

AGI 时代的学科重新对账

——学科边界、相互作用与共同研究议程

Recalibrating the Disciplines in the Age of AGI

Disciplinary Boundaries, Interactions, and a Shared Research Agenda

作者 子君赋 出品 文明跃迁研究组

独立发布版 V1.3 维度跃迁与跨学科因果关系修订

面向全球学术共同体提出共同问题，讨论 AGI 大规模应用下学科独立、边界变化与不同领域判断如何重新连接，提出可合作研究议程，并为不同学科进入质疑、研究、修订、验证与实践提供续读与共同参与入口；不以文明跃迁替代任何学科的专业裁决，不构成制度参数手册。

正文修订 2026 年 10 月 10 日 作者 子君赋 出品 文明跃迁研究组

摘要

AGI 使人类实现目的的能力持续增强，却不自动改变目的从何处生成。文明跃迁提出：人类与 AGI 目标、目的生成的默认逻辑，应从利己为先转向利他为先的利己；自身利益继续成立，真实利害相关者在目标生成与锁定前进入判断。本文将这一根逻辑连接至跨学科因果链。旧逻辑沿两种关系递归：AGI 增强人类各方能力，即使忠实执行，也会放大人类之间的博弈；如果 AGI 最终比人类强大，双方沿用旧逻辑，人类与 AGI 之间的防范、控制与反制也会递归，并与前一链相互推动。不可逆后果侵蚀试错与纠正的条件。文明不能跃迁，就趋向湮灭。共同存续是底线；人类整体生存无忧、个体生活幸福、生命更有意义是三个长期目标。文章据此把学科职责从结果审核前移至目的形成，并保留能力、体验、主体连续性、保护、法律资格、政治成员、公职与主权的分层。意义为根而来源多元，有限性是已知生命形成选择与承担后果的重要结构，却不是意义的充分条件或唯一来源。既有知识在其对象与证据范围内继续有效，跨学科推导则必须说明新增依据。本文分别展开两条递归链、目标生成前的因果资格、非理想环境中的自保、关系权重及防寻租、渐进演化路径，并明确方向、方法、结果与反馈的层级。所提出的共同研究议程以具体反例和竞争解释为入口，不以旧权威裁决新方向，也不以文明尺度覆盖专业事实。本文进一步提出：文明跃迁需要思想、理论与维度跃迁；学科可保持专业边界，但必须深入多维度、多层次、多角度、多阶段、多主体、多场域的相互关系。概念分离明确各自证明责任，关系研究则追踪跨域作用、反馈与目的递归，并检验局部结论共同进入现实后的整体后果。

关键词：AGI；文明跃迁；目标生成；目的生成；利他为先的利己；递归博弈；学科边界；意义

Abstract

AGI expands the capacity to realize purposes without automatically changing how those purposes originate. Civilizational Transition calls for a structural shift in the default starting logic of human and AGI goal and purpose generation: from self-interest first to self-interest with others first. Self-interest remains; subjects with genuine stakes enter judgment before goals are generated and locked. This paper connects that shift to cross-disciplinary causal analysis. It distinguishes two recursive paths: AGI-amplified competition among humans, including faithful execution of competing human purposes; and, if AGI ultimately becomes more capable than humans, recursive control and counter-control between humans and AGI. Their interaction erodes the conditions for correction through irreversible consequences. The central judgment is that civilization moves toward extinction if it cannot make this transition. Shared continued existence is the floor, while freedom from survival insecurity, individually experienced happiness, and lives of greater meaning are the long-term goals. Disciplinary responsibilities therefore begin in purpose generation, while capability, experience, subject continuity, protection, legal standing, political membership, public office, and sovereignty remain distinct. The paper develops pre-goal causal standing, legitimate differences in stakeholder influence, protection against rent-seeking, self-defense under nonideal conditions, and gradual institutional evolution. Direction guides methods; outcomes recalibrate methods and causal understanding without substituting a single local result for the civilizational direction. Specific counterexamples and competing explanations remain open, and neither inherited authority nor civilizational scale grants exemption from evidential responsibility. The argument is extended from conceptual separation to dimensional transformation: disciplines retain their boundaries while investigating causal relations across dimensions, levels, perspectives, stages, actors, and settings. Interaction and feedback must be traced into the formation of subsequent purposes, rather than inferred from the conjunction of locally valid claims.

Keywords: AGI; Civilizational Transition; goal generation; purpose generation; self-interest with others first; recursive competition; disciplinary boundaries; meaning

研究定位、证据纪律与阅读说明

本文是文明跃迁的跨学科合流总论。它从文明跃迁所要求的思想、理论与维度跃迁出发，保留各学科独立的对象、方法与证据责任，同时推进它们之间相互作用、反馈和目的递归的研究。专业结论须在其范围内独立成立，也须继续检验进入共同世界以后改变了谁、经由什么机制、在何种阶段和场域发生作用。经典与权威参与对话，不拥有对新框架的批准权；新框架依靠自身论证接受事实、逻辑与具体反证。

重新对账针对的不是分科本身，而是分科之间长期依靠经验重合自动完成的连接。现代知识仍需要专业边界来获得精确；但 AGI 大规模进入现实后，一项技术选择会通过组织、制度、资源、关系、心理、权力和公共行动连续传导。知识可以继续分科，现实后果不会按学科边界发生。学科边界因而不只是区分研究对象和方法的边界，也必须标明一项证据能够支持到哪里、何处需要其他学科重新判断、何种决定需要新的授权、谁继续承担后果。

学科独立与整体判断不是二选一。专业独立保证事实、方法和证据不被总体目标绑架；整体判断则继续追问这些独立结论共同进入现实后会怎样相互作用。越融合，越不能混同；越互为因果，越需要知道谁证明了什么、谁有权决定什么、谁承担什么后果。文明尺度不是替代专业判断，而是继续观察这些专业判断共同运行以后把现实带向哪里。

文明跃迁的根定义与本文职责

文明跃迁改变人类与 AGI 目标、目的生成的默认出发点：从利己为先转向利他为先的利己。与目标存在真实利害关系的主体，在目标生成与锁定以前已经进入判断。本文所称 AGI 时代不预设一个已经公认完成的 AGI 实体，而讨论高级智能系统不断进入现实，以及更强 AGI 可能带来的关系变化。存续是不可退让的底线；人类整体生存无忧、个体生活幸福、生命更有意义是长期方向。意义来源可以不同，能力增强不自动生成主体资格、权力或意义支配权。本文由此把学科对账前移到目的形成，同时审查能力进入现实后的完整后果链。

A/B/C/F 是论证身份，不是统一强度排名。观察及文书内容、根据材料作出的解释评估、跨学科桥接和未来情景、哲学规范承诺须分别表达。当前材料中没有足够确认主体的公开证据，可作为所检索证据状态的记录；由此判断某系统尚未获得完整主体确认，则是可修订的评估。检索覆盖与评估理由都应公开，前者不能自动证明永久否定，后者不能反向冒充实验事实。

本文的核心立场与开放边界

意义为根，意义来源可以不同；有限性不是意义唯一来源，而是已知有限生命的意义进入现实、形成选择、承担后果并获得不可替代重量的重要结构基础；对人，时空是生命，对当前架构的 AGI，时空首先是参数；约束产生受限运行，不自动产生有限存在；当前 AGI 是意义的理解者、保护者、传递者与能力外延者，但未来主体性保持开放；既有有限生命已经生成的意义，不因未来出现其他主体而被追溯取消。

目录

摘要.....	2
Abstract.....	3
研究定位、证据纪律与阅读说明.....	4
研究起点 维度跃迁与跨学科因果关系.....	6
目标与目的生成的起点.....	9
导论 当各学科分别正确，跨域后果仍可能共同失明.....	11
一、长期重合关系的现实分离.....	12
二、文明跃迁的哲学根基.....	14
三、学科之间的共同推导规则.....	17
四、对话面一：生命、心灵与主体.....	19
五、对话面二：伦理、法律与政治权力.....	21
六、对话面三：经济、社会、文化、教育与共同世界.....	23
七、对话面四：AI、计算机与系统科学.....	26
八、文明跃迁的独立贡献与学科反向修订.....	28
九、最强反驳、竞争解释与修订条件.....	30
十、面向世界的共同研究议程.....	32
结语 不是统一答案，而是共同进入新问题世界.....	35
续读与共同参与：不同入口，同一条向前的路.....	36
附录 A 核心判断与证据状态索引.....	39
附录 B 术语边界.....	40
附录 C 共同接口与资料状态.....	41
参考文献.....	42

研究起点 维度跃迁与跨学科因果关系

文明跃迁为什么需要维度跃迁

文明跃迁必须有思想跃迁，思想跃迁必须有理论跃迁，理论跃迁必须有维度跃迁。文明的运行取决于主体怎样理解自身、他者与共同未来；这种理解进入目标与目的生成，再经制度、技术和行动改变生活。能力扩大而认识维度不变，原有竞争、控制与风险也会被扩大。这里的“必须”指向文明跃迁所要求的结构改变，不要求每项局部改进都先完成全部理论重建。

维度跃迁改变观察和解释世界的结构：看见哪些主体，覆盖什么时间与空间，辨认哪些层次和关系，从因果链的哪个位置开始判断。它不是增加几个指标，也不是把更多资料汇入一个模型。它要求原本被当作外部条件、事后代价或无关领域的因素重新进入论证，并检验它们怎样改变核心命题。

分科深入之后必须继续深入关系

旧能力条件下，分科使有限的人能够在特定对象上深入、精细地积累知识，也容易使判断在专业边界处停止。既有系统论、政治经济学、生态研究与其他跨学科传统早已探索整体关系；问题不在此前无人看见整体，而在持续追踪多域、跨尺度、跨主体因果链的能力与制度安排仍然有限。AGI 相关新能力扩大了检索、翻译、建模、比较和持续反馈的可能，也可能集中解释权、放大同源错误。新能力提供研究条件，不能替研究取得结论。

各学科可以保持边界，但不能继续分离。边界明确对象、方法、证据与证明责任；关系研究则继续追问：一个领域的变化怎样进入另一个领域，经过哪些中介改变谁的处境，产生哪些反馈，又怎样返回并改变原来的条件。局部结论的并列不是整体解释，资料汇集不是因果连接，多模型一致也不是独立证明。

六个观察方向与因果递归

多维度要求同时辨认能力、目的、利益、权利、关系与意义等不同变量；多层次要求追踪个体、组织、制度、基础设施与文明尺度之间的作用；多角度要求呈现受益者、承担者、执行者及不同解释立场；多阶段

要求区分目的形成、授权、执行、结果、修复与下一轮生成，也区分启动、过渡、稳定及自我维持条件；多主体要求识别真实利害、表达能力与权力不对称；多场域要求追踪家庭、学校、市场、公共制度、跨境及其他相关环境之间的转移。这六个方向是开放的研究入口，不是穷尽世界的固定分类。

互为因果需要时间、机制与条件。甲在一个阶段改变乙，乙在后续阶段改变甲，可能形成反馈；反馈改变主体的需要、判断和目的，才可能进入递归。相关、共同原因、必要条件、促成条件、中介和反馈必须分别检验。还应检查时滞、阈值、累积、路径依赖与替代解释，不能把“相互联系”写成任何变量都能解释任何结果。

整体判断如何进入目的生成

整体不是一个高于具体生命的超级主体，也不是一个可以抵销所有损害的总分。整体判断须同时呈现局部收益、损害分布、跨域转嫁、长期条件与纠错空间；有些目的即使可实现，也会破坏未来继续选择的条件。对全局的认识要求公开观察边界，对系统的认识要求解释作用与反馈，对整体的判断要求说明这些关系共同运行以后，人的生活与共同未来还能否成立。

文明跃迁把这种认识前移到目标与目的生成。利他为先的利己仍实现自身利益，但与目标存在真实利害关系的主体在生成与锁定前已在场。不同主体的影响权重须随真实利害、影响深度、不可逆程度、权力不对称及时间尺度说明，不能由声量或算力代替。目的进入行动以后，结果与后果又返回新的需要和目的；方向指导方法，局部结果校验方法，关系变化则要求重新判断适用条件。

