

人类命运共同体的实现、演变与跃迁

V3.9.3 三支柱结构对称与运行闭环锁定版

作者：子君赋

出品：文明跃迁研究组

事实与版本基准：2026年7月21日

目录

决策摘要 | 当前建议与决策边界

总论：从国家命运推导人类命运共同体的必要性

相位纪律：共同体不能以稳定期图景胁迫启动期

共同体必须为不同主体提供无损进入理由

第一部分：现实锚点与战略判断

第二部分·实现（上）：从零和现实建立国家层共同存续底座

第三部分·实现（下）：为文明跃迁建立个体制度接口

贯穿国家层、社会层与个体层的五类运行机制

第四部分·演变：从护栏到结构，从恐惧到预期

从倡议到采用：政治启动与制度承载链

第五部分·跃迁：文明逻辑转换、长期接口与终极检验

结语

精选附录一：事实基准与现实接口

精选附录二：核心制度边界摘要

精选附录三：相位、停止与退出摘要

精选附录四：研究依据与版本变更

人类命运共同体的实现、演变与跃迁

——在零和现实中搭建非零和底座的结构性推演

V3.9.3 三支柱结构对称与运行闭环锁定版 | 作者：子君赋 | 文明跃迁研究组

版本说明：

本文定位为面向政策研究、国际治理与制度设计读者的文明级战略推演。它不是任何政府或国际组织的正式政策文件，也不是可直接执行的制度规范；其任务是使关键决策者仅阅读主文，即能看清现实条件、国家利益、制度结构、权利边界、相位纪律、风险控制、政治承载与当前应当作出的决策。全文区分【已核实事实】、【作者研判】和【制度构想·情景推演】。专业协议、参数模型、数据字段和执行手册继续作为研究储备，不构成理解本文核心判断的前提。

公共思想与体系展开由文明跃迁五卷书——《被需要的文明》《制度前语言》《文明跃迁白皮书》《文明永续》《意义共生》——承担。V3.9.3 在 V3.9.2 已经补足向善智能体系制度骨架、提升五类运行机制层级的基础上，只进行最后一次结构终整：将治理共同体的多中心权力结构集中呈现，使资源共同体、治理共同体、向善智能体系在结构上真正对称；补明五类运行机制的动态闭环；压缩现有治理工具局限的重复解释。理论主轴、相位纪律、贡献值防火墙、三权分离、失败熔断和当前决策请求均不重开。完成本版后，主文冻结：动态事实进入事实台账，新增对象进入对象研究，新增参数进入实验与专业储备。文明的跃迁不等待完美蓝图，但每一步都必须保留证据、边界、退出与纠错。

决策摘要 | 当前建议与决策边界

本页供关键决策者在进入全文前快速判断。本文当前处于建议与前瞻研究阶段，不请求立即批准任何完整制度系统。

决策问题	摘要
本文的核心判断	多国长期共存、国家命运内外相连，使共同安全、共同发展与共同持续不再只是国际道义，而成为各国维护自身安全、发展能力和长期未来必须面对的国家利益问题。
本文提出的结构	以资源共同体为现实基础，以治理共同体为制度手段，以向善智能体系为新能力支撑，使国家、社会和个人获得另一条增强安全、发展与持续能力的道路。
当前请求作出的决策	第一层只建议将该命题纳入国家级前瞻研究；其后依次是跨部门论证、候选城市基线调查与方案预研。只有法律、财政、技术、权利和社会可行性通过审查后，才另行决定是否进入城市级综合实验授权。
第一阶段成本与退出	第一阶段主要承担研究、论证、基线调查和方案比较成本；不改变现行产权规则，不强制个人、企业或地区参与，不部署不可逆制度。论证不成立即可停止，过程与失败证据完整保留。
本文明确不请求	不请求一次性改造现行社会，不请求立即建立全球治理中心，不请求以贡献记录决定基本权利，不请求让人工智能替代人类终审，也不把任何国家或群体的先行牺牲作为启动条件。

当前首先需要回答的不是“是否接受整套文明跃迁方案”，而是：这一命题是否值得进入国家级正式研究，并接受事实、法律、利益、权利与失败情景的系统检验。

总论：从国家命运推导人类命运共同体的必要性

【作者研判】

国家命运并不只存在于国境之内

世界由多个国家长期共同存在。没有任何国家能够消灭所有其他国家，也没有任何国家能够永远脱离其他国家，独自获得维持现代社会所需要的全部资源、技术、市场、知识与安全条件。

一个国家可以决定自己的部分行动，却不能独占行动的后果。它的安全政策会改变他国的安全判断；他国的反应又会重新塑造本国环境。它的产业、能源、技术和金融政策会穿过供应链进入其他国家；其他国家的战争、崩塌、封锁、疫情和技术失控，也会穿过边界进入本国。

因此，国家命运既取决于其内，也取决于其外；既取决于国家如何组织自身，也取决于各国彼此如何作用，以及共同构成的世界是否仍能容纳本国继续存在、发展和纠错。

安全、发展、持续共同构成国家命运

安全回答国家能不能继续存在。发展回答国家能不能不断获得新的生存能力。持续回答今天的安全与发展能不能跨越时间，而不在未来自我瓦解。

安全决定国家能否存在，发展决定国家是否拥有未来，持续决定这个未来是否真实。三者不能彼此替代：没有安全，发展成果可能被一次危机摧毁；没有发展，安全会逐渐失去物质与能力基础；没有持续，安全和发展都可能只是以资源透支、社会撕裂、外部敌意与代际损失换来的阶段性繁荣。

为什么必须是共同安全、共同发展、共同持续

在核武器、人工智能、生物技术、全球供应链和跨境基础设施深度连接的时代，一个国家不能通过让所有其他国家更加不安全，获得持久安全；不能在合理成本下独自拥有全部发展资源；也不能独自阻止生态、技术、金融和文明级风险跨境传导。

共同不是道德修饰，而是多国长期共存与跨境因果推出的现实结论。共同安全，是各国共同阻止竞争穿透共同存续底座；共同发展，是各国都保有通过自身努力进入未来的现实通道；共同持续，是今天的安全与发展不以耗尽下一代、摧毁共同系统和封死文明未来为代价。

资源共同体是现实基础

安全、发展与持续最终都依赖资源。这里的资源不仅包括粮食、能源、矿产、土地和水，也包括算力、数据、知识、技术、人才、组织能力、资本、市场、物流、公共基础设施与时间窗口。

各国事实上早已处于资源共同体之中：拥有矿产者依赖加工和市场，拥有模型者依赖芯片、能源与数据，拥有资本者依赖真实生产和社会稳定。问题不在于资源共同体是否存在，而在于现有相互依赖仍然不透明、不对称、可勒索、可武器化，闲置与短缺并存，危机时缺少共同调度，使用后缺少公平清算。

资源共同体不是资源共同所有，而是在所有权清楚、主权和合法权利受到保护的前提下，使部分资源的使用权、保障权、应急调用权与清算责任能够通过自愿、分层、可审计、可退出的接口协同运行。

治理共同体是制度手段

资源不会自动合理流动。谁可以进入、谁可以调用、危机时谁优先、如何防止强者支配、如何保护弱者、如何处理违约、争议和退出，都需要治理结构。

治理共同体不是全球中央政府，也不是国家主权替代者。它是对跨主体、跨区域、跨国家共同问题建立的有限、分层、多中心、可审计、可纠错、可退出的共同治理结构。它管理的是权利关系、授权接口、风险边界、使用程序和清算责任，而不是取得所有资源的所有权，也不是替国家治理全部国内事务。

向善智能体系是新能力支撑

传统行政协调、市场配置和跨境迁移无法在全球规模上持续发现、理解并连接海量分散的资源、能力、机会与真实需求。人工智能带来了此前不存在的新感知、新连接、新匹配、新推演、新协同和能力放大。

智能向善不是要求人工智能抽象地拥有善意，而是通过目标、权限、审计、责任、退出和人类终审，使智能优先用于减少共同伤害、释放闲置资源、扩大国家与个人能力，而不是只用于替代人、控制人和压倒他国。它为各国提供另一条增强安全、发展与持续能力的道路：不是替一个国家长期解决问题，而是增强其自主发现问题、组织资源、培养人才、应对风险和形成自身答案的能力。

人员与个体资源池的全域可达、可用与自主流动

共同体不能只连接国家和机构。每个人都拥有由时间、能力、知识、经验、作品、数据、关系、合法财产与人工智能能力外延构成的个体资源池。

人员与个体资源池的全域可达、可用与自主流动，是共同治理得以成立的重要基础，是共同体从国家与组织间协作走向每一个人真实参与的演变方向，也是文明从地域锁定、组织依附和被动配置走向权利随人、资源可达、能力释放与生命自主的跃迁方向。

人不是资源。个体资源池属于人，并服务于人的自由、发展与意义。个人有权决定是否进入、开放什么、授权多久、如何使用、是否撤回、何时退出与是否回流；不流动、不共享、不贡献，不影响基本生存与基本权利。

本文总定义与总路线

人类命运共同体，是在多国长期共存、国家命运内外相连、资源与风险深度跨境传导的条件下，各国为维护自身及共同世界的安全、发展与持续，通过资源共同体、治理共同体与向善智能体系建立的制度共同体。

它以资源共同体为基础，以治理共同体为手段，以向善智能体系为能力支撑，以人员与个体资源池的全域可达、可用与自主流动为共同治理的重要基础与演变方向，以共同安全、共同发展、共同持续为目标。

