、认定、执行、能力缺口、HOLD 范围、UNKNOWN 关闭理由、Human-COS 回退基线、修复证据、复核、升版与异议
结论状态	标准输出记录顺序、四层基础状态、T1/T2/T3/T4/T5/T6-R/T6-M、R 阶段、现实授权；方向／机制假设／证据发现／条件性支持／研究候选机制的准确用词

附录 B | 上游谱系、公开接口与历史原型

文稿或接口	当前身份	可提供	不得产生的效力
《当智能不再稀缺》V0.6	THEORY LINEAGE 理论谱系	总体方向、概念来源与文明边界	不证明具体机制成立
中文公众长文独立出版版 V1.1.1 第 9 章	DIRECT PUBLIC MECHANISM DISCIPLINE 当前唯一公开方向接口	结构先于人格、分配与俘获、现实机制不能由理论或工具代劳、工具可审计／替换／停止	不替代 V0.6，不构成经验验证；V1.1 只记录首次形成史，不具有独立现行接口效力
《当智能不再稀缺》V0.4 外部证据加固与方向性机制桥控制矩阵	EVIDENCE CONTROL RECORD 编制与证据控制记录	处理、排除与证据来源记录	非理论基线、非成立证据、非授权

文稿或接口	当前身份	可提供	不得产生的效力
《中美竞争与合作的可计算研究设计》v1.0	EARLY BILATERAL PROTOTYPE 早期双边原型	问题来源、风险—收益阈值与部分机制构件	非母本、非均衡证明、非部署依据
上述 v1.0 评估报告	REVIEW INPUT 审查与修订输入	暴露可达性、反事实、双轨、核验与第三方缺口	非独立经验验证
Human-COS	FUTURE TOOL INTERFACE 未来工具接口	记录、主体覆盖、因果推演、异议、复现与影子比较	不取得 R-A/L/G/S，不形成 L0/T5/T6 权利—授权分量，不作为 R-V/X 或 M-V/X 唯一—最终承载，不取得 M 平面现实权力

上述材料不是一条线性高低版本链，也不自动更新本母本。未来升版只能触发影响评估；若要改变本项目定义、阶段门、主张等级或权限边界，必须通过母本明示升版。

附录 C | 方法谱系与借用边界

本项目的治理架构与严格状态规则属于项目设计，不是已有国际通行标准或已经外部验证的最佳架构。以下来源只提供方法参照，不证明 WP-01 或本母本的机制主张成立：

- 英国财政部《The Green Book (2026)》提供敏感性、转换值、分配和不可货币化影响的现行政策评估参照；美国 OMB《Circular A-4》(2003-09-17 历史官方文本) 提供阈值／break-even、非量化影响与分配分析的历史方法参照。 Δp^i 是本项目对 break-even／switching-value 逻辑的项目化应用，不是新数学方法，也不继承上述规范的政策授权效力。
- Holland, 《Statistics and Causal Inference》(1986) 提供反事实／潜在结果因果推断的经典表述；选择偏差、共同冲击和样本可比性是本项目据此保留的历史比较识别要求，不声称该为该文给出的穷尽清单。
- Shapley, 《Stochastic Games》(1953) 提供随机博弈的经典原始表述；Axelrod, 《Effective Choice in the Prisoner’s Dilemma》(1980) 提供迭代囚徒困境计算机竞赛的原始报告。两者在模型或竞赛条件下的结果均不能直接证明现实主体可进入的均衡。
- Nosek 等, 《The Preregistration Revolution》(2018) 提供预测／事后解释及确认性／探索性研究的预注册区分，并说明预注册不会消灭低质量统计实践；“预注册不能补救错误变量、错误因果图或无效数据”是本项目据此设定的边界，不作为该文已经逐项验证的结论。
- 美国 FDA《Non-Inferiority Clinical Trials to Establish Effectiveness》最终指南 (2016-11) 提供预先设定非劣界值并将置信区间上界与该界值比较的临床方法参照。H3 只借用这一逻辑；临床界值、前提、证明标准和监管结论均不得直接移植。
- Altman 与 Bland, 《Absence of evidence is not evidence of absence》(1995) 只支持“未观察到显著差异不等于已经证明安全或无差异”的一般证据纪律，不单独提供 H3 的正面安全判据。
- 美国国家科学、工程与医学学院 (NASEM, 2019)《Reproducibility and Replicability in Science》提供“同一输入与计算步骤的计算可复现”和“各自取得新数据的复制研究”定义；