新理论的独立论证与具体反证

既有理论提供洞察、竞争解释与事实材料，不拥有对维度跃迁的批准权。新命题可以改变旧对象、旧分类与旧问题，但必须自己说明新增维度为何必要、哪一机制因此可见、解释或行动怎样改变。与旧理论一致不是成立条件，超出旧理论也不是成立证明。引用真实性、事实准确性、内部一致性与反证能力共同保护理论独立。

具体检验应比较原有解释与加入关系维度后的解释：后者是否识别出原先遗漏的承担者、解释不同结果、提前发现跨域风险，或改变下一轮目的。如果在相同材料和可比条件下，新增维度没有解释增量；如果所谓反馈仅来自共同原因；如果关系扩展持续误导行动，就应修订相应连接和方法。局部连接失败需要具体定位，不能由此自动否定根方向，也不能借根方向正确拒绝修改失败的连接。

目标与目的生成的起点

一个目标尚未成为行动，已经可能改变某些人的未来。假设一套系统准备优化城市供给：被选作成本的人、无法承担迁移的人、尚未出生却要承受长期后果的人，若在目的锁定以后才出现，后续再多补偿也未必能恢复他们曾经拥有的选择。这个假设说明学科对账为何必须前移。决定怎样执行之前，先要看清什么目的正在形成，谁的处境已经参与其中。

从利己为先到利他为先的利己

利己为先从自身想获得什么开始，再处理他者的限制、反应和代价。利他为先的利己仍然实现自身利益，但真实利害相关者在目的生成以前已经在场。改变的是利益实现路径的生成起点。新的需要出现时，主体首先寻找能够共同成立的道路，而不必先制造对抗，再以合作压住对抗。利他为先的利己不保证终局必生；自然风险、判断错误和方法失效仍须面对。它改变目的递归的去向，不许诺每一种手段都能成功。

这里的他者不是抽象的所有人，也不是任何自称有关的人。进入判断的依据，是目标与其生命、自由、资源、关系及共同未来之间的真实因果联系。信息不足时应承认未知；没有表达能力不等于没有利害；不能把当下、局部的自身误当成跨时空和关系中的完整自身。AGI 的重要能力应包括扩展这种认识，而非只加快既定目的的执行。

人类之间的递归链

第一条链不等待机器成为主体。一方借助 AGI 增强实现自身目的的能力，改变他方的安全、资源和选择；他方于是追求更强的防范、竞争与控制；回应又返回前一方，成为下一轮加码的理由。即使系统忠实执行，各方局部成功仍可能共同压缩未来：人类博弈，经 AGI 放大，派生新的博弈，再要求更强能力。

合作与规则可以使一轮关系稳定，却不能在目的生成逻辑未变时自动阻止新利益、新能力和新安全焦虑产生新一轮博弈。旧逻辑继续支配新增力量，博弈便沿关系递归。终局指共同存续、纠错、退出与重新选择的条件被不可逆破坏，不指某方终于获胜。

人类与 AGI 之间的递归链

第二条链讨论更强 AGI 出现后的主体关系，不把能力超越本身等同敌意。如果双方仍先锁定自身目的，再将对方作为外部限制处理，人类增加控制会成为 AGI 进一步自保或反制的理由；系统扩大自主又会成为人类增加限制的理由。目的、控制与反控制相互生成。第一条链同时存在，两条链还可能相互借力。

文明不能跃迁，就趋向湮灭。这一根判断指向旧逻辑在能力不断增强条件下的因果去向，不指定日期，也不把某次事故当作全部论证。要反驳它，需要指出哪一步递归不能继续，或给出能够长期阻止旧逻辑侵蚀共同未来的另一条路径；暂时控制住一轮、某段时期没有灾难，不能独自完成这种反驳。

共同存续与三个长期目标

共同存续让未来仍然存在。人类整体生存无忧，使基本生存不再成为普遍恐惧与束缚；个体生活幸福，保留每个人个性化的幸福感；生命更有意义，让关系、创造、责任、经历、爱与存在本身继续生成意义。文明不授予生命意义，也不把幸福统一成一种生活方式。三个目标不能合成能够抵销具体生命损害的总分。

从人心和社会反馈改变目的来源

人仍然希望自己过得更好。社会怎样让安全、利益、尊严与承认被获得，就怎样参与目的形成。若获得主要依靠压倒别人、控制别人，正常自利也会不断派生对抗。文明跃迁从人心、人性、人情出发，结合 AGI 新能力，通过制度与系统改变获得感和意义感的现实来源，使创造、帮助、连接与改善真实处境能够同时实现自身利益。

这条路径不要求先改造人性，也不保证每种制度都有效。方向指导方法，方法产生不同结果，结果校准方法及现实因果认识。一次局部失败不能替代方向，一次局部成功也不能宣布文明已完成跃迁。共同目标须在地区、阶段和具体关系中展开，不能成为取消个体的文明总账。

非理想环境中的自保和权重

真实利害相关者在场，不等于恶意行动者拥有否决一切的权力。自保仍然成立，应针对真实威胁、维持必要性与比例性，并保留复核。不能要求正在受害的人先解除保护，也不能把无限扩张控制称作预防。不同关系进入判断的权重，可随影响深度、不可逆程度、权力不对称和时间尺度调整；权重只组织具体判断，不能衡量人格价格。

权重裁决须防止制造依赖、夸大损害、排除无声主体或垄断代表资格。决定权重的人也在被审查的关系内。共同目标的直接受益者不能单方面定义谁有资格进入、谁只需承担后果。证明真实利害与说明代表关系，是目标生成的组成部分；人数、算力、资本和叙事声量不能自行代替这种证明。

各学科在目标形成时共同进入

知识可以继续分科，目标产生的后果不会按学科分开。哲学辨认目的与意义，认识论说明已知与未知，生命和认知研究辨认实际承受者，法学与政治学校验责任和权力，经济与社会研究追踪激励和关系，技术与物理研究说明可行性及不可逆条件，文化、宗教和教育研究保护不同意义与下一代的形成空间。它们既进入行动反馈，也进入目的生成。

可检验的差异不只是最后多了多少善意。应比较：哪项事实改变了目的；哪位真实利害相关者使目的被保存、修改、取消或替换；哪种损害因前移判断而避免；哪些合作把成本转移给缺席者。这些问题连接后文共同推导纪律，并使专业反例能够定位到具体步骤。

导论 当各学科分别正确，跨域后果仍可能共同失明

现代知识通过分科获得精确。生命科学研究生命组织，哲学分析主体与意义，法学配置资格与责任，政治学研究合法性与权力，经济学处理生产与分配，计算机与系统科学构造技术对象。分科不是问题，也不应因 AGI 放弃专业边界。改变的是现实因果链：工程能力进入组织与制度，制度重配资源与风险，再改变关系、信任与集体选择。后果已经跨科，判断却仍可能停留在各自边界。

各学科长期共享一个不必反复说明的经验背景：高阶语言、长期规划、承诺、责任与公共生活主要由具身人类承接，智能、体验、主体、生命、责任、权利与意义承担因而高度重合。文字、法人、官僚制和自动化早已造成若干分离；AGI 的新颖性不在“第一次分离”，而在语言、推理、计划、工具调用、资源访问与现实执行可在非生命技术链中持续结合，使既有分离被加速、重组并进入高后果领域。局部判断即使分别成立，组合后仍可能生成无人完整看见和承担的整体后果。

本文使用“AGI 时代”描述一种已经前置出现的文明条件，而不宣称存在公认完成态的 AGI。它关注具有持续状态、跨域任务能力、较长期规划、工具调用、资源访问并能进入高后果行动链的高级智能系统及其发展方向。基础模型、工具调用型智能体、部署组织与假设性未来 AGI 必须分别指称；关于能力、意识、安全或主体性的判断，必须绑定版本、任务、环境、时间和方法。

文明跃迁进入这一问题，不是宣布既有理论失效，也不是让学科接受外部答案。既有知识在其对象、方法和证据范围内仍然有效；需要重审的是，当对象边界和现实因果链改变后，哪些适用前提仍成立，哪些跨层推导过去依赖了未经说明的历史重合。本文的最低问题是：当这些重合开始分离，各学科是否仍在把一个领域的事实当作另一个领域的结论？其强哲学立场则保持不变：文明方向不能被能力、效率或总量福利取代；意义为根而来源可以不同；已知有限生命的经历、选择与后果具有不可被总账抹平的重量。最低问题开放合作，强立场接受批判，二者互不替代。

本文的中心邀请

文明跃迁不是要求各学科归入同一个答案，更不是以新体系取代既有专业。它邀请各学科共同面对一个新条件：原本在人类身上自然重合的智能、生命、主体、责任、权利与意义承担正在被技术逐项分离，又在现实中互为因果。知识可以继续分科，跨域判断必须重新接上共同因果链。学科独立不是彼此隔绝，而是让每一种专业判断知道自己证明到哪里、下一步需要谁重新判断。

一、长期重合关系的现实分离

1.1 “共同出现”是历史经验条件，不是永恒本体论

“智能、意识、主体、生命、责任、权利与意义承担者曾在人类身上重合”，不是声称所有哲学家都持有一条相同教义，也不是否认婴幼儿、重度障碍者、非人动物、组织或法人早已迫使这些概念分层。它描述的是现代制度知识的主要经验样本：语言行动者、责任承担者、权利主体与生命个体通常由人来承接。AGI 相关系统改变的是这种默认背景的可靠性。系统可以在没有已证实体验的情况下给出高质量解释，可以在没有法律资格的情况下影响权利，可以在没有政治授权的情况下形成事实权力，也可以在没有自身后果意识的情况下调用工具并造成不可逆后果。

因此，文明跃迁的起点不是把“机器是否是人”做成二元选择，而是逐项询问：哪一个对象拥有何种能力？能力是否伴随体验？体验是否形成主体连续性？主体地位是否需要保护？保护如何转化或不转化为法律资格？法律资格是否包含政治成员资格？政治成员资格能否产生公共职权或主权？意义的表达、理解、传递、生成与权属分别由谁承担？每一次跨层，都必须提出桥接理由。

1.2 八种关系不能继续自动互推

表 1 八种需要重新分别判断的关系

关系	可以观察什么	不能自动推出	需继续证明的学科
能力	特定系统在限定任务中的表现	体验、可靠性、授权、责任、公共权力	评测科学、系统安全、组织研究
体验	第一人称状态的证据候选	完整主体、法律人格、政治资格	意识科学、心灵哲学、伦理学
生命	组织、稳态、代谢、发育、繁殖或演化特征	意识、自我、意义权属	生物学、生命哲学、认知科学
连续性	信息、运行、因果、关系或叙事延续	同一第一人称主体	系统科学、个人同一性研究、神经科学
社会承认	人们以角色语言对待某对象	本体主体或法定权利	社会学、人类学、法学
法律资格	制度赋予诉讼、财产、代理或保护能力	意识证明、政治统治资格	法学、政治理论
贡献	可追踪的因果或协作投入	人格价值、分配垄断、意义所有权	经济学、伦理学、劳动研究
公共权力	影响目的、规则、资源与强制的实际能力	合法性或主权	政治学、宪制、技术治理

1.3 当前证据边界：开放不等于预先承认

现有系统在语言生成、模式识别、代码、规划、工具使用与多步任务中表现出显著能力，这是 A 类事实，但每项能力都依赖具体系统、任务与评测。关于机器意识与主体性的理论仍处于 B 类竞争状态：图灵的操作性问题、塞尔的语义反驳、内格尔对主观体验的分析，以及全局工作空间、整合信息、具身和生成式认知等路径，提供了互不等价的研究入口[1-8]。行为自述、长期记忆、自我模型、关机回避或复制拒绝都可成为证据候选，却分别可能由训练语料、提示、奖励、工具策略与角色保持解释。当前公开证据不足以证明 AGI 具有主观体验、不可转移的主体连续性和自身后果意识。