它最终服务的不是资源、治理和智能本身，而是让每一个国家、每一个社会与每一个人继续拥有可以期待的未来。

相位纪律：共同体不能以稳定期图景胁迫启动期

本文所有制度构想必须按启动期、初期、过渡期、稳定期理解。相位由上一阶段的真实结果、权利记录和独立核验触发，不由日历自动推进。稳定期的全域结构不能被写成启动期的现实义务，启动期的自愿邀请也不能被误读为共同体的最终形态。

相位	主要任务	规则性质	最低权利边界
启动期	建立理解、动力与最低合作接口	自愿、邀请性、只增加可感收益	任何主体的现状底线不得因方案出现而劣化
初期	在有限节点验证资源、治理、智能和权利闭环	可观察、可退出、可修正	退出不受减损，失败成本由方案承担
过渡期	形成多节点互认与可选择的制度引力	描述涌现，不设计胁迫	未加入者保有旧路径权利，加入通道持续开放
稳定期	形成经过验证的有限共同治理与宪制托底	一人一票守航道，专业经验管领域，规则可回滚	生存权先于贡献，刹车始终留在人和国家手中

层级说明：本表所述文明表决权 and “一人一票守航道”，主要适用于获得合法授权的城市级综合实验区及其后续成熟共同体，不表示主张在当前国家间治理层面直接实行全球一人一票。国家间共同治理遵循主权授权、分层参与和有限议题协同原则。

三条铁律贯穿全部相位：第一，相位标注强制——任何机制必须说明何时、在何种尺度成立；第二，全员无损——任何群体的基本权利和现实底线不得被当作过渡成本；第三，对立死结禁令——凡必须先击败、剥夺或压服某一群体才能启动的方案，说明方案仍在复制旧逻辑，必须退回重设计。

共同体必须为不同主体提供无损进入理由

共同体不能只说明整体最终会更好，还必须回答每一个现实承担者为什么愿意从今天开始进入。进入理由必须来自该主体真实的安全、发展、尊严、责任和未来利益，而不是要求其先牺牲，再等待遥远回报。

主体	真实动力	可感进入收益	不可突破的无损边界
国家与治理者	安全、发展、可控改革、历史责任	新的风险识别和资源组织能力；保有最终责任与停止权	主权核心、国内法实施、退出与安全例外不被架空
存量资本与企业	合法资产安全、长期确定性、新市场	闲置资源形成使用收益；获得新周期制度信用和发展空间	合法存量不被追溯清算，不以参与为名强制转移所有权
专业贡献者	经验正义、专业判断与真实署名	专业能力进入领域决策，贡献可追溯、可承接	专业判断不被货币、流量或统一评分稀释
普通劳动者与被替代者	生存安全、被需要感与新价值通道	照护、维护、社区在场和能力转型被看见	基础生存不依赖贡献，退出和暂不参与不受惩罚
青年与未入场者	新的成长坐标和创造空间	低门槛参与、学习成果可携带、AI 能力普惠	生存与教育权不以贡献记录为条件
弱资源国家与地区	自主发展能力、避免永久边缘化	本地语言、教育、医疗、产业和治理能力增长	不以支持为名抽取数据、控制模型或制造永久依赖
AI 开发与安全社群	创造自由、责任边界、避免军备裹挟	在透明、可审计边界内获得信任与长期创新空间	责任与可观测性相匹配，不承担无限责任

任何一类主体的损失若成为另一类主体的进入激励，即触发对立死结。正确路径不是要求谁先输，而是把旧逻辑下的“让步”重新设计为新逻辑中的“换入”。

第一部分：现实锚点与战略判断

【导读】总论已经从多国长期共存与国家命运出发，推导共同安全、共同发展、共同持续的必要性。本部分进一步把已经发生的事实、从事实推出的判断和现有工具的缺口分层摆清，说明这一逻辑并非脱离现实的抽象构想，而已经出现可以继续搭建的制度接口。标题三阶与本文结构的对应：实现=第二、三部分，使资源共同体、治理共同体、向善智能体系和个体制度接口第一次进入真实运行；演变=第四部分，使单一节点逐步连接为多地区、多国家、多领域的共同体网络；跃迁=第五部分，使国家安全、发展方式、资源价值与人的价值来源发生结构转换。

一、2026年7月16日至17日发生了什么

【已核实事实】

2026年7月16日，成立世界人工智能合作组织协定签署仪式在上海举行。29个国家代表签署协定；协定将世界人工智能合作组织界定为独立的政府间国际组织，总部设在中国上海。协定签署、法律生效、组织启动与实际运行属于不同制度阶段，本文不将其合并为同一事实。

7月17日，2026世界人工智能大会暨人工智能全球治理高级别会议开幕。习近平主席发表题为《携手构建公正合理的全球人工智能治理体系》的主旨讲话，提出坚持开放共赢、强化风险意识、鼓励包容并蓄、倡导和衷共济四点意见。同期，《人工智能合作发展行动计划》《国际人工智能伦理治理行动计划》发布；中国国家互联网信息办公室会同有关方面在大会主论坛提出《智能体互信互联互操作全球合作倡议》，围绕十个方面提出行动方向（全文由新华网等于7月17日公开发布）。

事实来源与现实接口见文末精选附录一。

二、世界人工智能合作组织意味着什么

【作者研判】

29个国家在一份文件上签字。这份协定可能成为AGI时代较早出现的、以共同发展与共同治理为目标的非零和制度接口之一。它的意义不在于竞争已经结束，而在于国家开始尝试为竞争建立不能穿透的共同安全与共同发展底线。

从"一份协定"到多份行动文件，谈判桌的功能正在从"建立最低共识"向"落地具体行动"演进。这是本文全部推演的现实锚点。

人类整体条款：

本文所述共同体以人类整体为对象——包括当前的竞争对手、尚未签署任何协定的国家、技术强国与最不发达国家。29国协定是接口的起点，不是阵营的边界；第七节的脆弱分析表明，中国、美国、欧洲与全球供应链彼此深度耦合——无论是否签署，技术强国在结构上已处于同一个风险共同体之内。因此，本文提出的共同体机制按普遍准入原则设计：加入不以承认任何一国主导地位为条件；规则应由参与主体共同制定；退出权应受明确程序与制度保障；参与是分层的——不同主体可按议题、按轨道选择性参与，不要求接受整套理论作为前提；不加入不受惩罚，安全层合作（重大事故通报、人道通信、灾难救援、共同存续核查）不以成员资格为前提、对非成员保持开放——但开放不等于无条件授予敏感数据访问权、核查权限、资源调用权或治理席位，具体权限仍须以风险必要性、协议授权、安全审查、责任承担和对等义务为前提；不加入不受惩罚，也不意味着未承担共同成本与协议义务的主体自动取得参与者共同建设形成的全部增量权益——资源调用、敏感数据访问、治理席位及其他增量权益，依据实际责任、资源投入、风险承担和协议义务配置。三层关系合为一句：基本安全接口不封闭，不加入不受报复，增量权益不脱离责任。反向的边界同样成立：共同体内部规则不得以惩罚未加入主体、制造外部技术封锁或固化地缘阵营为目的——成员可以根据共同投入配置内部增量权益，但不得把共同体机制异化为针对非成员的排斥、胁迫或集体施压工具。发起主体不垄断规则的解释权与执行权——初始框架由谁提出，不决定制度由谁拥有。以上描述的是本文的制度构想，不代表现有组织章程已经包含上述安排。共同体若演化为一个阵营对抗另一个阵营，即违反其自身定义——阵营化本身列入第二十章情景A的预警信号。

分层参与阶梯：不以一次性政治表态替代长期验证

共同体不预留一个等待某国坐下的固定空椅子，而建立公开观察→安全信息接口→技术观察与演练→单项协议参与→联合项目合作→有限议题治理→正式成员的分层阶梯。任何主体可以长期停留在某一级，也可以按条件升级、降级或退出。参与越深，增量权益越多，责任与审计义务同步增加；基本安全接口不封闭，不加入不受报复，增量权益不脱离责任。

三、全球 AI 能力与治理进程正在沿四条轨道同时推进

【作者研判】

当前全球人工智能治理并非没有行动，而是行动分散在四条彼此独立、节奏不同的轨道上：

产业市场轨道：以算力扩张、模型迭代、产品落地为核心驱动力，其核心逻辑是效率与先机。这条轨道速度最快，但在速度、资本回报和市场先机压力下，容易将部分安全、伦理和社会成本外部化。

学术共同体轨道：以可验证知识生产为目标，其核心逻辑是同行评议和可重复性。这条轨道为理解 AI 能力边界提供了知识基础，但产出周期慢于产业迭代速度。

政府技术网络轨道：以联合测试、风险语言统一、技术标准对接为内容，其核心逻辑是操作性安全。目前已在小范围内建立前沿人工智能安全联合测试、风险通报和红队演练机制，但覆盖范围有限、约束力弱。

政府间国际组织轨道：以发展、治理、能力建设与国家协作框架为内容，其核心逻辑是规则制定与合法性生产。2026年7月的上海大会和世界人工智能合作组织是这一轨道的最新进展。

核心判断：四条轨道各有其功能，但真正的全球风险在于它们各自加速运行，却尚未形成稳定接口。产业在加速能力，学术在理解能力，安全网络在测量能力，政府间组织在尝试治理能力——四者之间的落差正在扩大。本文的机制设计旨在为这四条轨道之间的接口提供结构性思路。