“独立再分析”是本项目在两者之间明示增加的运行分类，不声称为该报告定义的第三种统一类别。

· [NIST FIPS 180-4 \(2015-08\)](#) 规定包括 SHA-256 在内的安全哈希算法及消息摘要。在本项目中，SHA-256 摘要只用于相对已知基线检测文件字节变化；它不认证作者、来源、日期、内容正确性或现实权威。

历史通信、通知、核验或磋商案例若以后进入附件，只能作为“装置曾存在或运行”的存在性证据；不能单独证明总体战争概率下降、合作可达、可维持、正当或可部署。

附录 D | 一票否决终核

出现以下任一情形，当前主张、版本或阶段不得升级：

1. 关系状态被写成单向文明阶梯，或 R0—R6 被冒充为现实必然演化。
2. 多维利己通过事后增加维度吸收所有反例，失去可证伪性。
3. M/R 平面混用，使研究标签、HOLD 或工具输出拟触发现实行动。
4. 同一实际控制者在单一平面或跨 M/R 两平面贯通证据进入、核验、挑战、后验标签、正当性、决定、执行、阶段状态或重开；M 平面主体控制评价自身记录、决定或机制的 R 平面链条；或形成资格已经失效却仍保留原 L0/T5 标签的现行效力。
5. R-V 输出直接生成恶意、背叛、责任或现实回应判断。
6. H1/H2/H4 的正面证据把 H3/H5 或其他承重 UNKNOWN 冲成支持。
7. H5 代替 L0/T5 作规范判断，或 L0/T5 被计作 H5 的经验支持。
8. 双边或总体收益覆盖第三方不可补偿权利、潜在不可逆风险或未被计入的负担。
9. L0M:REJECT 被用来删除危险机制的安全、非介入式负结果研究；或反被当作无需新版本与 L0-R/L0-M 即可继续优化、介入、形成试点或政策材料的许可。
10. 潜在不可逆、不可补偿或权利硬阈值被普通净值跨轨抵销。
11. 调参、换窗口、删样本、换代理或改阈值不升版本。
12. R2 以禁止越级为由拒绝第三方反证筛查，或以做过筛查为由越级形成 R3/R4 正面结论。
13. 计算可复现、来源边界不足，或独立挑战未相对于本次被挑战对象及其实际控制者成立，却保留原条件性支持标签。
14. Human-COS 或功能等价工具被赋予 R-A/L/G/S、L0/T5/T6 权利—授权状态、现实核验信誉、政策决定、停止或执行权；作为 R-V/X 或 M-V/X 的唯一—最终承载；或以工具分身、代理链、共同提供者冒充独立复核。
15. 以 T6-R 研究流程可纠错替代 T6-M 现实机制的授权、退出、修复、申诉与重授权。
16. M-V/M-X 窄事实输出绕过 M-D 授权决定或直接命令 M-S 执行。
17. 把 M-E/M-V/M-X/M-D/M-S 功能位直接解释为现实机构方案，或以 R 平面成熟证明其已经可达、正当、可信或获现实授权。