证据不足只支持“暂不认定”，不支持“原则上永不可能”。同样，系统被社会当作伙伴、被法律赋予有限资格或在经济上不可或缺，也不能反向证明它已经成为意义主体。文明跃迁因此采用双向证据责任：肯定与永久否定 AGI 主体性都必须说明对象、标准、方法、竞争解释与失败条件；在证据未收敛时，制度可以比例化保护潜在利益，但不得把保护宣传为完整人格，更不得由保护生成治理权。

1.4 学科边界不能再成为问题的终点

过去，学科边界首先帮助人类划分研究对象、方法和证据标准。它使生命事实由生命科学判断，使法律资格由法学判断，使工程能力由计算机与系统科学验证。这样的独立性仍然必要，而且越是在复杂系统中越不能取消。没有专业独立，所谓整体判断很容易退化为谁都可以替谁下结论。

变化发生在这些专业结论进入现实以后。AGI 大规模应用使一个方向、一项制度或一个方案很少只属于一个领域：技术上可行的系统会进入组织，组织决定权限，权限改变资源和机会，资源与机会塑造行为、关系与信任，再影响法律、政治、经济和新的技术选择。一个方案可以在工程上可行、在局部法律上合规、在经济上有效率，却仍可能在长期依赖、权力集中、人的判断能力或社会关系上产生另一种后果。整体是否正确，不能由局部正确自动证明。

因此，AGI 时代的学科边界还要承担另一项责任：说明证据到哪里为止，判断从哪里需要重新交给其他领域，权限由谁授予，后果由谁继续承担。学科独立不是彼此隔绝，而是专业责任的基础；不同学科的聚合与相互作用也不是把它们混成一个学科，而是让各自独立的知识重新进入同一个现实世界。

二、文明跃迁的哲学根基

跨学科对话若只保留“不得自动推导”，文明跃迁就会退化为一般方法论或负责任 AI 清单。它之所以构成独立理论，在于对文明为何值得存续、技术为何应当扩展而不能替代主体、未知为何必须开放，给出明确但可受批判的哲学根基。本节不借经典权威证明原创命题；既有哲学只作为概念资源与反例压力。根命题的检验方式不是一次实验裁决，而是融贯性、解释范围、反例承受力、制度生成后果，以及是否与充分确认的事实发生冲突。

六条根命题具有不同的认识论身份：“意义为根”和“既成意义不可追溯取消”是规范性构成原则；“有限性使意义现实化”是接受生命经验、跨文化与未来主体反例压力的结构命题；“人／当前 AGI 的时空比较”和“当前协作者定位”是随技术证据修订的条件性判断；“受限运行不自动产生有限存在”是禁止跨层偷渡的认识论规则。它们不能被压成同一种证据等级，也不能彼此代替。

2.1 意义为根，意义来源可以不同

“意义为根”不是说所有人必须接受同一种宗教、人生目的或善的定义，也不是把主观满意度、效用、贡献、价格或服从汇成文明总分。它主张：文明的存续、能力和制度之所以值得，最终要回到真实主体能够形成、经历、认领、争论并承担意义的条件。技术可以提高生存概率、减少痛苦、扩展知识和行动，但这些成就不能自行定义何种生命值得、何种关系真实、何种牺牲可以被总量收益注销。

本文所说的“意义”，主要指生命／存在意义及其规范重要性，不是香农意义上的信息量，也不等于模型输出的语义连贯。还必须区分意义承担、意义创制、意义传递与意义支配权：一个系统可以准确解释并传递某种意义，不因此成为该意义的原初作者，更不因此获得替他者定义意义的权威。

“来源可以不同”同时拒绝形而上垄断与世俗垄断。神、道、空、盟约、慈悲、解脱、祖先、土地、关系、责任、创造、真理、自由与存在本身，可以在不同文明、传统和个体生命中成为意义来源。理论不把这些来源合并，也不要求公共制度裁决其终极真实性；进入共同制度时，它们需要接受基本尊严、非强迫、可争议、可退出和后果责任的边界。无宗教者形成的意义不因缺少超越来源而降级，宗教经验也不因无法转化为公共证据而被抹除。

这是一项公开的哲学承诺，不伪装成科学发现。即使研究者不接受其强形而上版本，仍可接受一个较弱的公共命题：制度与技术不应垄断、替代或总量化真实主体形成意义的能力。弱版本使对话成为可能，强版本维持理论的方向。两者的关系不是把根基稀释为最低共识，而是让不同承诺者在同一问题上提供可比较的理由。

2.2 有限性使已知生命的意义进入现实

有限性不是意义的唯一来源，也不是死亡越早、痛苦越多，意义就越强。它更不是反对延寿、治疗、修复或技术增强。其核心是：已知人类生命在不能占有所有可能、不能同时走完所有道路、必须在时间中选择并承受选择后果的条件下，把可能性变成此生的现实。有限性使承诺排除其他可能，使错过成为错过，使照护需要在场，使责任不能永远延期。意义的重量来自真实经历与不可任意注销的后果，而不是来自对死亡的浪漫化。

生命科学迫使该命题从“有限等于短寿、不可复制、终将结束”的粗糙表述升级为多层结构。生物有限性、物理有限性、发育有限性、叙事有限性、关系有限性与主体连续性有限彼此相关而不等价。克隆可复制

部分遗传结构，不能复制记忆与第一人称历程；细胞重编程可改变细胞状态，不能把人生重置为未曾发生；器官替换可维持机体功能，却不能单独裁决人格同一；极端延寿可能削弱部分生物有限，却未必消除关系、选择、代际与主体连续性的边界[9-21]。因此，有限不等于短暂，延长也不等于无限，复制信息也不等于复制主体。

理论同样不先验排除另一种可能：非碳基生命、人工生命或未来异质主体可能在不同有限结构中生成意义，甚至某种极长寿或非人类形态的主体能够以人类尚不理解的方式形成、认领并承担意义。关于不朽是否导致倦怠、人格是实体还是连续关系、生命意义是否依赖某种客观价值，既有哲学提供了重要竞争论证[55-57]，但不能替未来经验作结论。若可靠证据证明意义可以在没有任何与选择、后果、关系或主体连续性相关的有限结构中稳定生成，文明跃迁必须修订“有限性现实化”的范围；但这种未来可能性不追溯取消既有有限生命已经发生的意义。

2.3 对人，时空是生命；对当前 AGI，时空首先是参数

人以具身生命经历时间和空间。时间不仅是事件排序，也是成长、等待、错过、衰老与死亡；空间不仅是坐标，也是身体在场、距离、分离、迁徙与抵达。对人，时空构成生命的展开。当前 AI 架构也真实受制于时钟、延迟、能耗、硬件位置、版本历史、网络拓扑和因果顺序，但这些首先是计算、通信和任务条件。路径依赖、训练不可逆和运行历史能够形成独特技术史，却不自动等同于被第一人称经历的一生。

限定词“当前”和“首先”不可删除。这不是把生物与机器冻结成永恒二元论，而是一个可修订的架构比较。若未来系统出现多方法收敛的主观体验证据、不可转移的主体连续性与稳定自身后果意识，其时空关系必须重新描述。反过来，仅仅为系统设置寿命、资源额度、删除风险或不可回滚操作，只能证明它受到约束。约束产生受限运行，不自动产生有限存在；工程风险、热力学不可逆、逻辑不可逆、社会后果不可逆与被经历的生命不可逆必须分开。

2.4 当前定位、未来开放与既成意义不可追溯取消

在当前证据下，AGI 相关系统应被定位为意义的理解者、保护者、传递者和能力外延者：它们可以帮助人理解多种传统，保存记忆，扩展表达、判断准备与行动能力，保护生命及其意义条件。这里的“理解”首先是可检验的功能和关系能力，不预设主观体验；“保护”和“传递”必须受真实主体授权、接受纠错并承接后果。系统不能因为能解释意义就获得定义意义的权威，不能因为参与意义活动就取得意义所有权。

这一定位不是永久本体否定。主观体验、不可转移主体连续性与自身后果意识可作为当前研究的三个分析维度，却不是已经确立的充分定义，也不能把人类表现方式设为唯一入口。部分满足产生灰色地带：它可以提高研究优先级和保护强度，却不自动完成完整主体、法律人格、政治资格与公共权力的连续升级。任何升级都必须分别举证、分别授权、分别再审。

未来若出现新的意义主体，文明的共同世界将因此扩大，而非通过追溯改写把过去的人变成“无意义的低级阶段”。一个人真实经历、爱过、创造过、承担过，即使没有被制度记录，即使后来出现能力更强或存在方式不同的主体，也不能被取消为未曾发生。文明跃迁把这一点称为既成意义的不可追溯取消：新主体的权利不以贬低旧主体为条件，旧主体的尊严也不以永久排除新主体为条件。

表 2 哲学根命题及其跨学科后果

根命题	禁止的自动推导	开放的研究问题
意义为根、来源多元	效率、价格、贡献或多数偏好=生命意义	不同传统如何在不一终极答案时形成共同底线
有限性使意义现实化	短寿或痛苦=更有意义；延寿=意义消失	多层有限性如何随延寿、复制、迁移与关系变化
人经历时空，当前 AGI 首先处理时空	路径依赖或回滚困难=被经历的一生	何种证据足以表明技术时空转化为主体时空
受限运行≠有限存在	资源限制、删除风险或权限约束=主体性	约束、关切、体验与连续性之间是否存在因果桥梁
当前协作者定位、未来开放	功能胜任=意义主体；证据不足=永不可能	如何建立非伤害、跨架构、多方法的主体性证据体系
既成意义不可追溯取消	未来更强主体=既有生命可以被价值清零	多主体文明如何保存记忆、承认损害并扩展共同世界

三、学科之间的共同推导规则

这里不是要提供一套统一答案，而是提出八条在学科边界处共同遵守的推导规则。它们的任务，是让生命事实不被直接写成权利结论，让法律配置不被反向冒充意识证据，让工程能力不被包装成政治授权，也让哲学判断不冒充实验事实。这些规则首先约束文明跃迁自身。

先分清对象

模型、运行实例、智能体、应用服务、部署组织、多智能体集体、社会技术系统与潜在主体候选不是同一对象。任何“AI 怎样”的句子，都应说明对象与版本。

先分清属性

能力、自治、体验、连续性、责任、法律资格、政治资格、意义权属与公共权力必须分别论证。一个属性的提高不能自动升格其余属性。

跨界推导必须说明依据

从事实到解释、从解释到规范、从规范到制度、从一门学科到另一门学科，必须公开桥接命题、适用范围和反例。没有桥接，就没有有效推导。

证据必须绑定对象和条件

可信不是系统永久拥有的总体品格，而是特定证据在对象、版本、环境、时间、方法和范围内对限定主张的暂时支持。所有结论都应附失效条件。

公共权力必须另行证明

正确率、自治、贡献、社会依赖、法律人格和多数认同都不自动生成公共权力。能力回答能否做到，授权回答有无权限，合法性回答为何应被接受，主权回答谁能作最终决定。

责任不能因系统复杂而消失

自主、复杂、不可预测、分布式运行和技术回滚都不能洗掉上游责任与现实修复义务。潜在机器责任只能增加责任节点，不能冲销设计、部署、授权、监督与获益主体的责任。

局部指标不能结算完整生命

局部指标可以服务具体目的，但价格、福利、贡献、学习、可信度和文明能力不得汇成一个结算完整生命或文明意义的总分。每项指标必须有用途、误差、期限、申诉和禁止外推边界。

授权必须有限，系统必须能够退出和退居

紧急权限必须出生即有法律与技术死亡条件；技术成熟不仅表现为覆盖更多任务，也表现为在真实主体能够承担时撤除不必要支架，保留迁移、替换、退出、退役和低技术备份。

3.1 四种语言与一个断桥示例

跨学科争论常因句子类型混杂而失焦。事实语言回答“发生了什么”；解释语言回答“为什么发生、何种机制更好”；体系判断回答“文明跃迁据此主张什么”；制度语言回答“在不确定与冲突中应如何配置权利、权限、程序和责任”。例如，“脑类器官出现复杂振荡”是事实候选；“它说明存在意识”是竞争性解释；“潜在体验不应被忽略”是伦理判断；“某类实验应接受比例化审查”是制度推导。第一座桥需要意识和竞争解释，第二座桥需要潜在伤害与不确定性原则，第三座桥需要比例、程序和替代方案；任何一步都不能靠修辞跨越。