四、人类命运共同体为何由愿景进入生存结构

【作者研判】

"人类命运共同体"在过去更多地被表述为一种外交愿景、一种价值倡导或一种理想主义的终极目标。在核威慑平衡和缓慢全球化时代，这种表述足以维持其感召力。但 AGI 时代改变了这一概念的性质。人工智能在军事、生物和基础设施领域的深度渗透，使风险归因与责任确认的难度显著增加，关键系统可能在部分危机场景中决策窗口压缩至分钟乃至秒级。人类命运共同体不再是"我们应该追求的理想"，而是"我们必须在崩塌之前建成的生存结构"。

这不是突如其来的理想，而是非零和博弈圈层不断扩展的历史逻辑延伸。人类早期的合作边界通常局限于血缘、部落与近距离共同体；随着城邦联盟、主权国家体系与区域合作机制形成，承认外部主体持续存在并与之建立稳定规则，逐渐成为文明扩大非零和空间的重要方式。而今天，AI、生物风险和基础设施互联，正在把"区域"的边界推向"全球"。但这一次，扩展的速度必须快过崩溃的速度。

五、现有全球治理工具为何仍不够

【作者研判】

现有工具各有必要功能，却不能单独承担 AGI 时代跨境风险与共同能力的完整治理。市场能够提高局部配置效率，却不会自动保护共同存续底座；国家治理能够承担国内法律、财政与安全责任，却无法独自治理跨境风险和资源互锁；国际合作能够形成项目协调，但若缺少稳定的权利、审计、退出与共同能力层，便难以转化为可持续制度。

同时必须承认三项现实：主权国家仍是公共权力与国际承诺的核心容器；竞争不会因倡议而消失；关键资源与生态承载力存在边界。本文提出的不是绕过国家、竞争和稀缺，而是在这些现实之上建立有限、可审计、可退出的新接口，以处理决策窗口压缩、归因困难和风险跨域传导。

第一部分结论：共同体不是国家利益的替代

【作者研判】

人类命运共同体不是要求国家放弃自身命运，而是要求国家看见自身命运的完整结构。正因为每个国家都必须保护自身安全、发展与持续，而三者已经部分存在于彼此之间，各国才需要共同维护一个不会整体崩塌、不会永久封死任何国家未来、能够持续纠错的世界。

2026 年上海出现的制度接口不能证明共同体已经形成，也不能证明参与国已经具有同等承诺强度。它证明的是：在竞争继续存在的条件下，国家仍然可以开始建设超越单一竞争关系的共同能力。接下来的问题，不是理念是否动人，而是资源共同体、治理共同体和向善智能体系能否在权利、利益、违约与危机约束下真实运行。

第二部分·实现（上）：从零和现实建立国家层共同存续底座

【导读】本部分回答：在竞争永不停歇的现实条件下，共同体如何从概念变为可操作的制度——分离原则与升维条款、脆弱到存续的三层演进、第一张谈判桌、资源互助与三账清算、主权协同边界。

【制度构想】本部分及以后各部分的机制设计，均为结构性推演，接受代际验真，并受前述相位纪律约束。启动期只能形成自愿、有限、可撤回的接口；初期只扩大已经通过权利与事实验证的机制；过渡期描述制度引力，不设计胁迫；稳定期规则不得提前成为现实义务。方向正确不等于可以跳过权利、证据和失败成本，每一步都必须保留边界、退出、纠错与回滚。

六、共同安全、共同发展与共同持续的分层治理

人类命运共同体必须从"不可逆风险"开始构建，而不是从"共同繁荣"开始。前者是底线，后者是方向。

可竞争领域（允许你进我退）：芯片制造、AI 应用、电动汽车、市场份额、国际标准制定；产业政策、关税、技术封锁与反制。不可竞争领域（必须共同存续）：AGI"奥本海默时刻"的误判预防；生物实验室泄露的即时通报与联合调查；关键海底光缆、卫星链路、全球算力中心的瘫痪应对。

升维条款：

可竞争与不可竞争的边界不是静态清单，而是动态判据：发展竞争可以继续，但任何竞争手段一旦具有跨域传导性、不可逆扩散性和整体存续风险，即自动升维进入共同安全议程。判断权不属于任何单一国家，应由经参与方共同授权、程序公开且可复核的风险分级机制行使（最低证据、复核与暂停原则见精选附录三）。

机制设计：

建立"战略隔离墙"——在可竞争领域，允许甚至鼓励激烈博弈；在不可竞争领域，建立一经缔约即对参与方具有约束力的"红色热线+紧急刹车协议"。涉及战略误判风险的异常自动化行为、危机态势变化和联合核查触发信号，应建立最低必要通报机制。通报的不是全部部署，而是足以防止误判的信号。各国采取必要防御措施的权利不受影响，但应同步启动多方科学核查。

据公开报道，主旨讲话强调应共同反对在人工智能领域泛化国家安全概念、把本国安全凌驾于他国安全之上的做法——这与分离原则的方向一致。

七、从共同脆弱到共同存续

核时代的"相互确保毁灭"依赖一个前提：对方知道你是谁、知道你在哪里、知道如何报复你。在 AGI 时代，这一前提正在瓦解——部分攻击可能难以在危机所需时限内完成可靠归因，能力隐藏于算法更新，破坏伪装成技术事故。因此，共同体的现实动力不是"善意"，而是对共同脆弱的清醒认知。这一认知须经三层演进，不可停留在第一层：

第一层，现实诊断：相互确保脆弱。

每一方越是想独立，就越发现对方手里的东西自己短期造不出来；越是想卡脖子，就越发现自己的脖子也被卡着。任何一方崩塌都会连带伤害他者。善意不需要，但谎言藏不住；信任可以没有，但恐惧谁都逃不掉。深度耦合是可核查的行业事实：中国在全球多类稀土分离、精炼与加工环节占据高度集中份额；美国企业在芯片设计环节占据全球领先地位（按 SIA/BCG 等公开研究口径，美国企业在全球半导体设计与销售中保持领先）；先进制程制造高度集中于少数代工企业；欧洲在部分高端装备、精密制造及关键工业能力上拥有短期难以替代的节点——没有哪一方能在不承受巨大损失的前提下完成"完全脱钩"（相关数据口径、时间点与统计边界纳入研究储备持续核验）。其成立前提必须钉死：脆弱是共同敌人，不是单方面人质——脆弱性源于系统性深度耦合，而非任何一方的战略弱点；一旦脆弱被武器化为"我知道你离不开我"的勒索筹码，机制立即失效。复合相互依存理论（基欧汉与奈）为本文提供了重要分析工具，尤其是对敏感性、脆弱性、非对称依赖与多渠道关系的区分。但本文并不把当代大国关系整体等同于复合相互依存，也不认为经济耦合会自动产生和平；本文关注的是，如何把已经存在、且可能被武器化的不对称依赖，转化为多节点、可替代、可核查的共同存续结构。

第二层，制度转化：相互确保依存。

把不可控、可勒索的依赖，改造为多节点、可替代、可审计的互助结构。防勒索的结构设计是资源池的"多对多"原则：任何单一供应方持有的关键资源占比设定上限（集中度阈值须依据资源类别、替代周期和危机状态另行建模），确保不存在"唯一供应方"；当替代来源真实存在、且能够在可接受时限和成本内启用时，单一供应方的勒索能力将显著下降。透明度的目标不是要求各方预先互信，而是降低对主观信任的依赖，把合作基础尽可能转移到可核验数据、独立审计和重复验证之上：卫星监测、算力审计、数据脱敏共享构成可核查的透明度基线。

第三层，成熟状态：相互确保存续。

任何一方的安全不再依赖另一方更弱，而依赖共同系统保持开放、冗余与可修复。完整演进路径：看见共同脆弱→建立可控依存→形成共同存续。以"维持脆弱"为长期目标是危险的；脆弱只是入口，存续才是目的地。

相关博弈模型与行业数据口径纳入研究储备；主文只保留其已被验证的结构性结论。

八、第一判

人类命运共同体不需要从"世界大同"开始。它可以从一个最小化、可操作、不触及核心主权协同的谈判议题开始。这张桌子的议程仅限三项：

1. AI 误判紧急联络热线：任何一方的预警系统发现可能导致战略误判的高风险异常信号时，应立即提交本国依法授权的危机响应机构复核；经复核达到协议约定的通报阈值，或因时间紧迫无法完成充分核查但继续沉默可能造成更大风险时，通过热线发送最低必要的预警信息，并在协议预先约定的最短时限内启动联合核查。AI 可以检测和预警，不能自行完成国家间战略通报。这不需要信任，只需要恐惧——双方都害怕"自己没来得及解释"就开战。

2. 生物异常事件联合溯源机制：任何不可解释的公共卫生异常事件，在证据不足时，不宜公开作出不可逆的攻击归因；各方采取必要防御措施的权利不受影响，但应同步启动多方科学核查。

3. 关键跨境民用基础设施最低保护机制：对承担广泛民用导航、通信、金融、医疗、灾害响应与公共算力功能的设施，原则上不得实施以造成系统性民用瘫痪为目的的攻击，即便在冲突升级阶段；军民两用设施的范围、例外和核查标准须另行谈判。这不是仁慈，是理性——谁先砸烂这些，谁先瘫痪自己的全球依赖系统。