18. H3 以“未发生利用”替代机会可达、识别、能力、决定空间、延迟窗口和可观测性的承重证据；已经依据预注册规则确认关键暴露维度超过 δ_k 却不形成 FAIL；或 H3 UNKNOWN 时仍形成 T2 条件性支持。
19. H5 把冻结范围内的有界覆盖表述为穷尽全部第三方路径；达到可信性、实质性与边界相关性门槛的新路径被忽略；或未达门槛的纯推测取得无限停止权。
20. 没有可审计规范形成包，或足以改变准入／正当性结论的规范冲突尚未关闭，却形成 LOM: ADMIT 或 T5 PASS。
21. 最低治理能力或实际控制矩阵未满足，却通过扩大范围、外包、自动化、形式分立或“以后补齐”继续确认性研究或形成正面候选资格。
22. 以 RUN:CLOSE 把承重 UNKNOWN 改写为安全、不存在、失败、默认反对或默认许可。
23. Human-COS 或功能等价工具触发承重偏差、不可审计、泄露、中心化、责任逃逸或不可停止条件后，仍扩大用途、删除记录、自行认定或自行重开。
24. 任何语义变化被标记为补丁，或者既有附件在母本变化后未经兼容性记录与必要重验即被作为新母本下的现行材料采用。
25. 未满足授权主体、权限链、范围、期限、停止回退、复审申诉、责任记录等最低语义即报告 PILOT-SCOPE AUTHORIZED，或以试点授权推定机制有效、正当、可扩展、T6-M 通过或运行授权成立。
26. 没有可核验的常态运行授权、持续监督、定期复审、暂停撤销、责任修复和承重变化重授权即报告 OPERATIONAL AUTHORIZED，或以运行授权免除 H／T／L 复审。
27. 修改者、原团队或同一实际控制者未经独立 R-X 复核即单独宣告语义补丁资格或附件完整兼容，或在存在合理分歧时仍按正面结论采用。
28. 承重阻断、FAIL 或 UNKNOWN 未在正面 H／T／L 结论前集中展示，或在 LOR-REDESIGN／LIMIT／FAIL、LOM:REDESIGN／LIMITED／UNKNOWN／REJECT 或 T5 承重阻断下仍对同一范围标记 RUN:PROCEED。

附录 E | V0.1—V1.0 版本迁移与哈希

版本	状态	主要作用	当前效力
V0.1 立项基线	历史冻结件	建立六重测试、双轨、R0—R6、WP-01、证据与风险边界	不再作为现行条款库
V0.2 承重纪律修订版	历史冻结件	增加利己递归、正当性双门、权限分离、H3/H5 与严格未知	不再单独派生附件
MASTER/V0.3 完整整合候选稿	历史候选	单稿重述全部现行规则并关闭结构循环	不再作为现行条款库或附件父版本
MASTER/V0.4 锁定候选稿	历史候选	关闭制度蓝本歧义、机会与路径覆盖、规范与资源、长期未知、工具失效和版本接口	不再作为现行条款库或附件父版本

版本	状态	主要作用	当前效力
MASTER/V0.5 终审修订候选稿	历史候选／直接语义前身	封闭 REOPEN、H3 阈值、工具权限、按对象独立性、跨平面控制及附录／来源差异	不再作为现行条款库或附件父版本
MASTER/V1.0 FROZEN	当前正式锁定母本	完成连续终审并固定首个可派生研究母本身份	WP-01-A 及后续附件的唯一现行派生根

前序交付文件 SHA-256（哈希对象为下列正式交付 DOCX 的完整文件字节；渲染 PDF、提取文本与后续副本不共用该哈希）：

- 竞争合作联盟阵营共同体_可计算机制研究_V0.1_立项基线.docx：
7867da79cf0e0cb805b48f305d695814cceeaea327ab86d917e41890f1d34f3d
- 竞争合作联盟阵营共同体_可计算机制研究_V0.2_承重纪律修订版.docx：
e6ac854ff1989271700b518f1a43c0d561c19398725222f4964f16502db7ac17
- 竞争合作联盟阵营共同体_可计算机制研究_MASTER_V0.3_完整整合候选稿.docx：
a9d0abdf8e7b34f7f81389c0664de7bc45e8fcf24ce12c327d2a3e2653074926
- 竞争合作联盟阵营共同体_可计算机制研究_MASTER_V0.4_锁定候选稿.docx：
3b2750badbc0753ca21292d58756181f58e23e4773b4117398efc8291df77f9c
- 竞争合作联盟阵营共同体_可计算机制研究_MASTER_V0.5_终审修订候选稿.docx：
4f6d119969ce961024921c7e31392668014a1725ec0c0ca626e3dd8061871c04