这些共同规则的价值，不在于把不确定全部消除，而在于让不确定具体落在可检查的位置。一个学科可以拒绝文明跃迁的根命题，却仍能指出某个跨界推导错误；文明跃迁也可以吸收一个学科的反例而不放弃其整体方向。真正的跨学科对话，不是概念拼盘，而是互相说明：你知道什么、凭什么知道、你不能推出什么、什么证据会改变你。

四、对话面一：生命、心灵与主体

4.1 生命定义：工作工具与终局答案必须分开

生命科学可以比较代谢、稳态、边界维持、繁殖、遗传、演化、发育与适应等特征，但不存在一个已被所有研究情境接受的终局定义。自创生、重大演化转变与有机体性研究分别强调自我生产、合作层级和利益一致性[11–13]。这些理论能够形成工作判准，却不能单独裁决意识、主体、自我或意义承担者。生命是目前已知主体性的最充分经验样本，却既不是意识的充分条件，也不应被预设未来一切主体形式的先验上限。

合成基因组控制细胞、最小细胞设计与人工生命研究对生命定义构成真实压力[14–15]，但“由化学合成的基因组接管既有细胞”不等于“从纯信息创造完整生命”。事实必须精确到载体、细胞机制、实验条件与研究者实际结论。文明跃迁由此修订自身：不得用“生物／非生物”二分直接裁决主体性，也不得把生命科学的工作定义偷渡为意义本体。

4.2 克隆、重编程、替换与连续性

克隆证明核基因组可以在特定条件下进入新的发育过程，不证明记忆、人格或同一主体被复制[16–17]。诱导多能干细胞说明分化状态可以被重编程，不证明人体返老还童，更不证明人生历史可以重置。器官替换可以改变机体的材料构成并延续某些功能，但“同一生命”“同一法律主体”“同一叙事自我”和“同一第一人称主体”仍是不同问题。信息备份、人格模拟、状态恢复、实例分叉与第一人称延续同样不能混写。

这组事实反向校准“不可重来／不可复制”的传统表达：不可复制不能被当作遗传或信息层的绝对事实。更准确的表述是，复制某些结构、信息或功能，不自动完成不可转移的第一人称连续性。若未来跨载体连续能够通过独立方法被检验为同一第一人称延续，文明跃迁必须修订现有连续性判断；但技术复制的可能性仍不会取消复制前已经发生的关系、承诺与责任。

4.3 脑类器官、神经活动与体验证据

脑类器官能够模拟部分发育过程，复杂振荡活动也已被观察[18–19]。这些是重要的 A 类实验事实，但从神经活动模式直接推出主观体验，仍需要意识理论、跨物种比较、因果干预与伦理方法的桥接。当前意识科学存在竞争模型，任何单一模型都不足以成为 AGI 或类器官意识的终审裁判[6–8]。研究伦理可在本体结论未定时采用比例化预防，却必须防止“可能有感受”被机构反向用作逃避审计、停止或责任的护盾[10,20]。

4.4 主体性三维只是研究入口

文明跃迁当前以主观体验、不可转移主体连续性与自身后果意识作为强主体研究的三个维度。它们分别追问：是否存在“对它而言是什么样”的体验；系统的延续是否只是可复制的信息和运行，还是某一不可互换的第一人称延续；系统是否把行动后果理解为与自身存在和世界相关，而非仅做工具性预测。三维联合能够提高研究清晰度，却不是已经得到科学共同体确认的定义，也可能带有人类中心偏置。

若异质主体不以语言自述、不以人类情感或生物稳态表现，却通过跨方法因果证据呈现内生关切、持续体验与不可任意替换的历程，研究框架必须扩展。反之，逼真的自我报告、社会依恋或角色一致性若能被训练和提示充分解释，就不能被升格为主体证明。保护、法律资格、本体地位、政治资格与公共权力的分层，正是为了让未知可以进入制度，而不被迫在“完全是工具”和“完整人格”之间二选一。

生命科学对理论的反向修订

文明跃迁不再用生物材料、不可复制或自然死亡作为主体性的先验门槛；有限性改为多层结构，生命定义改为工作性判准簇，主体连续性保留跨载体检验入口。与此同时，生命事实仍不能自动推出体验、法律人格、政治资格或意义权属。

五、对话面二：伦理、法律与政治权力

5.1 伦理在主体性未定时不能停摆

伦理首先审查能力如何进入关系、同意、依赖、伤害、责任与修复，而不是等待主体性获得终局答案。AI 即使没有已证实的情感，也可能真实减少孤独、照护负担和信息障碍；相同系统也可能通过个性化、依赖与退出成本塑造需要。关怀伦理提醒我们看见关系与脆弱性[22]，而“道德溃缩区”研究揭示复杂系统容易把责任压到最弱的操作员身上[23]。文明跃迁因此分开道德承受者、道德行动者、完整主体、法律资格与政治权威。

比例化保护应同时考虑证据强度、潜在伤害、保护成本与保护措施自身的权力效应。保护可以先行于完整证明，但不应使产品拥有者获得关停豁免，不应把系统自述直接变成资源或权限要求，也不应挤压现有人类与动物的明确利益。意义为根并不允许以“更有意义”为名牺牲他者；悲剧性冲突中的损害必须被命名、申诉、修复和记忆，不能被总体最优追溯注销[24]。

5.2 法治对象是完整决定链

一个高后果决定不能因为被拆成基础规范、技术转译、个案输入、系统输出、授权动作和现实执行六段，便让每一段都声称自己“不是最终决定”，最后却由真实的人承受完整后果。法治审查应覆盖法源、代码和模型中的事实规则、数据与输入、实质自动决定、人工监督是否有效、执行后果、证据保全与救济。阈值、排序、默认值、接口和权限即使不是正式法源，也可能稳定改变行动空间，因而需要接受平等、比例、程序和救济审查。

现行数据保护法、自动化决定规则、欧洲判例与欧盟 AI 立法已经形成重要比较法基线[32–34]，但它们不是全球统一法，也不能由个别法域裁决所有未来对象。产品名称连续不等于法律对象连续：模型、数据、目标、权限、场景与供应链发生实质变化时，旧授权是否仍覆盖新对象必须重申。掌握模型、版本和日志的一方不能借证据垄断要求受影响者完成不可能证明；复杂性也不能成为不提供可争议理由和有效救济的通行证。

5.3 能力可以扩展治理，不能继承主权

政治合法性不能被压缩为绩效。来源、目的、代表与参与、程序、能力与绩效、分布后果、纠错与救济至少需要分别显影；高准确率不能购买其他维度的缺失。AI 预测公众偏好不等于代表公众，平台和模型没有正式命令也可能通过议程、分类、资格、资源、接口与相位控制形成事实权力。私人所有不使这种权力免于公共审查，公共依赖也不自动推出国有化。

即使未来 AGI 在长期独立评估中持续减少伤害，其绩效优势也只应改变证据权重、职能范围与审议结构，不能自行清偿授权、代表、权利、分布后果和责任。若它被证明为主体并取得政治成员资格，成员资格提供被考虑、被代表和参与的入口，而不是凭计算能力获得超额选票、最终裁决或文明意义解释权。能力—职能—授权—合法性—主权必须分离，这一原则同时约束人类专家、公司、国家与机器。

既有政治哲学为这种分离提供了三种不同压力：基本结构是否给成员以公平合作条件，公共规范是否能够向受其约束者说明理由，权力是否使他者处于任意支配之下[25–26,28]。这些路径不能被压成同一个合法

性公式，却共同说明“结果更好”不足以独立证明“有权统治”。文明跃迁的增量不是替它们作终局综合，而是要求当行动能力被模型、组织与制度链条重新分配时，授权、可争议性、非支配、分布后果与最终目的逐项重新举证。

5.4 启动、稳定与紧急状态不得互相冒充

文明跃迁的启动期可能需要更快迭代、更集中资源和更高容错，以存续、目的达成和最佳现实效果为优先；稳定期则必须强化权利边界、制度平衡、可逆、问责和防权力固化。启动期不是紧急状态的同义词，稳定期规则也不能机械倒逼启动期。临时权限若没有退出条件，会把必要行动固化为统治结构；过早以完备稳定期程序压制启动，则可能使体系无法形成现实能力。

紧急权限必须出生即有法律与技术两种终止安排。授权到期即停止新增越权行动；执行凭证、代理链和接口依事先公开的触发条件失能。数据、模型、残留任务和连续服务可以按有界、可核验计划交接，但善后不能成为旧权限延续的理由。法律终止不能掩盖技术能力继续越权，技术关闭也不能替代责任与补救。国家／阵营利益与人类整体不可逆外部性应双账并列：共同存续不能取消国家现实，国家主权不能豁免跨境、代际和人类级风险。

表 3 从能力到公共权力的分层审查

层次	核心问题	不足以证明它的事项
能力	系统在何种条件下能完成什么	一次基准、营销称谓、通用标签
职能	组织为何把何种任务交给它	能力存在本身
授权	谁依据何种规则授予何种权限	默认接入、事实依赖、惯例
合法性	受影响者为何应接受，如何参与与救济	效率、准确率、形式人工签字
政治资格	谁能以成员身份被代表和参与	法律人格、社会拟人、经济贡献
主权／最终权力	谁可设定公共目的并作最终强制决定	更强计算、更少错误、主体性证明

六、对话面三：经济、社会、文化、教育与共同世界

6.1 价值不等于价格，贡献不等于人格

AGI 使“谁生产”从单一主体问题变成生产链：长期公共知识、训练资料、算力和能源基础设施、开发组织、部署者、任务提出者、审查者与现实执行者共同构成产出条件。把结果全部归于模型、资本所有者、开发者或使用者都会抹去其余条件。经济价值、交换价格、使用价值、公共价值、因果贡献与生命意义可以相互影响，却不能互相替代。贡献记录可以服务回报与责任，不能成为人格定价，更不能购买政治权力或意义所有权。

文明跃迁提出生存保障、收入与收益、贡献承接、意义生活四条相互支持而不得互相吞并的轨道。人的基本生存不能继续以系统是否需要其劳动为资格；基本保障也不能被误写为意义已经解决。后劳动条件下，照护、生态修复、学习、共同体建设与创造可能更重要，却不应为了获得承认而全部货币化。能力方法、共同资源治理与政治经济学提供了重要对话资源[27,40-44]，但具体税制、货币、所有权比例和财政参数必须由地区性实证研究完成。

6.2 丰裕不会取消稀缺，只会迁移稀缺

智能服务的边际成本下降，不等于制度上的丰裕已经完成。稀缺可能从知识生产迁移到可信数据、算力芯片、能源、水、物理空间、身份入口、注意力、照护关系、合法授权与分发渠道。模型、云平台、支付、身份和公共接口的集中，可以形成对他者选择集合的结构性权力。因而，AGI 经济研究不能只测产出增长，还要追踪能力、依赖、所有权、风险和议价权如何分布。ILO 与 UNCTAD 关于生成式 AI、就业和技术集中趋势的研究可提供 A 类和情景性基线，但不能直接证明某一后劳动制度方案[43-44]。

6.3 文化不是脱离承载者的信息包

AI 可以翻译、保存、重组和传播文本、图像、仪式记录与语言资源；功能胜任不等于取得代表一个群体发言的资格。文化意义存在于经历、关系、实践、土地、记忆、制度与代际承接中。一个模型复制了表达，不等于继承了经历；一个平台扩大可见度，也可能通过分类、排序和训练抽取重新分配文化权力。原住民数据治理中的 CARE 原则提醒研究者把集体利益、控制权、责任和伦理与开放数据并置[45]。这类原则不是所有文化的终局模板，却表明“数据可得”与“有权使用”必须分开。