这三项议程可以在军备控制、危机沟通、信任建立措施与民用基础设施保护的制度谱系中找到先例——热线、事故通报、联合核查与有限不攻击承诺，为 AI 与生物风险治理提供了可借鉴的制度语言；但新风险的难归因性、自主性与跨域扩散，仍需要重新设计而非简单平移。

启动方式：发起者与参与者可以分离。不必由大国直接牵头，可以由联合国框架、区域性联盟或中立第三方以"科学风险评估"名义召集，消解"谁先提谁示弱"的政治压力。先开口的不是让步者，而是定义了"什么风险值得谈"的人。2026 年上海的实践表明，多边人工智能合作的组织性谈判桌已经出现；但本文提出的军事误判、生物联合核查与关键基础设施保护议题，仍需进入独立、专业且具有安全授权的谈判进程。

九、国家间资源互助与清算机制

"国家间资源池"是统称，内部不是一个仓库，而是由三类机制组成的资源互助网络：

机制类型	适用对象	权利转移性质
实物储备与应急调拨	粮食、医疗物资、能源储备、关键矿产	实物消耗后所有权发生转移
能力预约与服务兑现	算力、物流、制造能力、通信冗余、应急医疗能力	被调度的是服务能力，非实物所有权
授权访问与许可互认	数据、专利、软件、技术标准、模型接口	转移的是有限许可，非所有权

资源池原则上不改变存入主体对原始资产的基础权利，但不同资源类型适用不同的权利安排：实物消耗类资源在调用时依法完成所有权或处置权转移；能力服务类资源按预约和交付规则结算；数据与知识产权类资源仅开放约定范围内的访问、使用或许可权。

三账分离：

资源池记账实行所有权账、使用权账、清算账的严格分离：

- 所有权账：登记进入资源池本身不改变所有权；实物消耗类资源在成员调用、交付和验收完成后，依据协议办理所有权或处置权转移；能力服务类和授权许可类资源不发生原始资产所有权转移。
- 使用权账：记录哪部分资源在何时何地可被调度。
- 清算账：记录谁存入多少、谁调用多少、谁还持有多少调用权。

三账模型的最低结构见精选附录二；跨品类清算、延期交付、应急调用与违约补偿须根据资源类别和真实运行数据另行建模，不在本文预设。

跨国资源共管并非无先例的构想：奥斯特罗姆关于清晰边界、参与式规则、分级处置、低成本争议解决与嵌套治理的研究（《公共事物的治理之道》，1990；2009年诺贝尔经济学奖），为资源池提供了重要制度参照。但这些经验不能直接证明国家间战略资源共管必然可行——主权竞争、安全例外、能力不对称与国际执行缺口，构成从地方公共池治理扩展到国家间尺度时必须重新解决的问题。本文的三账分离与补充规则，是对这些原则的跨尺度尝试，而不是其有效性的既成证明。另须说明："国家间资源池"是统称，并不意味着所有资源都具有传统公共池塘资源的经济属性；奥斯特罗姆框架主要提供治理原则，不是对全部资源类型的统一分类。

核心原则：

- 存入是自愿的；
- 调用资格具有非歧视性，具体资源调用受库存、优先级、应急状态与清算余额约束；
- 退出是允许的——任何国家可按规定程序退出。

资源池运行还需配套超额预约、价值折损、违约交付、紧急冻结、跨品类换算与退出结算等规则。相关参数属于专业验证内容，不作为理解本文的前提。

资源机制的有效前提与复杂度边界

三账分离能够把所有权、使用权与清算责任说清楚，却不能凭账本制造不存在的履约意愿和现实交付能力。任何国家间资源机制都以参与方具有最低履约意愿、真实交付能力和可执行的国内授权为前提；制度的作用是降低不确定性、识别违约、配置替代和组织救济，而不是假装可以绕过国家与现实能力。

共同授权机制也不能在启动期被预设为一个已经存在的全球中心。启动期从双边、少数多边和单项资源协议开始；初期建立联合项目、审计和清算接口；过渡期再形成多节点互认；只有经过长期运行与明确授权，才可能进入有限共同治理。

制度复杂度本身也是风险。任何资源、贡献和审计系统的复杂度，不得超过参与者能够理解、治理者能够维护、独立审计者能够复核的上限。无法在合理成本和时限内有效审计的机制，不得正式运行，只能停留在影子记录、人工抽样或有限验证阶段。

差异化接入：

国家层面的资源投入与公共能力不分规模和形式；接入规则则应考虑发展阶段、资源禀赋、技术能力与现实需求，避免以既有能力差距固化结构性不平等。有限算力条件下的有效产出可探索按稀缺度和公共价值加权。本土医药、农业适应技术及其他传统知识，首先由本地专业机构完成来源、有效性、知识权利与社区授权确认；进入跨国调用与清算体系前，还须完成专业复核、利益分享、知情同意与争议处理。具体加权和保底额度属于后续验证内容。

差异化接入解决的是发展中国家的结构性弱势；对技术强国与发达经济体，接入的必要性同样成立，但进入理由、责任承担、利益结构与合作路径并不相同——分离原则、最低必要通报和共同安全议题对所有主体开放，具体参与方式则应依据其能力、风险与责任分别设计。必须承认现实阻力：地缘政治阵营化可能使部分国家不愿与对手共同存入资源；技术垄断集团可能抵制资源互认；发展中国家在"存入—提取"体系中可能处于结构性弱势。这些约束需通过差异化准入、普惠基金以及反歧视、反支配、利益分享与弱势成员保护条款逐一对冲。

十、主权协同边界

共同体资源池不构成主权让渡。所有机制均基于自愿参与，保留退出权。但"主权"不是一个开关，而是一个洋葱——核心层寸土不让，外壳层可以协同。

层级	内容	协同程度	退出机制
核心层	核武密码、核心战略数据、国防机密	零协同——信息绝不披露	无

层级	内容	协同程度	退出机制
缓冲层	非涉密算力冗余、民用数据样本、非敏感技术参数	信息协同，物理保留——数据可追踪、可删除、可审计	按协议退出：终止后续访问与处理，尚未使用的数据删除或返还；已合法完成的调用记录、审计证据按最小必要原则保留
交互层	跨境资源提取、应急调度、联合溯源	根据资源类型，协同配置实物所有权、服务受益权、有限使用权或许可权；具体权利变化以调用协议为准	协议约定的缓冲期限；剩余资源按约定返还或折价清算
外壳层	标准互认、透明度基线、伦理框架对齐	规则协同，无资产让渡——共同制定规则，各自国内实施	依照成员国内程序和共同协议生效

关键原则："存入"不是"上交"——资源池原则上不改变存入主体对原始资产的基础权利，不同资源类型适用不同的权利安排；信息可撤销——缓冲层披露的数据必须可追踪、可删除、可审计；退出有路径——退出成本透明、可预期，不因退出而遭受报复性制裁。主权保留的绝对底线：核心战略资源数据不公开权、国家安全例外条款、协议退出权、国内制度自主权。

治理共同体：资源共同体的制度运行层

共同体资源池不由单一全球中心占有和控制。更安全的形态，是由国家节点、区域节点、专业资源节点、城市与实验区节点、国际清算节点和独立审计节点构成的共同体资源协同管理体系。

治理共同体至少承担：规则共建、权利确认、资源登记、供需匹配、调度与清算、风险分级、争议仲裁、独立审计、违约响应、退出结算和规则纠错。管理权、审计权、资源决策权与文明表决权必须分离；任何中心不得以提高效率为由取得未被明确授权的资源所有权、个人决定权或主权利力。

治理共同体的多中心权力结构

治理共同体不只是一组治理功能，还必须把决定规则、执行规则、审查规则、处理争议和终止偏航的权力分开。其最低骨架如下：

权力位置	核心职责	主要承担主体	刚性制衡
规则制定权	制定有限议题规则、授权边界与修改程序	参与国家、合法公众代表与专业主体共同组成的规则节点	不得同时控制执行与审计；重大修改须公开理由、影响与异议
执行与协调权	履约协调、资源接口、公共服务与危机处置	分布式国家、区域、城市及专业执行节点	只能在授权内执行，不得自行修改规则或扩大权限
独立审计权	核验资源、数据、算法、程序与权利影响	与执行主体分离的多法域审计节点	不得承担被审计事项的运营、收益或仲裁

权力位置	核心职责	主要承担主体	刚性制衡
争议仲裁权	处理履约、权利、权限与退出争议	多元仲裁与复核结构	不得由争议一方控制；裁决须可复核、可申诉
文明否决与终止权	阻止不可逆偏航，启动暂停、回滚与相位重申	合法授权的公众表决、人类终审与紧急制动结构	不得被资本、技术能力或贡献值购买；紧急权必须自动失效

上述权力结构遵守四项原则：多中心，不建立单一全权中心；授权有限且可撤回；相邻否决性节点不得由同一主体控制；紧急权只用于限时止损，期限届满自动失效并接受独立复核。

资源共同体回答资源如何进入真实使用；治理共同体回答权力如何获得授权、受到约束并在失败时被终止；向善智能体系回答新能力如何建设、调用、审计与拉闸。三者相互依赖，但任何一根支柱都不得吞并另外两根。

向善智能体系：共同体的新能力层

资源共同体和治理共同体要在全球尺度上运行，需要人工智能帮助发现分散资源、识别真实需求、解释不同规则、进行风险推演、连接多种语言与制度、降低协调成本并放大国家和个人能力。

向善智能体系必须满足六条刚性边界：人类保有最终决定权；国家与个人保有核心数据和授权控制；基础安全智能不以完整成员资格为前提；技术支持不得制造永久依赖；重要模型、规则和结果可审计；智能不得决定人的基本权利、人格价值和生存资格。