V0.3 相对前序的语义变更台账：

编号	变更	责任含义
V03-01	增加五种关系状态的操作性定义	阻断线性文明阶梯解释
V03-02	拆分 M 平面现实机制与 R 平面研究治理	阻断研究标签直达现实行动
V03-03	将多维利己明确为可反驳压力测试	阻断事后吸收所有反例
V03-04	冻结四层状态、研究阶段与现实授权标记	阻断不同状态互相冒充
V03-05	将 L0 拆为 L0-R 与 L0-M	分离研究行为安全与机制正当性
V03-06	将旧称 P^* 校准为 $\Delta p_{i,r,t}^{\dot{c}}$	明确主体、风险、时间与不可计算边界
V03-07	增加下游反证筛查规则	禁止以阶段门遮蔽第三方反例
V03-08	将 H2 拆为识别效能 H2a 与激励效应 H2b	禁止由分类改善直接推出行为改变
V03-09	分离 H5 经验判断与 L0-M／T5 规范判断	阻断事实与正当性互相替代
V03-10	区分计算可复现、独立再分析与复制研究	统一证据术语与阶段责任
V03-11	将 T6 拆为 T6-R 与 T6-M，并增加 M-D／M-S	阻断研究纠错替代现实授权纠错

V0.4 相对 V0.3 的语义变更台账：

编号	变更	责任含义
V04-01	明确研究宪章身份与 M 功能位非制度蓝本	阻断研究治理或功能分解被平移为现实机构设计
V04-02	冻结失败优先的标准输出顺序与试点／运行授权最低语义	阻断状态省略、总灯号和空壳授权

编号	变更	责任含义
V04-03	明确 L0M:REJECT 后负结果研究边界	同时阻断删除负结果与借负结果继续优化危险设计
V04-04	增加规范形成包与承重规范冲突规则	阻断 R-L 单方选择规范或多数压平硬边界
V04-05	增加实际控制矩阵与最低资源能力门	阻断名义分工、外包或自动化冒充真实治理能力
V04-06	完整化 H3 机会链、竞争性解释与 H→T 传播	阻断零事件冒充安全及 H3 UNKNOWN 下的 T2 升级
V04-07	将 H5 覆盖定义为有界程序并增加新路径三重门	同时阻断无限路径否决与预注册后任意忽略反证
V04-08	收紧客观触发器并冻结 PROCEED/LIMITED/HOLD/CLOSE 选择规则	阻断把解释藏入字段或自由选择运行状态
V04-09	建立长期 UNKNOWN 的 RUN:CLOSE 与重开纪律	允许停止无限投入但不消灭未知
V04-10	建立 Human-COS 影子停止、回退与关闭链	阻断工具在失效后扩大用途或自行重开
V04-11	分开语义升版、非语义补丁与附件兼容记录，并纳入独立复核链	阻断修改者自行宣告补丁或附件静默继承
V04-12	增加附录 F 非规范速查层	降低外部接手门槛而不制造第二套规则源

V0.5 相对 V0.4 的语义与自足性变更台账：

编号	变更	责任含义
V05-01	将 RUN:REOPEN 限定为一次性版本过渡记录并设置默认 HOLD	阻断以“重开”绕过 PROCEED/LIMITED/HOLD 的适用条件
V05-02	补齐 H3 已确认超过 δ_k 的 FAIL 状态	阻断把已证实阈值突破降格为 UNKNOWN
V05-03	禁止 Human-COS 与功能等价工具承担 R-L、形成正当性状态或借工具分身制造独立	阻断工具经 L0/T5 间接取得阶段否决或现实信誉
V05-04	将 R-X/M-X 与独立再分析/复制的独立性绑定具体被挑战对象	阻断对另一对象无冲突被冒充为本项独立复核
V05-05	增加 M/R 合并实际控制矩阵与跨平面回避后果	阻断双帽、共同上游和形式防火墙掩盖自我评价
V05-06	校准附录 C 的机构归属、术语来源与项目自设边界	阻断把项目纪律误写为外部来源已经证明的结论
V05-07	校齐附录 A/B/E/F 的主张等级、授权标签、角色摘要与版本链	阻断速查或迁移表形成第二套口径