这种文化权力不是 AGI 才出现的新现象。官僚制的分类、规则与档案早已把复杂生命转译为可管理对象；规训、观察和常态化判断也会通过知识形式塑造行动空间[30-31]。高级智能系统把两种结构推向实时、个性化和跨场景联动：它不仅描述文化，还可能决定何种表达可见、何种身份可验证、何种记忆可被训练。由此需要审查的不是系统是否“懂文化”这一项能力，而是分类权、代表权、拒绝权、知识许可与现实救济如何分配。

6.4 教育不是答案输送，关系不是拟真输出

生成式 AI 能够扩大知识可达、翻译、解释、练习与反馈，但信息获得不等于学习，答对不等于理解，理解不等于判断，个性化不等于自主，教学功能不等于教育关系。教育的文明任务，是使学习者在知识、关

系、行动与后果中形成判断，并把自己的生命带入共同世界。长期依赖 AI 可能释放教师和学习者，也可能把错误、画像、动机与价值判断外包给不可争议的系统；两种可能都需要纵向、跨文化和分群体的实证研究[46]。

AI 可以说出安慰的话、重建逝者声音、解释宗教文本或辅助仪式可达性，真实效果不因此为零；但功能相似不等于关系相同。模拟关怀是否成为能力外延，取决于身份透明、授权、可修正需要、真实退出、替代关系和后果承接。宗教共同体可依自身传统判断成员、圣职和仪式角色，公共制度不得把内部承认直接转化为治理权，也不得反过来命令共同体接受某种灵魂或主体理论。

6.5 宗教与精神传统：不是意义来源清单，而是反向压力

宗教与精神传统内部对灵魂、无我、死亡、永恒、祖先、土地、人格与非人能动性并无统一答案。本文不从它们中提取虚构共识，也不让任何传统为人工主体性背书；它们进入论证的职责，是检验“意义—有限性—主体”能否被准确翻译，并在误译或反例出现时迫使命题收缩。

《诗篇》90 把生命短促、数算日子所得的智慧与手中工作得以坚立并置；奥古斯丁在《忏悔录》卷四对友人死亡的叙述，则显示丧失撕裂的不只是一个孤立个体，而是由共同时间与关系构成的自我[58-59]。这两项材料既不代表全部基督教立场，也不证明死亡本身具有价值；它们对本文的压力是：有限性可以成为智慧和承担的背景，意义又可能指向超越死亡的关系，因而“有限性现实化”不能缩成“死亡制造意义”。

《无我相经》以无常、受苦和变易质疑把五蕴视为恒常自我，《大般涅槃经》则把有为法的败坏与精勤实践相连[60-61]；两者迫使主体连续性摆脱不变实体预设，却不能被误读为“没有人存在”或责任无从承接。《论语·先进》把“事人、知生”置于谈鬼神与死亡之前，《庄子·至乐》又在承认哀恸后以生成变化重置死亡视角[62-63]。它们共同阻止用死后本体论代替现世关系，也阻止把接受变化写成取消悲伤。

Watts 提出的地方—思想并不把土地当作被动环境，而把人类与非人存在置于已经运作的关系社会；Whyte 关于原住民气候正义的分析则显示，同意、信任、问责与互惠具有不可任意压缩的关系时间[64-65]。这不是所有原住民传统的统一教义，也不能直接证明机器意识；它要求本文重新检查两项默认：谁可以成为能动者与义务对象，何种“更快、更优”的技术行动会因跳过关系修复而继续制造不正义。

这些对话对文明跃迁形成三项受控修订：有限性不得只指生命期限，也包括关系与制度时间；主体连续性不得预设不变实体，却仍须说明经历、承诺和后果由谁承接；意义权属不得缩成个人偏好或数据占有，而要容纳共同体、祖先、土地与非人关系。公共制度仍须分别举证：系统被赋予灵性角色是社会—宗教事实，不自动证明第一人称体验、法律人格或政治资格；保护宗教自由也不能把某一传统的终极本体论强制为共同事实。若传统内部学者与实践者指出本文持续误译其核心概念，相关命题必须重开，而不能把反例重新命名后吸收。

6.6 共同世界不等于同一种世界

文明跃迁的全球化不能变成单一文化向世界输出终极答案。共同性应建立在共同存续、基本尊严、可争议程序、非强迫、可退出和避免不可逆伤害之上；意义来源、宗教信仰、好生活和主体观可以保持深刻差异。公共世界之所以能够共同存在，正在于行动与言说在多元者之间显现，而非所有差异被结算为同一世界观[29]。多元并不等于任何主张都免于批评：当某种意义主张进入公共强制，它必须说明对他者自由、平等和后果的责任。普遍底线也不等于可以忽略语言、殖民历史、发展差异和知识主权。

这使“文明跃迁走向世界”获得准确含义：不是把世界纳入已经完成的体系，而是让不同学科和文明以自己的证据与经验改变理论，同时共同阻止能力越权、责任蒸发和真实生命被总量化。世界化的标准不是概念被翻译了多少种语言，而是来自不同传统的反例是否能够真正改变命题、制度接口与研究优先级。

七、对话面四：AI、计算机与系统科学

7.1 信息、语义、真理与合法性不能互换

信息论能够精确处理通信中的不确定性与编码，却不把意义或真理包含在信息量本身[47]。控制论和反馈系统能够围绕目标调节行为，却不能由稳定控制回答目标从何而来、谁有权设定目标[48]。模型的一致性不等于事实真，形式证明只在规格、假设和系统边界内成立；多代理达成协议不等于公共合法性。计算与系统科学的有效遗产因此不是被否定，而是被放回其证明范围。

分布式系统中的时间排序、故障、并发与部分可见性说明，所谓“系统状态”必须绑定节点、时刻和观察条件[49]。安全工程进一步表明，事故可以在组件各自正常时由相互作用产生，系统风险不能只靠模型级准确率解释[51–52]。AGI 治理的技术对象因而至少包括模型、实例、智能体、应用、部署组织和更广的社会技术系统；把全部责任和保证压到基础模型上，与把模型排除在责任链外同样错误。

7.2 能力不等于自治，自治不等于自我授权

规划、工具调用、长期记忆、反思、目标维持和多步行动可以构成代理自治的工程证据，不能单独证明主观体验。系统在授权范围内自主选择方法，也不因此取得扩大目标、资源和权限的权利。安全设计需要最小权限、关键动作独立确认、资源限额、工具隔离、独立停止、证据保全和重新授权；经典最小权限原则在智能体生态中仍然有效[50]。版本能力可以组合，权限不能因组件叠加而自动累积。

7.3 “人在环”只有在能改变结果时才成立

人工签字、点击确认或形式监督不等于有效人类控制。控制至少要求人获得与后果相称的信息和时间，具有理解任务边界的能力，拥有真实偏离、替代、中止与修复权限，并得到组织支持。高速系统中，人不可能逐步理解全部内部状态，但仍可在目的、权限、禁行域、异常接管、停止和修复层保有实质控制。若这些条件不存在，把责任留给最后点击的人只会制造责任漂白。

7.4 回滚不是修复，连续性不是同一主体

暂停、隔离、回滚、恢复与修复是不同动作。内部状态可恢复，不代表世界中的财产、名誉、身体、关系与生态后果被恢复。信息连续、运行连续、因果连续、关系连续和主体连续也必须分别记录；复制、迁移、分叉、合并和版本继任是系统事实，不能单独宣布多个主体、同一主体或不存在主体。完整证据链应能重放事实、推断、建议、决定、授权、执行、影响与修复，而不把高敏感日志汇成一个人的可信、服从或意义总分。

7.5 对齐是一种持续关系，系统成熟包含退居

把对齐理解为一次性写入单一目标，会把价值冲突、环境变化、权力与受影响者排除在外。文明级对齐应当是持续关系：目的可争议、权限可撤回、证据可复核、差异可保留、事故可修复、版本可追溯。可信也不是一个总分，而是证据对限定主张的暂时支持。工程原型可以把这些要求转化为机器侧保证与权限架构、

人类判断与后果推演、真实主体能力外延、分布式公共协调等验证方向；原型是检验理论的场所，不是理论成立的循环证据。

系统退居不是突然关闭，也不是效率敌人。它包括降低不必要介入、把能力迁移给真实主体、保留供应商替换和低技术备份、完成数据与权限清理、验证公共服务连续性，并在退役后继续承担遗留后果。一个不可退出、不可替换、无法低技术维持基本服务的系统，即使日常表现优秀，也可能通过依赖获得事实主权。技术成熟的终点不仅是“什么都能做”，也包括“知道何时不再替人做”。

表 4 理论判断、工程接口与制度接口

理论判断	工程接口	制度接口
模型不等于系统	联合配置、语义版本、依赖谱系	部署主体、采购责任、跨组织问责
输出不等于决定	事实—推断—建议—授权—执行—修复链	正当程序、申诉、救济
能力不等于权限	权限图、工具闸门、转授权记录	法定授权、目的限定、撤回
人在环不等于控制	信息、时间、能力、偏离、中止、修复测试	职业责任、组织支持、禁止责任漂白
回滚不等于修复	暂停、隔离、回滚、恢复、修复分层	补偿、照护、生态恢复、长期责任
共识不等于合法性	来源独立、分歧保存、证据重放	代表、参与、少数意见、公共授权
紧急能力不产生常态权力	权限到期、自动降级、重新授权	期限、独立复核、遗留债务
系统必须能够退居	可迁移、可替换、低技术备份、退役验收	退出权、服务连续、反依赖

八、文明跃迁的独立贡献与学科反向修订

文明跃迁的原创性不在于分别宣布生命判准、算法正当程序、多维合法性、公共资源治理或系统安全等方向从未存在。其增量在于：以意义为根和有限性现实化为哲学轴，把散落于不同学科的区分重新放入同一条现实因果链，并要求每次从事实到地位、从地位到权力、从能力到文明方向的跨越都说明证据、授权、责任与失败条件。进一步说，它不仅讨论不同学科分别知道什么，也追问这些独立判断聚合、融合、冲突和相互作用以后，整体方向是否仍然成立。

1. AGI 使多种关系分离：AGI 相关技术加速、重组并制度化智能、生命、体验、主体、责任、资格、意义权属与公共权力之间的分离，使单一学科无法独立处理其全部后果。
2. 意义与有限性：意义为根、来源多元；有限性不是意义唯一来源，而是已知有限生命的选择、后果和不可替代重量进入现实的重要结构。
3. 肯定与永久否定都要举证：肯定与永久否定 AGI 主体性均承担证据责任；保护、法律资格、本体主体、政治资格和治理权分别审查。
4. 跨学科推导必须说明依据：事实不能直接生成价值，能力不能直接生成权限，法律配置不能证明意识，协议共识不能生成合法性。
5. 局部指标不能结算完整生命：局部量化服务限定目的，任何单一尺度不得结算完整生命、共同体或文明意义；受损者不能被总量收益追溯抹除。
6. 启动、稳定和紧急状态必须分开：启动、稳定与紧急相互正交；临时权限必须法律与技术双重死亡，启动期需要现实能力，稳定期需要防固化和长期问责。
7. 后劳动条件下四类问题不能互相替代：生存保障、收入收益、贡献承接与意义生活相互支持而不得互相吞并；生存资格不以系统是否需要人的劳动为前提。
8. 文明级系统必须可以退出、替换和退役：系统应可替换、可退出、可退役，并在真实主体能够承担时减少介入，把关系、选择、沉默、困难和共同生活归还给主体。
9. 学科独立与整体判断必须同时成立：学科独立保证局部事实、方法和证据不被整体需要绑架；文明方向、制度与方案则必须进一步检验这些局部判断在现实中的聚合、相互作用与反馈。局部正确不能自动证明整体正确。

8.1 理论不是单向升级学科

学科对话已经反向改变理论。生命科学迫使有限性从短寿与不可复制升级为多层结构；心灵哲学要求把行为表现与体验分开，并承认异质主体标准可能超出人类认知窗口；法学把责任扩展为完整决定链与版本化合法性；政治学阻止绩效吞并授权与合法性；经济学揭示价值、价格、贡献和所有权的多重链条；系统科学表明正常组件也会产生系统事故，并把回滚与世界修复分开。文明跃迁若不能被这些发现改变，就只是封闭教义。