制度结构：多节点建设、权限分离、独立审计与人类终止权

向善智能体系不由单一国家、单一企业或单一模型垄断建设和运行。更安全的结构，是由国家公共节点、区域与城市节点、专业领域节点、受监管企业节点和独立审计节点共同构成；不同节点可以使用不同模型与技术路线，但必须遵守共同的权利边界、证据格式、异常通报和退出规则。

核心目标、权限边界、数据使用范围和不可逆规则的修改权，必须与模型开发权、日常运行权、审计权和终止权分离。开发者可以改进能力，不能单独改变文明目标；运行者可以调用系统，不能自行扩大权限；任何高影响修改都应留下版本、理由、影响评估、异议和回滚点。

审计至少包括技术安全审计、数据与授权审计、权利影响审计、长期依赖审计和独立外部复核。审计者同样必须接受权限限制、利益冲突披露和审计质量复核，防止“审计”本身成为新的不受约束权力。

多个智能系统之间应通过开放、可替换、可验证的接口协同，而不是汇聚为单一全球模型或单一智能中枢。互操作的对象是最低必要的身份、授权、证据、异常和清算格式，不是要求各国交出全部数据、模型权重或治理自主权。

出现重大错误、权限越界、数据滥用、不可解释歧视、系统性依赖或不可逆风险时，国家、合法治理主体和受影响个人应分别保有与其权利范围相适应的暂停、拒绝、申诉和退出权。高影响系统必须具备人工终止、降级运行、版本回滚和证据保全能力；提出建议、批准执行、持续监测和最终拉闸不得由同一主体独占。

人工智能的文明价值，不只是让少数国家更快领先，而是让更多国家获得原本无法以同等成本获得的教育、医疗、产业、风险识别与治理能力，并保有形成自身答案的自主性。

第三部分·实现（下）：为文明跃迁建立个体制度接口

【导读】本部分回答：国家层的资源安全为何不能替代个体层的意义制度——核心命题的准确表述、个体资源机制、贡献值的性质与边界、三权分离与动态仲裁、三大元目标审查。

【制度构想】

十一、国家间资源池为何不能替代个体制度

国家间资源池解决的是国家层面的资源安全互助。但国家层面的互助并不能自动解决个体层面的危机。当 AGI 替代了越来越多功能性劳动，"被需要感"的消逝成为比贫困更深的危机——但必须指出，这一判断成立的前提是基本生存已得到保障。在物质匮乏尚未解决的地方，贫困仍是第一位的生存问题。

核心命题的准确表述：

旧逻辑：我要更安全，就让你更弱；我要更有价值，就比你更强。新逻辑不能表述为"我被需要，所以我存在"——那仍然把人的存在资格绑定在功能之上。准确的表述是：我存在，所以我的基础价值不可否定；我关联、照护、创造、承担，所以我的生命不断涌现意义。

三种"被需要"的严格区分：

- 功能性被需要——系统因为你能产出而暂时保留你；
- 关系性被需要——生命之间真实的关心、依赖与回应；
- 存在性价值——不因能力、产出和贡献而被取消。

文明跃迁要终结的，是功能性被需要对生存资格的垄断；要张扬的，是关系性被需要这一意义引擎；要用制度锁死的，是存在性价值——生存保障是出厂设置，先于一切贡献叙事。

效率兼容论证：新逻辑不是消灭竞争，而是升级竞争。

旧逻辑中的竞争机制——国家竞争、市场博弈、技术竞赛——在人类数万年演化中并非"错误"，而是特定资源约束下的适应性策略。竞争具有筛选功能：淘汰低效制度，推动技术进步，激发创新活力。简单取消竞争，会导致系统性效率损失与搭便车问题。新逻辑的目标是改变竞争的方向：在可竞争领域，芯片竞赛、AI 竞赛、市场竞争继续存在，甚至更加激烈，但获胜标准从"让对方无法生存"改为"为共同体资源池贡献更多增量价值"；在不可竞争领域，竞争被协同取代，目标是"共同不失败"而非"我比你更安全"；在个体层面，贡献值体系保留了"争上游"的人性动力，但把"上游"重新定义。人可以因贡献被看见、资源得到支持而更愿意创造、照护与承担，但系统不得以贡献值诱导过度劳动、自我损耗、关系表演和功利性利他。任何使人为了分数持续压榨自身的机制，都属于贡献值体系的失败。

十二、人员与个体资源池的全域可达、可用与自主流动

个体资源池的定位

个体资源池不是由平台建立的个人全景档案，也不是把人的行为、能力、关系和价值折算为统一分数。它是由个人主导、分域保存、按需显现、逐项授权、可以撤回的资源与能力集合，包括时间、知识、技能、经验、作品、数据、关系、合法财产、关心与照护能力，以及借助人工智能形成的能力外延。

人不是资源。个体资源池属于人，并服务于人的自由、发展与意义。基础医疗、基础教育、基础照护和其他基本公共服务，不得因个人是否登记、共享、贡献或流动而产生等级差异。

全域可达

全域可达解决“能不能进入”。在合法、安全、承载能力允许的条件下，个人应当能够看见并进入更广泛的地域、行业、组织、教育、医疗、生活、创造和公共参与场景。可达不等于取消边境、主权与公共安全管理，而是减少不合理的地域、身份、信息、资格与组织壁垒。

全域可用

全域可用解决“进入后能不能真实生活、参与和创造”。共同体资源应依法向个人开放；个人进入新的节点后，其资格、学习成果、合法数据、作品、专业记录、社会保障与必要权益应当能够在协议范围内继续使用。个人自己的知识、技能、设备、空间、数据和服务能力，也可以在本人授权下跨域使用，而不因离开原有组织自动失效。

自主流动

自主流动解决“谁决定进入、停留、改变与退出”。个人有权决定是否迁移、是否远程参与、开放哪项能力、授权多久、向谁开放、何时撤回、是否返回原节点。流动权同时包括不流动、拒绝系统匹配、退出与重新选择的权利。任何人不得因拒绝迁移、拒绝某项工作或拒绝共享个人资源而失去基本生存与基本公共服务。

三种流动同时存在

第一，人员自主流动：个人自主选择生活、学习、工作、照护与创造场景。第二，个体资源自主流动：知识、能力、作品、时间、设备、数据与AI能力外延，在本人授权下跨越空间和组织边界。第三，共同体资源向个人流动：教育、医疗、算力、设备、资金、服务和发展机会主动抵达资源不足的人与地区。

人工智能的能力作用

AGI-COS 及其他受控智能系统可以帮助发现资源、识别需求、进行资格与规则解释、提供多方案匹配、说明风险和长期影响、保存授权与使用证据、发现异常与权利侵害。人工智能可以提出可能，不能替个人决定命运；可以协治理，不能垄断治理；可以放大人的能力和关心，不能把人变成系统效率的附属物。

运行机制的层级说明

人员与个体资源池是共同体进入个人生活的关键接口，但共同体的运行机制并不从属于个体资源池。资源使用与向善智能贯穿国家、区域、组织和个人；贡献确认与意义承接主要进入社会和个体层；制度前语言、完整实验与高地核验则约束全部层级。其完整结构在下节独立展开。

国家间资源池与个体资源池的关系

国家间资源池解决共同体关键资源能否稳定流动、保障和清算的问题；个体资源池解决个人能否自主携带、调用、开放和使用自身资源的问题。两者通过有限授权接口连接，不建立自动兑换关系：国家间资源池清算额不可兑换个人贡献值，个人贡献值不可兑换国家间资源池治理权重；上层资源共同体不得吞并个体资源池。

贯穿国家层、社会层与个体层的五类运行机制

资源共同体、治理共同体、向善智能体系和个体制度接口只有通过一组跨层级机制彼此衔接，才能从概念转化为真实运行。以下五类机制不是彼此孤立的模块，也不由任何单一层级独占。

运行机制	主要适用层级	核心作用	不可突破的边界
资源发现、授权与有效使用	国家、区域、组织、个体	发现闲置与短缺，建立自愿授权、匹配、使用、清算与退出接口	不以有效使用为名改变未授权所有权，不建立单一资源控制中心
贡献确认	组织、社区、个体为主	依据公开规则和可核验事实确认公共价值，为基础权利之外的相关增量资源提供参考	不评价人格，不影响基础权利，不形成跨领域总分，不兑换政治权力
意义承接与回应	个体与社会关系层	使照护、陪伴、创造、守护与真实回应不被系统遗忘	不强迫反馈，不把情感关系量化为竞争分数，不以意义记录决定生存资格
向善智能能力	国家至个体全层级	提供感知、匹配、推演、翻译、预警、审计和能力外延	人类终审，数据与授权可控，多节点可替换，不制造永久依赖
制度前语言、完整实验与高地	整个共同体	把原则转为可理解、可验证、	按相位推进，全员无损；失败

运行机制	主要适用层级	核心作用	不可突破的边界
核验		可失败、可修正的制度语言与运行事实	可暂停、退出、回滚并永久保留证据

五类机制的动态闭环

五类机制不是并列摆放的五个模块，而是共同体持续运行和纠偏的循环：资源发现与授权使沉睡资源进入真实使用；真实使用产生可核验的贡献事实；贡献确认与意义回应使参与动力获得承接；向善智能提高发现、匹配、推演与风险识别能力；制度前语言、完整实验和独立核验再检验上述机制是否偏离权利边界与三大元目标，并把纠正结果反馈到下一轮资源发现、授权和规则设计。