V1.0 相对 V0.5 的锁定变更台账：

编号	变更	责任含义
V10-01	完成逐条连续审读、逻辑反证、权力—责任、规范—资源、来源—主张、正文—附录及成品回归审计，P0/P1 清零	冻结资格来自可复核审计，不来自版本名称或作者宣告
V10-02	将版本身份改为 MASTER/V1.0 FROZEN，并固定其为首个可派生母本	后续附件必须绑定本版本及其外部交付哈希，不得从历史候选稿派生
V10-03	除锁定身份、谱系、派生根与接力接口外，不改变 V0.5 的机制语义	防止在“正式锁定”名义下夹带未审计规则

哈希只证明当前文件字节与交付记录一致，不证明作者、日期、理论正确性或现实权威。MASTER/V1.0 的成品哈希不得回写被哈希文件；应记录在独立版本清单、交付记录或同名 .sha256 文件中。被哈希母本只记录前序文件哈希与外部清单规则，以避免自指变化。

附录 F | 框架总览与标准输出速查

速查层效力 | 本附录只帮助外部团队定位正文，不创设、放宽、替代或汇总正文规则。任何简写都必须回到对应正文条款解释；若出现不一致，以正文的完整限定为准。本附录不得被单独抽出作为制度蓝本、授权文件或自动判定器。

F.1 一分钟定位：先分平面，再分状态

先问什么	正确位置	最常见误读
现实中谁有权决定与执行	M 平面；必须另证现实授权、能力与责任	用研究标签或工具输出代替现实决定
研究中谁保全、核验、挑战、认定与守门	R 平面；受实际控制与不可兼任约束	把成熟研究治理直接平移为现实机构
证据现在说明什么	窄事实、H、T、L0/T5 分列报告	用一个总分或多数结果冲销 UNKNOWN
研究现在能做什么	当前运行许可：PROCEED/LIMITED/HOLD/CLOSE；REOPEN 仅为一次性过渡记录，必须同步目标状态	把研究运行状态写成现实停止命令
研究是否核验到满足最低语义的现实授权	NO M-AUTHORIZATION/PILOT-SCOPE AUTHORIZED/OPERATIONAL AUTHORIZED	把研究候选或试点授权写成运行资格
文件在什么版本下有效	母本、附件、父版本、哈希与兼容记录	静默继承、回写父版本或把语义变化称为补丁

F.2 两个平面的功能位：名称相似，权力不同

功能位	只形成什么	绝对不能因此形成什么
R-E	研究材料、来源、版本、哈希与访问记录	H/T/L 判断、阶段决定或现实行动
R-V	预注册窄命题的四类证据输出	动机、责任、背叛或现实回应
R-X	独立挑战、遗漏与竞争性解释	无限噪声否决或自行决定停止/重开
R-A	可挑战、可撤销的 H、T1—T4，以及 T6-R/T6-M 的技术可操作性参考标签	T5、现实事实中心、阶段控制或授权
R-L	L0-R、L0-M、过程台账、T5，以及 T6-R/T6-M 各自的权利—授权部分	经验替代、抽象价值任意否决或阶段执行
R-G	根据既有状态决定研究运行范围	生产证据、修改阈值或改写状态
R-S	机械执行冻结、PROCEED、LIMITED、HOLD、REOPEN、CLOSE 与记录	解释条件、删负结果或单独重启
M-E	现实运行证据保全	事实核验、现实决定或行动执行
M-V	操作性窄事实输出	直接命令惩罚、回应或停止

功能位	只形成什么	绝对不能因此形成什么
M-X	跨方复核与冲突暴露	取代授权决定或凭异议自动执行
M-D	经现实授权作限域决定并承担后果	伪造事实、越权执行或转授给 R/工具
M-S	在授权范围内执行并保留责任记录	改写决定、扩大范围或自我授权