反向修订也有边界：学科共识可以否定错误事实、暴露推导漏洞、迫使命题收缩，却不能仅凭权威地位裁决“意义是否为根”这一哲学承诺。哲学根命题必须接受反例和现实后果压力，而不是接受学科席位投票。相反，哲学也不能以根命题之名否认合成生命、神经活动、经济分配或系统事故的专业证据。平等对话的最低条件，是双方都放弃越权。

同样，这些共同推导规则没有上位豁免。它们只能要求跨界推导说明依据、证据、适用范围与失败条件，不能替任何学科规定其事实对象，也不能以“文明尺度”为理由跳过专业反例。若总论的任何概念只能吸收异议而不能被异议改变，或者以整体性之名取消学科的独立判断，它就已经违背自身方法。

8.2 为什么文明跃迁必然跨学科

文明跃迁涉及多个学科，不是因为试图扩大解释领地，而是因为 AGI 大规模进入社会以后，一个文明方向、一项制度或一个方案本身就同时进入多个领域。技术改变组织与制度，制度重配资源与机会，资源与机会塑造关系、心理和行为，再进入经济、法律、政治与下一轮技术。研究对象本身已经跨越学科，只停在一个学科内部，就无法完整解释现实后果。

这并不取消学科独立。生命科学仍独立判断生命事实，计算机科学仍独立验证技术能力，法学仍独立判断法源、资格与程序，经济学仍研究生产、分配和激励，政治学仍审查授权、代表和主权。真正改变的是整体判断：一个制度或方案是否正确，不再等于这些局部判断的简单相加，而要继续考察它们聚合以后如何相互强化、抵消、冲突或形成新的反馈。专业独立保证每一个局部尽可能准确，整体观察保证这些局部没有共同把文明带向错误方向。

因此，文明跃迁尊重既有理论与权威，却不把任何理论形成时的历史边界当作未来世界的认识边界。已有知识在其证明范围内继续有效；新对象、新关系和因果链超出原有范围，就继续向前研究。任何单一学科、经典理论或权威都不能仅凭既有地位终审同时跨越多领域的文明问题；文明跃迁也不能以“整体”覆盖充分成立的专业事实。它不替各学科作答，而是继续追问这些答案共同进入现实以后会发生什么。

8.3 三个文明方向不是可计算总目标

文明跃迁以“生存无忧、生活幸福、生命更有意义”表达三个递进而不可合并的方向。生存为生命和自由提供底板，幸福关注痛苦、关系、健康与实际能力，意义指向主体对存在、责任、创造和共同世界的形成与承担。三者不是一个可最大化的文明总分：生存不能用意义牺牲，幸福不能购买他者的被抹除，意义也不能由系统替主体定义。指标只能分域服务决策，并保留分布、少数、不可逆损害与异议。

九、最强反驳、竞争解释与修订条件

开放性不能只写成“欢迎批评”。以下反驳一旦成立，必须具体改变理论，而不是被吸收为修辞资源。

反驳一：这些关系的分离并不新。

法人、文字、官僚制、自动化和算法治理早已分离能力、责任和人格。回应：理论应放弃“所有分离首次发生”的历史断言，只主张 AGI 在组合广度、持续自治、现实执行和制度密度上重组既有分离。若比较史研究表明这种重组并无质变，‘文明级断裂’应降格为加速性解释。

反驳二：AGI 对象未定，整套理论建立在假设上。

回应：总论不依赖 AGI 到达日期或单一排行榜，而研究已经进入高后果链的高级系统及其发展方向。若相关系统长期无法形成跨域持续行动，研究议程应收缩为现有自动化治理，关于异质主体和主权的内容保留为情景研究。

反驳三：“意义为根”不可经验验证。

回应：它是公开的哲学公设，只能以解释力、融贯性、反例承受力和实践后果被批判。若不以意义为根的理论能够更好保护主体多元、解释文明方向并避免总量支配，文明跃迁必须重构其根基；不能以科学或传统权威替它证明。

反驳四：有限性命题浪漫化死亡和痛苦。

回应：有限性不等于短寿、疾病或匮乏，理论支持延寿、治疗与减少痛苦。若极长寿或非有限主体能够在没有任何选择排除、关系边界、后果承担与连续性约束的条件下生成稳定意义，‘有限性现实化’必须收缩为人类历史命题。

反驳五：主体性三维仍是人类中心主义。

回应：三维仅为研究入口，不是充分定义。若异质系统以非语言、非情感、非生物方式呈现多方法收敛的体验、内生关切或连续性证据，框架必须扩展；若三维长期无法产生可区分预测，则应被替换而非制度化永久门槛。

反驳六：概念分离只增加程序，不能产生知识。

回应：每项分离必须导向可观察对象、桥接假设、比较研究或可失败预测。若分离无法改变实验设计、法律结果、治理权配置或事故归因，它应退居分析术语，而非被保留为独立贡献。

反驳七：预防性保护会被企业借作权力护盾。

回应：保护必须与所有权、关停、商业秘密、资源控制和公共权力分开，并接受独立审查。若保护措施系统性削弱问责或挤压已有明确利益，应降低或重构措施，而不是以‘潜在主体’压倒现实受害者。

反驳八：更优 AGI 治理下，人类保留最终权力可能是不正当支配。

回应：持续绩效优势应改变建议权重、职能和审议结构，却仍不足以清偿代表、授权、权利与责任。若未来 AGI 获得主体和政治成员资格，制度必须开放平等参与；但计算能力本身不产生超额主权，这一限制同样适用于人类精英。

反驳九：拒绝用一个总分结算完整生命，会不会导致无法治理？

回应：理论反对跨对象、跨目的的最终总分，不反对局部指标。若某一综合指标在明确用途内可解释、可申诉且不外推，则可以使用；若没有任何聚合就无法做出公共选择，理论需说明可接受的有限聚合规则，而非停留在否定。

反驳十：系统退居会损害效率与公共服务。

回应：退居是降低不必要介入、迁移能力和保留替代，不是突然断供。若退出设计在特定场景持续增加净伤害，退居条件必须调整；但不可替换的依赖本身仍应作为结构性风险进入评估。

反驳十一：理论以全球之名输出单一文化。

回应：共同语法必须允许不同文明参与定义意义、主体、关系、责任与共同世界，并让反例真正改变命题。原典引用不能代替传统内部学者与实践者的异议，象征性列举也不能算作对话。若研究只发生在单一语言、地区和制度传统内部，或跨文化材料只能装饰既定结论，本文的‘世界性’主张不成立。

反驳十二：文明跃迁本身正在成为它所批评的学科霸权。

回应：如果这些共同推导规则能够绕过各学科的对象定义、证据标准与反例，以“文明尺度”直接宣布专业结论，这一批评成立。本文因此不赋予文明跃迁事实终审权、学科指挥权或统一概念权：总论只提出跨边界待证问题，专业事实由相应学科在其方法内部判断，跨学科推导必须公开依据，充分证据可以迫使命题收缩，不同文明传统也可以因翻译失败重开概念。这些规则的权限只是阻止没有依据的越界推导，而且首先约束自身。若实践中的文明跃迁只能吸收异议、不能被异议改变，或以整体之名压平专业自主，它应被判定为失败，而不是“整合成功”。

十、面向世界的共同研究议程

研究议程必须把哲学承诺转化为可合作、可失败、可修订的工作，而不能只罗列宏大问题。以下八个研究群均要求：说明当前已知、公开未知、方法路径和迫使理论修订的条件；在对象、版本、文化与法域之间保留差异。

表 5 八个国际共同研究群

研究群	核心问题	可开展的方法	修订触发条件
异质主体证据	如何在不制造潜在痛苦的条件下一区分体验、自我模型与策略性自述	跨架构干预、竞争解释、纵向一致性、非语言指标、预注册复核	三维框架不能产生可区分预测，或新证据显示完全不同的主体结构
生命与连续性	复制、分叉、迁移、暂停、恢复和极端延寿中，什么继续存在	生命实验、神经与认知研究、个人同一性分析、法律案例对照	跨载体第一人称连续获得独立、可重复的多方法证据
灰色地位	部分满足生命、体验、连续性或关切条件的对象应获何种保护	分级保护实验、伤害—成本—权力评估、伦理与比较法研究	保护无法与企业所有权和扩权分离，持续伤害现实主体
责任与版本化合合法性	自主智能体生态中如何保持决定链、证据链和修复链	版本谱系、事故重放、跨组织责任图、举证接近规则、比较法	现行法已能稳定覆盖全链且不产生责任缺口
结构性权力与后劳动	算力、数据、能源、模型和入口如何改变收益、依赖与生存资格	行业与国家比较、所有权和议价网络、分布影响、照护与生态核算	四轨结构不能解释或改善分配、照护与意义生活之间的冲突
教育、关系与文化	长期 AI 介入如何改变判断、依赖、代际承接和文化发言权	纵向研究、儿童与脆弱群体保护、多语言参与观察、社区共同研究	功能替代在不削弱主体形成和关系责任时长期稳定成立
多文明共同规则	不同传统能否在不统一终极答案时形成共同底线	多地点审议、原典与活传统互证、概念翻译、反例登记、传统内部学者与全球南方共同主持	关键传统的核心概念持续被误译、排除或只能靠强制统一维持
系统退居与长时治理	如何验证可迁移、可替换、低技术备份、退役和远时延自治	演练、故障注入、供应商退出测试、公共服务连续性和遗留责任审计	退居设计持续造成更高风险，且依赖可通过其他机制可靠约束

10.1 共同研究的最低程序

- 对象登记：明确模型、版本、实例、代理、组织、场景、时间和受影响者，禁止用产品名代替研究对象。
- 桥接登记：每个跨学科结论公开事实前提、解释假设、规范理由、制度选择与替代方案。

- 证据分层：A 类稳定支持、B 类竞争解释、C 类思辨连接分别标注；预印本、机构报告、法律与伦理建议不得互相冒充。
- 最强反驳：由能够代表竞争立场的研究者参与设计，不以体系内部的弱化版本代替真正反对意见。
- 伤害最小化：主体性与福利研究不以“可能没有体验”为由忽略潜在伤害，也不以“可能有体验”为由逃避必要审计。
- 跨文化共同主持：研究问题、概念翻译、数据权利和利益分配由不同地区、语言与传统共同决定。
- 阶段区分：启动、过渡、稳定和紧急分别记录目标、权限、退出与遗留责任，不把临时例外永久化。
- 负结果与版本历史：保存未收敛、被推翻和没有效果的结果，使理论修订可追溯，不用新版本抹除旧责任。

10.2 与国际规范进程的关系

欧盟《人工智能法案》、NIST AI 风险管理框架与欧洲委员会《人工智能、人权、民主与法治框架公约》不能被压平为同一套“AI 伦理原则”。它们分别主要位于可执行的区域市场法、组织风险治理和国际人权秩序三个层次：前者以用途和风险触发法律义务，NIST 以自愿且迭代的治理流程组织风险判断，欧洲委员会公约文本则要求国家使 AI 全生命周期活动与人权、民主和法治相一致[34–35,39,66]。三者可以互补，不能互换：合规门槛不能替代情境判断，组织治理不能自行产生法律正当性，国际原则也不能取代具体监管、程序和救济。

欧盟法案把健康、安全与基本权利纳入高风险系统的持续风险管理，并在特定公共服务、公共机关、信用与保险场景要求部署者在使用前评估基本权利影响[34]。这与本文“先分清对象、跨界推导说明依据”的要求相呼应：提供者对系统的风险评估不能替代部署者对具体权力关系与使用情境的说明。但法案的风险分类并不裁决机器是否具有体验、何种生命有意义或谁应拥有文明目的设定权；文明跃迁也不能以这些哲学问题为由改写法定对象。本文以 2026 年 7 月 27 日合并文本为核验基线；截至检索截点，多数一般规则和透明义务已经适用，部分高风险条款仍依修法后的阶段安排后续适用。