资源进入使用 → 贡献获得确认 → 意义形成承接 → 智能放大能力 → 制度核验纠偏 → 重新配置资源。任何一环失真、垄断或不可退出，整个闭环都必须暂停复核，而不能以其他环节的效率掩盖。

五类机制共同回答三个问题：资源和能力如何抵达真实需求，人的贡献与意义如何被承接，制度和智能如何在不吞并权利的条件下持续运行。任何机制若只能依靠其他群体先受损、依靠不可审计的复杂度或依靠单一智能与权力中心才能成立，均不得进入下一相位。

十三、贡献值的准确性质与意义回应记录

贡献值（CV）是依据公开规则、可核验事实与可申诉程序，对个体行为所产生公共价值作出的制度性记录，用于确定基础保障之外、与贡献类型合理相关的增量资源使用优先级。它不评价人格，不影响基本权利，不兑换政治权力与公民身份，也不建立跨领域统一总分。

体系注：中文理论体系继续使用“贡献值”这一术语；国际传播中的英文名称不作机械直译，须经独立语义测试、法律含义审查与反向翻译验证，尤其避免使用可能被理解为人格信用、社会评分或综合评级的 credit、score、rating 等表达。英文正式名称在完成测试前不锁定。

领域化与用途限定（刚性条款）：

贡献值原则上实行领域化记录与用途限定：不把医疗、教育、消费、社交、政治表达、守法记录和公共贡献汇总为一个人的统一分值；每类贡献记录只能用于事先公开、与其具有合理关联的增量资源场景，不得跨领域自动迁移；一项领域内的低贡献、争议或违规，不得自动降低个人在无关领域的资格、服务与社会评价。任何将不同生活领域汇总为统一人格、信用或公民等级的系统，均不属于本文所述贡献值体系。

需求被回应即意义记录：

弱势群体的需求被有效满足，记录为意义回应，进入独立的意义回应账，不形成贡献值、不兑换资源使用优先级——它不是生存的理由，是意义的确认。生存无条件，先于一切贡献叙事；接受帮助不需要被包装成贡献才能获得尊严，尊严先于贡献。

体系衔接注：本文"意义回应记录"即文明跃迁实验区文档中"需求被满足即贡献"原则在记账层的实现形态——原则陈述的是因果链（没有接受者，利他无法完成），记账实行的是双账分离。两种表述属同一原则的不同层次，不构成两套规则。

价值标准、双账结构、人工复核、造假处置和跨区域互认属于后续专业研究内容；主文锁定的最低边界是：不建立人格总分，不以贡献决定基本权利，不跨领域自动连带，不把贡献记录兑换为政治权力或公民身份。

可审计简化原则

贡献记录不是越精细越好。若一个领域的记账、核验、调用和审计复杂到普通参与者无法理解、独立审计者无法在合理时间内复核，系统就已经产生新的黑箱。应优先采用最少字段、最小用途、最低侵入的记录方式；无法简化的领域暂不进入正式资源配置。

十四、三权分离与动态仲裁

适用层级说明：

本文所述三权分离、文明表决权 and 人心与意义审查，首先适用于国家级综合实验区及其他自愿采用该结构的治理单元。进入国家间合作层面后，任何表决、资源权利和强制性义务均须以国际协定、成员国国内法和明确授权为前提。国际组织可以建立共同原则、审查接口与争议解决程序，但不得在未经授权的情况下越过主权国家，直接行使国内公共权力。

共同体的制度运行需要三种权力类型，彼此制衡、不可集中：文明表决权——一人一票，贡献值清零，用于不可逆文明航道决策；领域决策权——依据经验证的专业能力、实践经验、责任承担和相关贡献综合确定，贡献记录可以证明参与资格和实际经验，但任何资金、资源或声誉贡献均不得自动兑换专业判断权；资源使用权——按贡献值分级配置增量资源，与前两种权力严格隔离，不得兑换表决权重。

权力类型与程序节点不可混同：启动、执行、暂停、终止是任何重大事项都要经过的四个程序节点，每个节点由相应权力类型行使。最低冲突仲裁原则见精选附录三，完整程序纳入研究储备。

有限责任豁免：

若终止程序触发后推动者一律重罚，会导致"多做多错、不做不错"的逆向激励。因此：决策者已尽到合理注意义务、完成必要信息披露、遵守程序、披露利益冲突，且不存在故意隐瞒或重大过失的，即使结果失败，也不承担惩罚性责任，只要求公开复盘并纳入经验库。

区分责任追究与资格复评：

- 责任追究：故意、重大过失、隐瞒、违规；
- 资格复评：能力不足、连续判断失误、风险控制能力下降。

不能把能力不适任直接等同于恶意责任。

层级注释：

本文三权针对文明航道级决策（表决层）；《国家级综合实验区建议书》中治理者宪制化持有的实验区启动权与终止权，属行政授权层的先行先试安排——二者层级不同、对象不同，并行不悖。

十五、三大元目标审查

一切制度提案，除效率、资源与风险审查外，须通过人心与意义委员会的审查。但否决权必须分级，不可由单一委员会无限垄断：

刚性否决（委员会可直接暂停）：

- 基础生存权被贡献值条件化；
- 未经授权的人心与心理监测；
- AGI 黑箱直接剥夺基本权益；
- 不可申诉、不可退出、不可纠错；
- 人格总评分或系统性歧视；
- 明确将人压缩为评分对象。

暂缓否决（委员会只能要求暂停、补充论证和公开重审）：

- 幸福体验可能下降；
- 关系结构可能变薄；
- 意义结构可能失衡；
- 技术便利与人际关系发生冲突；
- 不同群体对幸福和意义理解存在分歧。

第二类事项委员会不能永久否决，只能暂停—公开说明—补充调查—重新论证—交由相应决策程序裁定。委员会自身的产生方式、任期、组成与责任边界由宪制程序规定，接受独立审计与罢免——人心高于效率，但人心的解释权不归任何少数人垄断。

第四部分·演变：从护栏到结构，从恐惧到预期

【导读】本部分回答：底座搭建之后如何演化——三阶段路线、实验区的试验场角色、高地的可见性与可学习性、失败情景与熔断。

【作者研判+制度构想】

十六、共同安全—共同发展—共同持续

共同体的演变不是从一个口号走向另一个口号，而是三种国家能力逐步进入共同运行。

第一阶段，共同安全：守住不被共同毁灭的底线。落地热线、联合核查、最低保护承诺、关键资源应急保障和基础安全智能接口；首批国家间资源池、个体资源池与城市级完整实验启动。共同安全不要各国停止竞争，只要求竞争不能穿透共同存续底座。其初始动力主要来自共同风险与共同脆弱。

第二阶段，共同发展：使更多国家、地区与个人获得真实的发展通道。资源池从应急保障扩展到能力预约、知识与技术授权、算力与公共智能基础服务；人员与个体资源池的资格、数据、成果和权益开始跨节点衔接；弱资源国家和地区获得形成自身能力的支持，而不是形成对单一提供者的永久依赖。其主要动力来自可计算的互利和能力增益。

第三阶段，共同持续：使安全与发展能够跨越时间。共同体开始对资源循环、生态承载、技术风险、人口与财政结构、代际影响和制度异化进行长期推演与纠错；高地效应规模化扩散，参与者开始依据共同结构安排长期投资、安全和发展。其主要动力来自对未来的共同预期。

共同存续是共同安全的底线，共同预期是共同持续形成后的结果。安全决定国家能否存在，发展决定国家是否拥有未来，持续决定这个未来是否真实。

十七、从专题合作到共同体结构的五级演变

演变不仅是时间推进，更是制度形态的逐级升级——合作内容、权力结构、信任来源与长期预期在每一级都发生质变：

第一级，危机联络：热线、事故通报、联合核查——信任来源是共同恐惧。第二级，项目合作：联合测试、算力备份、灾难协作——信任来源是单项互利。第三级，规则互认：评测方法、审计原则、数据接口——信任来源是程序可预期。第四级，有限共同治理：争议解决、独立审计、共同基金——信任来源是制度本身。第五级，共同预期形成：参与者开始依据共同结构规划长期投资、安全与发展——信任来源是对未来的共同押注。每一级都不要求跳级，也不承诺自动升级；升级发生在上一级的运行记录足以支撑下一级的风险敞口之时。

从倡议到采用：政治启动与制度承载链

国家级文明实验不能只靠一位领导、一个部门、一个企业或一次批示。个人可以打开入口，却不能独自承受法律、财政、数据、舆论、失败和跨任期连续性。真正需要建立的是一条可以跨越个人任期和部门边界的制度承载链。

位置	核心责任	不能替代的功能
政治发起者	把议题送入国家级前瞻研究与决策程序	承担首次提出和启动的政治责任
行政承载者	提供真实城市、人口、公共服务和跨部门协调	把授权转化为可运行的社会系统
法律与财政保障者	界定试验边界、权利保护、失败成本和救济	保证试验不是无法律保护的冒险
专业实施共同体	负责资源、智能、数据、审计和社会研究	防止政治意愿替代专业能力
独立核验与公众合法性承载者	保护异议、申诉、退出并公开失败	防止试验退化为政绩工程或自我证明