这些是功能、冲突检查位与证明义务，不等于十二个机构。具体组织可以采用不同名称、数量和拓扑，但不得消解正文冻结的不兼任、复审、申诉、退出与责任边界。

Human-COS 与任何功能等价工具只能提供有来源、可追溯、可挑战的输入；不得承担 R-A/R-L/R-G/R-S、形成 L0/T5/T6 权利—授权状态、作为 R-V/R-X 或 M-V/M-X 的唯一/最终承载，或以工具分身满足独立性。

F.3 标准输出记录：每个字段都必须出现

- 1. 主张 ID、主体、领域、时间范围与准确主张等级；
- 2. 文件版本、父版本、哈希与差异来源；
- 3. R0—R6 当前研究阶段及其阶段上限；
- 4. M 授权状态与授权文件核验结果；
- 5. 每个窄事实的 SUPPORTED/CONTRADICTED/INSUFFICIENT/NOT VERIFIABLE；
- 6. 承重阻断、FAIL 与 UNKNOWN 总表：H 的 FAIL/UNKNOWN；T1—T4、T6-R、T6-M 的 FAIL/UNKNOWN；L0R-REDESIGN/LIMIT/FAIL；L0M:REDESIGN/LIMITED/UNKNOWN/REJECT；T5 的 FAIL/UNKNOWN。该表先于任何正面 H/T/L 结论显示；较早呈现的窄事实或授权核验不得冒充机制正面结论；
- 7. 每个承重 H 的 FAIL/UNKNOWN/CONDITIONAL SUPPORT；
- 8. T1—T4、T6-R 与 T6-M 的三值状态；
- 9. L0-R、L0-M 与 T5 的当前状态及规范包版本；若形成资格失效，另列原标签、失效原因、聚合处理与待独立 R-L 重形成状态；
- 10. 当前 RUN:PROCEED/LIMITED/HOLD/CLOSE 及限域对象；若发生重开，另列 REOPEN 过渡记录、新版本与同步目标状态，REOPEN 不得单列为当前许可；
- 11. 最强反证、残余异议、残余路径与竞争性解释；
- 12. 状态形成主体、证据窗口、挑战与独立复核；
- 13. 修复材料、重开触发器、下一复核点与不得外推范围。

不适用必须写 N/A+理由；应检而未检只能写 UNKNOWN。任何字段缺席都不得被解释为已经通过。

F.4 状态家族：各自只回答一个问题

状态家族	回答的问题	不得冒充
窄事实四类输出	当前证据与预注册窄命题是什么关系	主观动机、责任、H 或现实回应

状态家族	回答的问题	不得冒充
H 三值状态	工作包承重假设失败、未知还是条件性支持	六重测试、正当性或现实授权
T1—T4／T6-R／T6-M	相对更优、可达、维持、可验与纠错是否获支持	T5 或总体部署资格
L0-R	研究动作本身是否安全、可实施或须限域	机制有效或现实停止权
L0-M／T5	候选机制准入与阶段性正当性	经验因果成立或永久正当
RUN	当前研究运行、主张与版本派生能走到哪里	现实系统状态或认识论结论
R0—R6	研究成熟度与允许证据活动	文明演化顺序或现实机制成熟度
M 授权状态	研究是否核验到现实主体在特定范围内满足最低语义的授权	机制有效、T6-M 通过或永久许可
版本兼容记录	附件能否在另一母本下继续采用	新的证据、研究运行或授权状态

F.5 实际控制矩阵：至少检查八类杠杆

控制杠杆	必须记录
经费	谁拨付、中止、附条件、审计与替代
任命与撤换	谁任命、考核、续约、撤换与影响职业后果
数据	谁产生、选择、进入、访问、删除与阻断
模型与知识产权	谁控制代码、参数、计算环境、许可与复算条件
权限与凭证	谁签发、撤销、保管与查看日志
人员与技术上游	共同母机构、外包、云服务、平台与关键依赖
判断与发布	谁能改变假设、阈值、规范包、成功定义和发布范围
停止与重开	谁能触发、决定、执行、申诉、复审、修复与恢复