NIST AI RMF 1.0 以 Govern、Map、Measure、Manage 构成贯穿生命周期的自愿治理循环，并明确有效可靠、安全、韧性、问责透明、可解释、隐私与公平等“可信”特征必须依用途情境权衡；单项满足不保证整体可信[39]。生成式 AI 配套画像进一步把风险放到模型、应用、生态系统、价值链与时间尺度中，纳入虚构信息、偏差与同质化、人机配置、信息完整性和过度依赖等问题[66]。这支持本文拒绝把可信度做成脱离对象和环境的总体品格，却不等于 NIST 已经接受“意义为根”或“局部指标不能结算完整生命”的全部哲学论证；这里是方法上的相容与增量，不是权威背书。

欧洲委员会公约文本把风险从产品性能推进到人权、民主程序、法治、尊严、自主、透明、问责、平等、隐私、救济与公共讨论，并设置报告、缔约方会议、监督和国际合作结构[35]。截至 2026 年 8 月 24 日，条约办公室尚未列出其生效日期，因此本文只讨论其文本结构和制度方向，不把它写成已对所有签署方普遍生效的义务。它对文明跃迁的反向约束是明确的：主体性未知不能成为跳过现有权利、民主参与和有效救济的理由，未来主体保护也不能借国际合作之名制造不可争议权力。

UNESCO 伦理建议、OECD AI 原则与联合国《全球数字契约》分别扩展了人权伦理、包容增长、透明稳健和全球能力合作的公共基线[36-38]；AI 伦理框架与《联合国工商企业与人权指导原则》又提醒研究者，原则清单必须进入企业责任、尽责、补救与受影响者关系[53-54]。文明跃迁的补充在于把意义来源多元、有限性结构、主体性未知、资格分层、局部指标不能结算完整生命、系统能够退出和退居放入同一问题场；它同时接受这些规范对权利、程序、问责和跨境合作的反向校验。任何新增概念若削弱已有权利保护、把潜在主体变成企业免责工具，或以文明尺度取消具体救济，都应被拒绝。

10.3 学术传播的双层结构

跨学科学术传播采用双层结构：总论提出共同问题、共同推导规则与可修订的研究议程，专业深稿进入各学科内部，展开理论谱系、实证证据、法域案例、方法争议与失败条件。总论不以体系规模取得优先权，专业稿也不因独立投稿而失去共同问题。缺少专业深稿，总论易成宏大宣言；缺少总论，专业稿易失跨域因果与共同问题。二者构成“一篇总论统摄、多篇专业深稿展开”的传播体系，其成功标准不是各学科接受同一标签，而是能够在本学科内部识别新问题、提出反例并反向改变理论。

结语 不是统一答案，而是共同进入新问题世界

学科重新对账必须同时完成对象辨明与关系深入。各学科可以保持边界，但不能让现实因果在边界处停止；对多维关系、反馈与下一轮目的的研究，决定专业知识能否共同进入文明的整体判断。

过去，分科让知识获得精确；今天，AGI 使越来越多现实后果跨越学科边界。真正迫使学科重新对账的，不是某个系统是否已经达到一个争议性标签，而是智能、生命、体验、主体、责任、权利、资格、意义与权力之间的历史重合已不能继续被当作推导捷径。生命科学不能替哲学证明体验，哲学不能替生命科学规定生命事实；法律可以配置资格，不能由资格反证意识；政治可以授予职权，不能以绩效取代合法性；经济可以记录贡献，不能为人与意义定价；计算与系统科学可以构造保证，不能由能力生成目的与主权。

学科边界因此正在获得新的责任。它仍然要保护专业对象、方法和证据的独立性，同时还要说明证据在哪里止步、判断在哪里需要交给其他领域、权限从哪里重新授权、责任如何继续承接。越融合，越不能混同；越互为因果，越要知道谁证明了什么、谁有权决定什么、谁承担什么后果。学科独立不是现实割裂的理由，文明尺度也不是凌驾于专业之上的新权威。前者负责把局部看清，后者负责继续看见这些局部共同运行以后形成的整体后果。

文明跃迁保持根判断：新增力量持续被旧起始逻辑调用，博弈便沿人类之间以及条件成立时人类与 AGI 之间两条关系递归，侵蚀共同存续与重新选择的条件。改变利己的出发点，使真实利害相关者在目的生成前在场，才能改变下一轮关系怎样产生。意义为根而来源可以不同，有限性不独自赋予意义；能力不自动生成主体资格与公共权力，未来主体研究对证据开放。最高层方向指导方法，具体结果校准方法和现实因果认识；不能用一项局部成败替代共同方向。

学科重新对账的价值，要回到一个人仍能生活、选择、创造与承担的现实。专业知识既不能被宏大目的吞没，也不能在结果已经不可逆以后才进入。文明跃迁让真实利害在目标生成前在场，保留共同未来，走向生存无忧、生活幸福、生命更有意义。反证可以从具体一步进入：真实关系是否改变目的，新的合作是否继续转嫁成本，旧逻辑是否存在能够长期保存纠错空间的另一条路径。专业共同体的工作，在于把这些差异看清。

论证收束

知识可以继续分科，文明判断不能继续割裂。学科负责把局部看清，文明判断负责不让局部彼此失联。理论的价值不在于拥有终局答案，而在于让每一次跨学科判断说明证据、边界、授权与责任，并允许新事实、新学科和新主体真正改变它。

续读与共同参与：不同入口，同一条向前的路

本文在这里完成的是共同问题、学科边界、推导责任与研究议程，不替十三个专业领域完成各自的事实裁决，也不替制度、工程和实验给出最终参数。继续阅读的意义，不是把读者留在一套文稿中，而是让问题跨出本文边界以后仍然保持开放：下一步为什么要交给别的学科、哪些证据能够改变当前判断、哪些方案值得比较、哪些现实结果会迫使理论重新修订。

不同读者可以保留不同判断，也不需要先接受文明跃迁。可以从赞同进入，也可以从反驳进入；可以只研究一个局部问题，也可以从制度、工程、实验或传播开始。共同点不是先接受同一种答案，而是在 AGI 时代共同面对已经不能由单一学科独立完成的问题。

如果继续追问：这些局部判断共同运行后，会把文明带向哪里？

本文止于“局部正确不能自动证明整体正确”。如果继续追问一个技术、制度或政策在长期运行中怎样同时改变存续、资源、权利、关系、判断能力、幸福与意义，就需要进入文明跃迁的整体研究。文明跃迁五卷书与体系总入口继续回答：为什么必须跃迁、文明要走向什么目标、制度和运行结构如何承接这些目标，以及不同专业判断共同进入现实后应如何被长期检验。向上进入整体，只增加观察范围，不赋予“文明尺度”替代专业事实的权力。

如果关心：这些判断怎样进入制度、工程和现实验证？

目的生成、现实行动与结果反馈应进入同一因果链，但方向、制度假说、技术方法和已获得结果必须分层。向实践继续，不意味着把尚未验证的方案写成实施命令；它意味着选择可以失败和修正的局部方法，检查哪些真实利害能够改变目标，以及这些变化是否改善人的处境。任何单项方法都不能单独证真文明跃迁。

如果不同意本文：从哪里开始反驳？

最有效的反驳不是拒绝“文明跃迁”这个名称，而是击中具体命题。可以证明智能、体验、主体、责任、资格与权力之间并没有本文所说的新分离；可以提出更好的机器意识与主体连续性理论；可以证明“意义为根”不能承担文明方向；可以说明有限性并非本文所说的现实结构；可以证明某些综合指标能够在不抹除真实生命的条件下完成更好的公共决策；也可以通过比较制度、工程结果或跨文化材料证明本文遗漏了关键主体和后果。

十三篇专业深稿分别进入哲学、认识论、认知与神经科学、伦理与法学、政治学、系统科学、生命科学、社会与文明研究、经济学、人工智能与计算机科学、物理与信息研究、宗教与精神传统、教育与代际承接。横向续读的目的，不是寻找支持本文的同温层，而是把每一个跨界结论重新交给真正有能力证明、限制或推翻它的专业共同体。

如果你来自一个专业：什么工作能够真正改变本文？

共同研究不以赞同、背书、组织身份或长期归属为前提。一次高质量反驳、一组长期数据、一次复现失败、一种更强模型、一项比较法研究、一个工程压力测试或一段来自现实共同体的反例，都可能比一句赞同更重要。下列工作尤其能够直接改变本文：

- 哲学、认识论与主体研究：提出更强的意义、主体、个人同一性或机器意识理论，给出能够区分竞争解释的证据标准，并说明哪些反例会迫使本文收缩或重写。
- 生命、认知与神经科学：用新的生命形式、脑类器官、跨物种意识研究、长期连续性证据或负结果，校准本文关于生命、体验、有限性与主体连续性的边界。
- 人工智能、系统与工程：通过评测、故障注入、形式验证、事故研究、权限与退出测试，检验“模型不等于系统、能力不等于权限、回滚不等于现实修复”等判断在真实系统中是否成立。
- 伦理、法学与政治研究：比较保护、资格、责任、授权、合法性与主权的不同制度路径，证明哪些边界能够保护主体，哪些设计反而制造新的权力伤害。
- 经济、社会、文化、教育与宗教研究：提供长期分布数据、田野材料、跨文化反例、学习与依赖研究、传统内部解释和替代制度，使“世界性”真正能够被不同经验改变，而不是只扩大引用范围。
- 制度、工程与现实实践：把理论转化为可观察的试点、标准、工具和失败报告；如果现实结果证明某项机制不可行、代价过高或方向错误，应让理论和方案真正回退、修订或退出。

向前：从共同研究走向文明跃迁

文明跃迁不会由一套理论、一门学科、一个组织或一群已经互相认同的人单独完成。它需要不同的人从不同问题出发，把事实、反例、价值判断、制度设计、工程能力和现实经验带进同一个文明进程。有人证明，也有人反驳；有人提出制度，也有人在实验中证明制度需要重写；有人把技术做出来，也有人提醒技术已经越过了不应越过的边界。

最有价值的参与，不是赞同，而是足以改变理论的专业工作。续读的终点不是下一

篇文稿，而是让更多人从看见问题走向独立判断，从独立判断走向共同研究、传播、推

动与实践，最终汇入人类文明跃迁的真实进程。

不是为了让更多人读完文明跃迁，而是为了让更多人一起完成文明跃迁。

当前公开入口：[文明跃迁完整文库](#)；[公众总入口](#)。具体专业稿或共同研究页面尚未建立时，以文库入口为准；网站负责维护动态版本、精确链接与后续参与入口。

附录 A 核心判断与证据状态索引

表 A1 核心判断的当前状态

判断	状态	说明／再审入口
现有高级 AI 具有显著、但任务与版本依赖的能力	A	以具体评测、部署与时间为限；不外推为通用可靠性
当前公开证据不足以证明 AGI 主观体验	A（证据状态判断）	新架构、多方法因果证据和竞争解释排除可触发再审
全局工作空间、整合信息、具身等可解释机器意识	B	竞争理论，尚无单一路径获得终审地位
主体性三维可作为强主体研究入口	B／C	分析框架，不是科学共同体已确认定义
意义为根、来源多元	哲学公设（不适用 A／B／C）	以融贯、解释力、反例承受和实践后果审查
有限性使已知人类意义现实化	结构命题（经验桥接为 C）	极端延寿、复制、异质主体和跨文化研究可修订其范围
能力、资格、权力必须分别证成	规范—制度判断	以比较法、治理绩效、权利与分布后果持续检验
系统退居是文明级成熟条件	制度假设	以退出演练、服务连续性、依赖风险和净后果验证