现阶段适当的决策请求不是一次性批准完整新系统，而是分四级推进：先决定是否将其纳入国家级前瞻研究；再决定是否建立跨部门研究与论证机制；随后选择候选城市开展基线调查与方案预研；只有在法律审查、公众沟通、最小闭环验证和独立评估通过后，才进入城市级综合实验授权。目标不降维，决策不跳级。

十八、国家级文明跃迁综合实验区

资源共同体、治理共同体、向善智能体系以及人员与个体资源池的全部机制，需要一个缩域不降维的完整试验场。实验区不能只是单一应用试点，而应具有完整人口、产业、公共服务、资源供给、社会保障和治理结构，至少达到规模化城市级承载。

缩域是为了控制风险，不降维是为了保留真实社会的复杂度。社区、企业和平台可以验证单项机制，却不能验证资源、权力、技术、人心、公共服务和长期时间如何共同运行。城市级承载不是追求规模，而是为了让新逻辑接受真实利益、真实冲突、真实普通人和真实失败的检验。

实验区的方案本体见《AGI 时代文明跃迁国家级综合实验区建议书》；本文的国际层机制与实验区的关系是：实验区验证资源权属与使用分离、共同体资源协同管理、三权制衡、向善智能、人权边界、人员全域可达、个体资源池自主授权和安全—发展—持续三目标能否形成闭环；国家间资源池为验证提供外部资源安全和跨区域接口；实验区的运行数据、失败记录、申诉、退出和纠错结果反过来为国际层参数校准提供事实基础。

实验区必须从第一天建立国际独立核验接口。它要证明的不是“系统能跑”，而是系统能在不压缩人的尊严、不制造新的依赖、不把贡献值变成人格评分、不把人工智能变成最终裁决者的条件下，真实提高国家、社会 and 个人的安全、发展与持续能力。

十九、高地可见性与信息真实性

高地效应揭示了一条不同于“打破”旧逻辑的路径：不是靠对抗，而是靠超越。高地是率先实现新逻辑闭环验证的局部。高地不需要攻击低地，只需要保持自身运行并保持可见；当足够多的人亲眼看见“另一种可能性真实存在且确实更好”时，旧逻辑的解释力就会下降。

必须保护：

- 公开批评；
- 负面报道；
- 对数据和指标的质疑；
- 讽刺、反对和独立调查；
- 对实验失败的公开讨论。

可以追责：

- 伪造实验数据；
- 篡改审计证据；
- 冒充公共机构；
- 网络攻击；
- 故意制造虚假文件；
- 破坏数据完整性。

高地的第一条条件不是自我宣布成功，而是其真实运行结果能够被怀疑者乃至敌意主体独立核验。高地不仅要证明效率提高，还要证明安全水平提高、资源闲置与短缺同时下降、个人实际选择增加、授权能够撤回、申诉真实运行、AI 支持没有制造永久依赖、失败可以公开、退出不受报复、弱地区获得了自主能力。

国际可信度与跨制度可验证性

一个在中国制度环境中启动的实验，不能仅凭中国内部成功就宣称适用于世界。它的国际可信度应分四层建立：第一，国内真实性——运行、申诉、失败与退出均留下可核验记录；第二，独立可核验性——允许不同立场的专业主体访问必要证据；第三，国际观察与参与——逐步开放观察员、联合课题、单项共同试验与审计节点；第四，跨制度复制——证明底层原则和接口能够在不同法律、产权、政治与文化条件下重新实现，而不是要求复制单一中国模板。

多国共同设计、共同出资可以增强可信度，但不能成为启动前提，否则会形成“先有国际参与才能证明可信、先证明可信才有国际参与”的死结。中国可以自主启动，外部核验接口必须从第一天存在；高地输出的是可验证的原则、边界和运行事实，不是要求世界照搬一种制度形式。

高地信息真实性与独立核验机制：

- 高地广播协议——高地的制度规则、评价方法、总体运行结果、失败记录和公共资金使用情况原则上公开；涉及个人隐私、商业秘密、国家安全、未公开知识产权和系统防护细节的内容，按最小必要、独立审计和可解释豁免原则处理，任何保密安排不得用于掩盖制度失败、数据造假或权力滥用；
- 高地观测基金——资助独立记者、学者、公民实地访问；
- 事实核查联盟——由多国独立机构交叉验证；
- 多节点镜像存档——独立学术数据库、第三方审计机构与依法运行的冗余发布渠道。

二十、失败情景与退出机制

三种失败情景及熔断机制：

情景 A，国家间资源池沦为政治工具：强国将资源池作为地缘杠杆，弱国被迫"自愿"存入，退出机制名存实亡。预警信号：单一国家存入占比超限；弱国退出申请被技术性拖延；评估标准被单方面修改；成员构成长期阵营化、准入被用于地缘划线。熔断：当单一供应方集中度超过风险阈值时，暂停继续提高其系统权重，启动替代供应方引入、增量采购多元化与风险敞口压降程序；既有资源仍按原协议处置，不得未经所有权主体同意强制转移。退出申请的具体时限根据资源类型、未履行义务、系统风险和替代周期另行确定。配套应对：存入记录实时公开可查；每次存入协议显性标注"不构成主权让渡"；退出设缓冲期；共同体设独立评估基金，用途限定为公开协议文本、翻译政策文件、支持第三方评估、组织公众听证、提供退出成本和资源流向说明、资助独立研究。

情景 B，CV 体系被算法垄断。预警信号：开源承诺未兑现；元审议被架空；申诉积压超时。熔断：强制开源审计；元审议在制度约定的连续周期内未能依法召开、或审议程序持续失效时，自动暂停新标准实施并回退至最近一次通过完整审议和公开审计的版本；人工复核权升级为法定权利。

情景 C，高地效应被系统性遮蔽。预警信号：广播数据被封锁；核验申请被无理拒绝；独立报道被系统性删除。熔断：启动多节点镜像存档与独立学术数据库发布。

情景 D，向善智能退化为数据抽取和永久依赖。预警信号：核心模型不可迁移；本地能力建设长期停滞；重要数据被默认回传；关键公共服务只能依赖单一外部供应商。熔断：暂停扩展部署，启动数据与模型主权审查，强制提供可替代接口、迁移方案和本地能力建设计划。

情景 E，人员与个体资源池被效率逻辑工具化。预警信号：拒绝系统匹配影响基本权益；跨域数据被默认拼接；个人撤回授权困难；流动被行政或算法强制。熔断：立即暂停相关匹配与调用，恢复最近一次通过权利审查的版本，启动独立调查、个人救济和公开复盘。

第五部分·跃迁：文明逻辑转换、长期接口与终极检验

【导读】本部分属于远景接口预研，不构成当前政策建议、近期工程路线或已证实科学结论。其作用是检验前述制度原则在更长时间尺度和更大文明尺度上是否仍然成立，并提前识别今天的制度设计不得穿越的终极禁区。当前决策不以接受本部分全部远景判断为前提。

【情景推演】以下内容均为条件式预研：只有相应技术与社会条件真实成立时，相关制度接口才进入进一步论证。

二十一、跃迁究竟发生在哪里

文明跃迁不是合作规模扩大，也不是国际组织数量增加，更不是资源池越建越大或人员跨境越多。它首先发生在国家安全、发展方式、资源价值、人的价值和智能用途的来源逻辑发生转换之中。

在国家层面，国家利益不再被理解为只存在于国境之内；安全不再主要依赖削弱、封锁和压倒另一方，而逐步依赖共同系统的开放、冗余、可核查、可修复与对方仍有未来。在发展层面，发展不再主要依赖排他占有、永久封锁和把他国锁定在低位，而更多来自资源有效使用、能力互补、知识创造和更多主体获得发展入口。在持续层面，安全与发展不再以透支资源、生态、社会和下一代为代价。

在资源层面，价值不再主要来自占有多少，而越来越取决于资源是否被真实使用、释放、调度、再生并产生公共价值。在个体层面，人的生存资格不再由功能价值决定；人员与个体资源池能够在权利保障下全域可达、可用和自主流动；人的价值感不再只能来自胜过他人，而可以来自关联、照护、创造、承担与真实回应。

在人机关系层面，人工智能不再主要用于替代人、控制人和压倒他国，而成为人的能力外延、共同治理工具、资源连接系统和文明风险预警系统。智能向善的检验，不是模型是否声称善意，而是它是否真实扩大国家和个人的自主能力，减少共同伤害，并始终受权利、审计、退出和人类终审约束。

实现建立资源、治理与智能底座；演变使资源、人员、能力与规则跨节点连接；跃迁改变文明为何运行、为谁运行以及依靠什么继续运行。

二十二、身份叠层，但不建立贡献等级公民制

身份的扩展是一个渐进过程，每一阶都在前一层之上叠加新的认同、责任和共同体关系；是否产生新的法律权利，须由相应国家法律、区域制度或国际条约另行确认。世界公民是一种文明认同和公共责任身份，不自动产生超越国家法律的权利。跨境医疗、公共服务、迁移与参与权利，必须由国家间协议、区域法律或国际条约另行确认。

碳硅契约，即人类与人工智能、AGI 等非生物智能主体围绕权利、责任、控制边界、共同安全与长期持续建立的文明级约束关系。它不生成、授予或取消世界公民身份；世界公民认同或声明对所有人开放。贡献值只影响个体在相应资源池中的增量资源使用优先级，不影响公民身份本身，也不决定人与非生物智能主体的基本权利边界。