每一项都要追溯最终控制者，记录能否发起、阻止、改变、延迟或撤销，以及独立替代渠道。机构名称不同、签字增多或外包给同一依赖网络，都不证明已经分权。

同一工作包、事件和范围必须合并审计 M／R 两平面；独立性必须相对于本次被挑战对象及其实际控制者成立，不能用另一项无冲突关系替代。

F.6 停止—复核—重开的最短正确链

1. 任何主体报告信号；信号本身不改变状态。
2. R-E 保全材料；R-V 形成窄事实；R-X 独立挑战。
3. 唯一机械结果才能成为 OBJECTIVE_TRIGGER；解释、归因、价值或比例权衡进入 G_REVIEW。

4. R-S 只能对客观硬触发机械执行限域 HOLD；R-G 回避冲突后书面决定四类当前运行状态，或决定一次性 REOPEN 过渡并同步目标状态。
5. PROCEED 只能用于 L0R-PASS、L0M:ADMIT 或经记录不适用、T5 无承重阻断且无其他承重 FAIL/UNKNOWN 的同一范围。HOLD 期间只开展经 L0-R 允许的最低证据保全、反证、修复与关闭未知活动；L0R-FAIL 时只保留不延续风险的最低证据与责任记录。LIMITED 期间可另开展经 L0-R 允许的范围界定、历史性、非介入式、负结果、独立再分析/复制与新版本重构准备，不得升级原版本正面主张。
6. 修复材料重新经过 R-E、R-V、R-X，并由 R-A/R-L 重形成受影响状态。
7. R-G 决定是否重开；R-S 以新版本登记一次性 REOPEN 过渡，并由 R-G 同步赋予 PROCEED/LIMITED/HOLD 之一。未同步赋值或受影响门未重过时默认 HOLD；REOPEN 不许可研究动作，也不恢复旧正面标签。
8. 证据线穷尽时可以 RUN:CLOSE；FAIL/UNKNOWN 与异议永久保留。

任何时候都不得发生以下短路：R 状态直达 M 行动；M 窄事实直达惩罚；H 多数冲销承重 UNKNOWN；T6-R 替代 T6-M；L0/T5 替代 H5 经验判断；Human-COS 或功能等价工具形成 R-L 状态、生成/执行/重开自身权力或以分身制造独立；试点授权自动变成运行授权。

个性化续读指引

P05-COOPMECH-001 | MASTER V1.0 FROZEN | 研究正式资产

本页为文库导航附页，不改变冻结研究母本正文。

向上游文明判断 | 《唯一之路》公众思想版

为什么读：继续追问旧竞争、安全与自保逻辑为什么可能共同制造更大的不安全；本研究只处理机制何时可达、失败或未知。

<https://www.civitas.top/documents/p03-only-002/>

向责任与现实判断 | Human-COS：让后来才看见的，早一点被看见

为什么读：把机制主张重新放回后果、责任、关键异议与现实有权有责主体；Human-COS不因此取得政策或停止权。

<https://www.civitas.top/documents/p05-humancos-004/>

向不可逆边界 | 共同窗口协议：开放未来与三约束

为什么读：当机制研究进入高后果领域时，继续检查回滚、退出、非操控与未来仍能重新选择的最低条件。

<https://www.civitas.top/documents/p05-window-005/>

研究边界

本稿是研究母本，不是理论Canonical、政策授权、现实制度蓝本或自动判定器。MASTER/V1.0处于PRE-HUMAN-COS相位：研究设计可以推进，但现实机制、工具效力、试点资格与现实授权均不得预写。失败、UNKNOWN、反例与一票否决结果必须永久保留。

后续研究：P05-FAIRNEG-001 多主体公平谈判与利益平衡仍处研究基线，尚未建立公开稳定链接。

完整公开文库

<https://www.civitas.top/library/>

文明跃迁官网

<https://www.civitas.top>