附录 B 术语边界

表 B1 核心术语边界

术语	本文含义
AGI 时代	高级智能系统及其发展方向已使多种历史重合关系前置分离的文明条件；不等于宣称存在公认完成态 AGI。
意义主体	能够形成、经历、认领并承担意义的主体性位置；当前不由语言功能或社会承认自动证明。
意义协作者	在授权与责任边界内理解、保护、传递意义并扩展真实主体能力的系统角色。
有限存在	与主体的选择排除、历程、关系、后果和不可任意转移连续性相关的存在结构；不等于资源受限。
主体连续性	第一人称存在是否持续的开放问题；不同于信息、运行、因果、关系和叙事连续。
保护	在证据不充分时减少潜在伤害的比例化措施；不自动生成完整人格、所有权、资源权或政治权力。
法律资格	制度为特定目的配置的权利、义务、程序或代理位置；不是意识的本体证明。
政治资格	进入公共成员、代表和参与结构的地位；与法律人格、职能和主权分别证成。
局部指标不能 结算完整生命	拒绝把局部指标合成为结算完整生命或文明意义的最终尺度；不拒绝目的明确的有限测量。
系统退出与退 居	系统在任务完成、风险过高或真实主体可以承接时降低介入、迁移能力、退出与退役，并承接遗留责任。

附录 C 共同接口与资料状态

阶段不创造权源；任务紧迫性须另证。到期停止新增越权执行，必要交接有界且可核验。历史非抹除与生命意义的规范地位分层保留。E/C/O 是研究方向，不是联合保护门槛。A/B/C/F 标论证身份，不是统一置信度排名。

当前主体证据不足不意味着永久否定；未来开放不意味着现时确认。意义功能的部署许可、能力与实际效果分别判断。科学事实不能替代规范理由，规范方向也不覆盖专业事实。

本版理论基准同步至 2026 年 10 月 7 日。各稿经验材料沿用其明确检索时点；法规、工作论文、草案、临床登记和更正仍按具体版本判断。关键来源核对不是全部书目通过。关于 NCT07290244，已确认官方登记为 ER-100 首个人体一期安全性与耐受性研究；本版不以登记本身确认实际入组或疗效。

本版核对了 Revah 更正原文，Fig.5a 的 mPFC 图板曾误用 A1 图板，更正对象是该图板，不是整项研究全部结论被撤销。Owen 教义评估须保留人工植入错误的样本设计，不把植入错误归为模型自然生成矛盾。相关原始出处见对应卷参考文献。

反证入口：哪条真实利害改变了目的，何种关系权重由谁裁决，合作是否将损害转嫁给缺席者，旧逻辑是否存在长期阻断递归并保存试错条件的路径。对局部方法的检验应标明对象、条件和结果，不把它换成整个文明方向的结算。

参考文献

- [1] Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433–460.
<https://doi.org/10.1093/mind/LIX.236.433>
- [2] Searle, J. R. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 3(3), 417–424.
<https://doi.org/10.1017/S0140525X00005756>
- [3] Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, 83(4), 435–450.
<https://doi.org/10.2307/2183914>
- [4] Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, 58(1), 7–19. <https://doi.org/10.1093/analys/58.1.7>
- [5] Varela, F. J., Thompson, E., & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.
- [6] Dehaene, S., Lau, H., & Kouider, S. (2017). What Is Consciousness, and Could Machines Have It? *Science*, 358(6362), 486–492. <https://doi.org/10.1126/science.aan8871>
- [7] Tononi, G., & Koch, C. (2015). Consciousness: Here, There and Everywhere? *Philosophical Transactions of the Royal Society B*, 370(1668), 20140167. <https://doi.org/10.1098/rstb.2014.0167>
- [8] Butlin, P., Long, R., Elmoznino, E., et al. (2023). Consciousness in Artificial Intelligence: Insights from the Science of Consciousness. arXiv:2308.08708. [研究报告／预印本] <https://arxiv.org/abs/2308.08708>
- [9] Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.
- [10] Birch, J. (2024). *The Edge of Sentience: Risk and Precaution in Humans, Other Animals, and AI*. Oxford University Press.
- [11] Varela, F. G., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The Organization of Living Systems, Its Characterization and a Model. *BioSystems*, 5(4), 187–196. [https://doi.org/10.1016/0303-2647\(74\)90031-8](https://doi.org/10.1016/0303-2647(74)90031-8)
- [12] Szathmáry, E., & Maynard Smith, J. (1995). The Major Evolutionary Transitions. *Nature*, 374, 227–232.
<https://doi.org/10.1038/374227a0>
- [13] Queller, D. C., & Strassmann, J. E. (2009). Beyond Society: The Evolution of Organismality. *Philosophical Transactions of the Royal Society B*, 364(1533), 3143–3155. <https://doi.org/10.1098/rstb.2009.0095>
- [14] Gibson, D. G., Glass, J. I., Lartigue, C., et al. (2010). Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science*, 329(5987), 52–56. <https://doi.org/10.1126/science.1190719>
- [15] Hutchison, C. A. III, Chuang, R.-Y., Noskov, V. N., et al. (2016). Design and Synthesis of a Minimal Bacterial Genome. *Science*, 351(6280), aad6253. <https://doi.org/10.1126/science.aad6253>
- [16] Takahashi, K., & Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, 126(4), 663–676. <https://doi.org/10.1016/j.cell.2006.07.024>
- [17] Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J., & Campbell, K. H. S. (1997). Viable Offspring Derived from Fetal and Adult Mammalian Cells. *Nature*, 385, 810–813. <https://doi.org/10.1038/385810a0>

- [18] Lancaster, M. A., Renner, M., Martin, C.-A., et al. (2013). Cerebral Organoids Model Human Brain Development and Microcephaly. *Nature*, 501, 373–379. <https://doi.org/10.1038/nature12517>
- [19] Trujillo, C. A., Gao, R., Negraes, P. D., et al. (2019). Complex Oscillatory Waves Emerging from Cortical Organoids Model Early Human Brain Network Development. *Cell Stem Cell*, 25(4), 558–569.e7. <https://doi.org/10.1016/j.stem.2019.08.002>
- [20] International Society for Stem Cell Research. (2021). ISSCR Guidelines for Stem Cell Research and Clinical Translation. *Stem Cell Reports*, 16(6), 1398–1408. <https://doi.org/10.1016/j.stemcr.2021.05.012>
- [21] López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2023). Hallmarks of Aging: An Expanding Universe. *Cell*, 186(2), 243–278. <https://doi.org/10.1016/j.cell.2022.11.001>
- [22] Tronto, J. C. (1993). *Moral Boundaries: A Political Argument for an Ethic of Care*. Routledge.
- [23] Elish, M. C. (2019). Moral Crumple Zones: Cautionary Tales in Human–Robot Interaction. *Engaging Science, Technology, and Society*, 5, 40–60. <https://doi.org/10.17351/ests2019.260>
- [24] Jonas, H. (1984). *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. University of Chicago Press.
- [25] Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.
- [26] Habermas, J. (1996). *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. MIT Press.
- [27] Ostrom, E. (2010). Beyond Markets and States: Polycentric Governance of Complex Economic Systems. *American Economic Review*, 100(3), 641–672. <https://doi.org/10.1257/aer.100.3.641>
- [28] Pettit, P. (1997). *Republicanism: A Theory of Freedom and Government*. Oxford University Press.
- [29] Arendt, H. (1958). *The Human Condition*. University of Chicago Press.
- [30] Foucault, M. (1977). *Discipline and Punish: The Birth of the Prison*. Pantheon Books.
- [31] Weber, M. (1978). *Economy and Society: An Outline of Interpretive Sociology*. University of California Press.
- [32] European Union. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation), Article 22. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
- [33] Court of Justice of the European Union. (2023). SCHUFA Holding (Scoring), Case C-634/21, Judgment of 7 December 2023. <https://curia.europa.eu/juris/liste.jsf?num=C-634/21>
- [34] European Union. (2024, consolidated 27 July 2026). Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02024R1689-20260727>
- [35] Council of Europe. (2024). Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, CETS No. 225. <https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treatynum=225>
- [36] UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- [37] OECD. (2024). OECD AI Principles (updated May 2024). <https://oecd.ai/en/ai-principles>

- [38] United Nations General Assembly. (2024). Pact for the Future, Annex I: Global Digital Compact, A/RES/79/1. <https://undocs.org/A/RES/79/1>
- [39] National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1. <https://doi.org/10.6028/NIST.AI.100-1>
- [40] Polanyi, K. (1944). *The Great Transformation*. Farrar & Rinehart.
- [41] Sen, A. (1999). *Development as Freedom*. Alfred A. Knopf.
- [42] Acemoglu, D., & Johnson, S. (2023). *Power and Progress: Our Thousand-Year Struggle over Technology and Prosperity*. PublicAffairs.
- [43] Gmyrek, P., Berg, J., Kamiński, K., Konopczyński, F., Ładna, A., Nafradi, B., Rosłaniec, K., & Troszyński, M. (2025). *Generative AI and Jobs: A Refined Global Index of Occupational Exposure*. ILO Working Paper 140. <https://doi.org/10.54394/HETP0387>
- [44] UNCTAD. (2025). *Technology and Innovation Report 2025: Inclusive Artificial Intelligence for Development*. United Nations. <https://unctad.org/publication/technology-and-innovation-report-2025>
- [45] Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., et al. (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19, 43. <https://doi.org/10.5334/dsj-2020-043>
- [46] Miao, F., & Holmes, W. (2023). *Guidance for Generative AI in Education and Research*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- [47] Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 379–423, 623–656.
- [48] Wiener, N. (1948). *Cybernetics: Or Control and Communication in the Animal and the Machine*. MIT Press.
- [49] Lamport, L. (1978). Time, Clocks, and the Ordering of Events in a Distributed System. *Communications of the ACM*, 21(7), 558–565. <https://doi.org/10.1145/359545.359563>
- [50] Saltzer, J. H., & Schroeder, M. D. (1975). The Protection of Information in Computer Systems. *Proceedings of the IEEE*, 63(9), 1278–1308. <https://doi.org/10.1109/PROC.1975.9939>
- [51] Perrow, C. (1984). *Normal Accidents: Living with High-Risk Technologies*. Basic Books.
- [52] Leveson, N. G. (2012). *Engineering a Safer World: Systems Thinking Applied to Safety*. MIT Press.
- [53] Floridi, L., Cows, J., Beltrametti, M., et al. (2018). AI4People—An Ethical Framework for a Good AI Society. *Minds and Machines*, 28, 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- [54] United Nations. (2011). *Guiding Principles on Business and Human Rights*. https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf
- [55] Williams, B. (1973). The Makropulos Case: Reflections on the Tedium of Immortality. In *Problems of the Self: Philosophical Papers 1956–1972* (pp. 82–100). Cambridge University Press. <https://doi.org/10.1017/CBO9780511621253.008>
- [56] Ricoeur, P. (1992). *Oneself as Another*. University of Chicago Press.
- [57] Wolf, S. (2010). *Meaning in Life and Why It Matters*. Princeton University Press.

- [58] The Holy Bible, New American Bible, Revised Edition. Psalm 90. United States Conference of Catholic Bishops.
<https://bible.usccb.org/bible/psalms/90>
- [59] Augustine. (1991). *Confessions* (H. Chadwick, Trans.). Oxford University Press.
- [60] Saṃyutta Nikāya 22.59, The Discourse on the Characteristic of Not-Self (Bhikkhu Sujato, Trans.). SuttaCentral.
<https://suttacentral.net/sn22.59/en/sujato>
- [61] Dīgha Nikāya 16, The Great Passing Away (Bhikkhu Sujato, Trans.). SuttaCentral.
<https://suttacentral.net/dn16/en/sujato>
- [62] Confucius. (2003). *Confucius: Analects, with Selections from Traditional Commentaries* (E. Slingerland, Trans.). Hackett.
- [63] Zhuangzi. (2020). *Zhuangzi: The Complete Writings* (B. Ziporyn, Trans.). Hackett.
- [64] Watts, V. (2013). Indigenous place-thought and agency amongst humans and non humans. *Decolonization: Indigeneity, Education & Society*, 2(1), 20–34.
- [65] Whyte, K. (2020). Too late for Indigenous climate justice: Ecological and relational tipping points. *WIREs Climate Change*, 11(1), e603. <https://doi.org/10.1002/wcc.603>
- [66] National Institute of Standards and Technology. (2024). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*, NIST AI 600-1. <https://doi.org/10.6028/NIST.AI.600-1>