反贡献等级公民制条款：

公民身份的任何一阶，均不以贡献值高低分级、不以贡献值购买或兑换。

二十三、行星文明接口

若人类未来具备稳定的跨行星活动能力，文明跃迁需要提前为其保留制度与意义接口——这是条件式预研，不是“物理必然”的断言。在该条件成立的情景下，星际接口的原则如下：

星际基础生存救援、人道通信、灾难避险和文明存续接口，不得因成员身份、制度差异或历史参与状态而关闭。涉及增量资源调用、工程参与、治理席位和长期开发权益的事项，可以依据参与责任、实际贡献与协议义务进行分配。

三条原则立于正文，长期技术与制度细则纳入研究储备：宇宙公民身份是面向多天体文明的普遍主体资格，不设贡献、健康与能力门槛；星际任务执行资格按能力与团队需要选拔——技能可以决定谁适合执行任务，不能决定谁更配成为文明主体；具有文明存续备份属性的稀缺迁移资源，按公开、公平、可申诉的程序分配，权属文明表决权管辖，不由资源池清算额计权、不以贡献值排序。禁区清单在此为刚性约束：不让贡献值购买一切，不让任何人因贡献低而失去基本生存保障，不以贡献值评价人的生命价值。

二十四、物理保障与心智意义持续

迈向深空，需要两套生命维持系统：一套维持身体，一套维持意义。物资、能源、空气和生态循环保证人活着；心智、关系、记忆、自由和共同愿景保证文明仍然是文明。

意义维持系统的五项制度构件：长期孤独与封闭环境下的关系承接机制；代际使命的自愿传递而非强制继承——后代拥有重新选择的自由，任何世代不得用启程时的誓言绑架尚未出生的人；文明记忆库的跨行星同步；意义反馈通道的延时适配；共同愿景的周期性重建——每一代都拥有重新讨论航行目的、共同生活方式和未来方向的权利，使命可以承接，但不能冻结。物质系统失效，人死去；意义系统失效，文明死去——而第二种死亡可以发生在所有人还活着的时候。

二十五、终极检验

文明跃迁是否发生，终极检验只有一个问题：国家、群体与个体，是否重新拥有了可期待的未来？

在国家层面：共同安全、共同发展、共同持续能够同时改善；关键资源与智能能力不被单一节点永久锁死；国家获得了自主解决问题的能力；国家存续不依赖对另一方的压倒。

在群体层面：不同文明、不同制度的群体不把对方存在视为对自身的否定；弱地区不再只是资源和人才的输出地；共同治理能够容纳分歧、异议、退出与重新进入。

在个体层面：人员与个体资源池能够全域可达、可用与自主流动；自己的能力、时间、创造、关心与合法资源可以在本人授权下抵达真实需求；自己可以拒绝、撤回、退出和重新选择；自己的存在不因“无用”而被否定。

在智能层面：人工智能真实减少了共同伤害，放大了国家与个人能力，没有以效率为名剥夺人的最终判断，没有以支持为名制造永久依赖。

结语

人类命运共同体的实现、演变与跃迁，不需要所有国家先接受同一种价值、同一种制度和同一个组织再行动。

它可以从国家对自身命运的重新计算开始：安全能否只靠威慑与封锁，发展能否脱离跨国资源与能力，今天的优势能否在一个不断崩塌的世界中持续？只要把这三个问题推演到底，共同安全、共同发展与共同持续就不再是外部强加的道德要求，而成为国家长期利益的完整形态。

它可以从一张谈判桌开始，从一次资源和能力的有限授权开始，从一个公共智能工具帮助弱资源地区形成自主能力开始，从一个城市级完整实验开始，也可以从一个人能够带着自己的权利、能力和个体资源池，自主进入更广阔的生活、学习、工作、照护与创造场景开始。

资源共同体使已经存在的无序依赖转化为可保护、可使用、可清算、可退出的共同能力；治理共同体守住主权、权利、责任与制衡；向善智能体系让资源、机会、能力与真实需求彼此抵达。它们共同服务的不是一个更庞大的管理系统，而是国家、社会与每一个人都继续拥有可以期待的未来。

2026年7月上海出现的制度起点，尚不足以证明这一转变已经完成，却足以证明：在竞争继续存在的世界里，另一种支持国家安全、发展与持续的道路已经拥有可以继续搭建的现实接口。

人类命运共同体不是国家利益的替代，而是多国长期共存时代国家利益的完整形态。

文明的跃迁不等待完美蓝图，但每一步都必须保留证据、边界、退出与纠错。

精选附录一：事实基准与现实接口

本附录只保留理解主文所必需的事实基准。动态组织状态、成员承诺强度、完整政策文本和后续变化应持续核验，不以本稿基准日冻结现实。

主要公开来源（截至 2026 年 7 月 21 日）：中华人民共和国外交部 2026 年 7 月 16 日《成立世界人工智能合作组织协定签署仪式在上海举行》；2026 年 7 月 17 日相关主旨讲话与行动计划官方材料。本文只将其作为现实接口，不据此推定本文制度构想已被任何政府或国际组织采纳。

事实	本文使用方式	边界
2026 年 7 月 16 日，29 国代表在上海签署成立世界人工智能合作组织协定	作为多边 AI 合作谈判桌已经出现的现实锚点	签署、法律生效、组织启动和实际运行必须分别核验
2026 年 7 月 17 日，2026 世界人工智能大会暨人工智能全球治理高级别会议开幕并发布相关行动文件	作为共同发展、伦理治理、能力建设与智能向善的政策接口	不据此推定本文全部制度构想已获采纳或背书
全球 AI 能力、治理、产业和安全网络仍然分散	说明需要跨轨道接口	不据此推定各国立场统一或合作必然形成

精选附录二：核心制度边界摘要

结构	最低定义	不可越过的边界
资源共同体	在所有权清楚前提下协同部分使用、保障、应急调用与清算关系	不等于共同所有、统一再分配或单一中心占有
治理共同体	有限、分层、多中心、可审计、可纠错、可退出的共同治理	不替代国家，不集中管理、审计、仲裁、执行和终止权
向善智能体系	以人类终审、数据主权、反依赖和可审计放大国家与个人能力	不决定人格、基本权利、生存资格和不可逆文明方向
人员与个体资源池	个人控制、分域保存、逐项授权、可携带、可撤回	人不是资源；拒绝流动、共享或贡献不影响基本权利
贡献值与意义回应	基础权利之外、领域相关的增量资源记录；意义回应独立记账	无人格总分、无跨域自动连带、不兑换政治权力和公民身份
三权分离	文明表决、领域决策、资源使用彼此隔离	资金、资源和贡献不得自动兑换专业判断或一人一票权重

精选附录三：相位、停止与退出摘要

任何机制进入下一相位，须同时提交事实证据、无损证明、独立评审、失败与回滚方案以及不升级的替代路径。相位切换涉及不可逆方向和全体根本权利时，必须进入一人一票的文明表决，贡献记录不得计权。

刚性停止条件	最低处置
基础生存或基本公共服务被贡献、评分或服从条件化	立即暂停并恢复权利
个人数据跨域默认拼接、形成统一人格画像	停止关联、删除未授权映射、独立审计
人工智能直接作出不可逆权利或文明决定	回退至人工授权版本
临时或危机权限无限续期、无独立复核	权限自动失效并公开复盘
退出受阻、撤回困难或因退出遭受报复	暂停扩展，完成救济和退出验证
稳定期规则被提前施加于启动期主体	退回重设计，不得以“未来利益”辩护
某一群体的净损失构成方案启动前提	认定为对立死结，重新设计动力与回报

精选附录四：研究依据与版本变更

研究依据包括：复杂相互依存与非对称依赖研究、公共池治理、多中心治理、危机沟通与信任建立措施、技术风险治理、公共权利与可审计制度设计。学术锚点用于建立可讨论的概念接口，不替代本文机制的现实验证。

数据纪律：所有事实、指标和组织状态必须标注来源、日期、口径、可信等级和更新触发条件；推断不得伪装为事实，制度构想不得伪装为现行规则。

V3.9.3 三支柱结构对称与运行闭环锁定：在 V3.9.2 已经完成向善智能制度骨架与运行层级归位的基础上，集中补足治理共同体的规则制定、执行协调、独立审计、争议仲裁与文明否决五权结构，明确多中心承担主体及其相互制衡；补明五类运行机制从资源使用到制度核验再回到资源配置的动态闭环；压缩现有治理工具局限的重复表达。完成本版后，主文正式冻结：新事实进入事实台账，新对象进入对象研究，新参数进入实验与专业储备；只有核心因果、权利边界、相位纪律或三支柱结构发生变化，才重新开启主文升版。

从本文继续前行

以下入口按本文的自然问题链承接，不重复正文已有的体系说明；均指向 www.civitas.top 当前稳定文稿页。

本文网站文稿页

<https://www.civitas.top/documents/p03-strat-001/>

<https://www.civitas.top/go/p03-strat-001/>

01 文明跃迁十问

帮助第一次接触体系的读者先看见现实问题、核心判断和下一步阅读方向。

<https://www.civitas.top/documents/p01-entry-001/>

02 文明跃迁研究文稿汇编：首次阅读版

帮助第一次接触体系的读者先看见现实问题、核心判断和下一步阅读方向。

<https://www.civitas.top/documents/p01-entry-002/>

03 文明跃迁入口路径图

帮助第一次接触体系的读者先看见现实问题、核心判断和下一步阅读方向。

<https://www.civitas.top/documents/p01-entry-003/>

完整公开文库

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

公众阅读总入口

<https://www.civitas.top/documents/p01-entry-006/>