

文明跃迁理论体系

AGI 时代人类文明跃迁的机遇、方向、方法与目标实现

——为什么必须改变、可以走向哪里、怎样抵达

AGI 给人类的最大机会，不只是获得一种更强的工具，而是第一次有可能重新安排生产、知识、时间、关系和人的位置。

文明方向合流深稿 V1.1

作者：子君赋

出品：文明跃迁研究组

证据截点：2026 年 8 月 26 日

摘要

人工智能正在从单项工具扩展为跨越科研、劳动、教育、医疗、治理、文化、关系、安全与基础设施的通用能力条件。2026 年《AI Index》显示，受访组织中的 AI 采用已达到 88%，生成式 AI 进入至少一个业务功能的组织占 79%；与此同时，代理型系统的现实部署仍显著落后于模型能力展示 [1]。国际劳工组织 2025 年的任务级研究估计，全球约四分之一就业位于某种生成式 AI 暴露范围，但现阶段更有依据的结果是工作转型而非简单整岗消失 [3]。IEA 则预计，在其基准情景中，全球数据中心用电将由 2024 年水平增长至 2030 年的约 945 TWh，AI 是增长的重要驱动因素 [9]。这些事实并不证明某一种未来必然到来，却足以说明：AI 已不再只是一个行业变量，而正在成为多个学科、制度和现实主体共同面对的文明级条件。

本文承接《AGI 时代的人工智能应用前景与风险研究——从能力扩展到文明后果》对现实传导关系的分析，不再重复七类应用场景，而进一步讨论：当这些变化共同发生时，人类为什么必须主动选择文明方向；哪些变化已经可见，哪些在当前能力、成本和组织采纳趋势延续时具有较高概率，哪些未来仍应保持开放；AGI 及其前身技术究竟为人类打开了哪些文明机会，又可能把文明自然推向哪些并不值得的方向。

本文公开区分 A/B/C/F 四种论证身份。现有采用、劳动暴露、能源约束、正式国际原则等属于 A 类事实或制度材料；对教育、组织和社会长期机制的竞争解释属于 B 类；“判断能力稀缺上升”“职业意义来源重组”“系统性依赖可能形成”等跨域未来判断属于 C 类；“存续是底线，生存无忧→生活幸福→生命更有意义构成文明递进方向”“技术应扩展人而不能静默取消人的主体位置”等属于作者公开承担的 F 类规范承诺。本文不把文明方向伪装成科学事实，也不让事实自动替代价值选择。

本文提出，AGI 时代最重要的机会不是“让机器替人做更多”，而是把新的生产、知识、科学、教育、医疗和协作能力转化为人的真实生活能力：使基本生存越来越少依赖对单一劳动岗位的强制依附，使更多人拥有可支配时间、健康、安全、关系与现实选择，并扩大创造、照护、学习、公共参与、信仰、探索和被需要等多种意义来源的可进入条件。相反，如果生产率提高并未转化为广泛自由，如果组织持续把判断和责任外包，如果教育变成答案配送，如果基础设施与公共权力高度集中，如果关系与文化越来越被少数同类系统中介，文明可能在系统更强的同时让人变薄。

因此，本文把文明跃迁理解为一种方向性的现实重组，而不是一次性制度替换。方法上强调从现实问题和真实受影响者出发，新旧并行、比较验证、允许失败与退出；让 AI 扩展人的理解、判断、创造、责任与共同方向生成；在不同阶段改变重点，避免将启动期的紧急安排永久化。目标实现不建立一个“文明总分”，而观察三类方向性变化是否持续出现。本文最终主张：技术打开可能，思想决定方向，现实检验道路。

关键词：AGI 时代；文明跃迁；人工智能；文明级事件；人类发展；结构意义；主体能力；生存无忧；生活幸福；生命意义

Abstract

Artificial intelligence is becoming a general-purpose condition that cuts across science, work, education, health care, government, culture, relationships, security, and infrastructure. This paper does not attempt to predict an exact arrival date for AGI, nor does it assume machine consciousness, superintelligence, or post-scarcity. Instead, it asks a more immediate civilizational question: as increasingly capable AI systems enter real institutions and reshape knowledge, labor,

responsibility, resources, relationships, and power, what directions remain open to human choice?

The paper distinguishes what is already visible, what may plausibly continue if current capability, cost, and adoption trends persist, and what remains genuinely open. It treats the following as normative commitments rather than scientific findings: civilizational survival is a non-negotiable floor; beyond survival, the desired direction is progressively toward freedom from basic insecurity, a happier and richer lived life, and wider conditions for people to generate meaning through relationships, creation, care, learning, responsibility, public participation, faith, exploration, and existence itself. These goals cannot be collapsed into one score or traded against basic rights.

The proposed method is deliberately non-blueprint: begin with real problems and affected people; allow old and new arrangements to coexist; compare outcomes; preserve the capacity to fail, exit, and revise; use AI to augment human judgment, responsibility, growth, and direction-setting rather than silently displace them; and adapt priorities across phases. Civilizational direction is therefore not a forecast. It is a choice that must remain answerable to evidence, professional knowledge, lived experience, and failure.

Keywords: AGI era; civilizational transition; artificial intelligence; human development; structural meaning; human agency; well-being; meaningful life

研究定位、对象边界与论证纪律

本文是“学科升维至文明”两篇合流稿中的第二篇。第一篇《AGI 时代的人工智能应用前景与风险研究——从能力扩展到文明后果》V1.2 负责把 AI 应用从模型能力一路追踪到组织、人机分工、资源、权力、关系与文明后果；本文只引用其已经建立的现实条件，不再逐章复述应用事实。本文的职责是：当这些现实变化已经汇合并形成文明级事件时，正面回答人类希望它们通向哪里、怎样把技术机会转化为更好的生活条件、怎样判断我们是否真的正在抵达。

现实按因果发生，不按学科发生。学科分别求真，文明共同承担后果。

这两句话不是要求所有学科融合成一个“超级学科”。计算机科学仍负责能力与系统保证，经济学负责生产、分配与激励，生命科学与医学负责生命事实与健康证据，教育研究负责学习与成长，法学与政治学负责资格、授权、权利与公共权力，哲学和伦理负责公开讨论目的与规范。文明尺度只继续追踪：这些分别成立的判断进入同一个现实以后如何相互作用、形成反馈与聚合，并最终改变人的生活。

论证身份	本文中的典型内容	能够支持什么	不能自动推出什么
A 稳定事实/正式材料	AI 采用、就业暴露、能源需求、正式国际文件、现实制度状态	限定时间、对象与条件下的事实判断	完整社会后果、终极方向、主体性
B 前沿/竞争解释	教育长期效果、组织激励、能力保留、关系影响	提出竞争机制与研究方向	写成已经收敛的普遍结论
C 跨域结构判断	高概率延续情景、意义来源重组、能力集中与事实权力	形成可反驳的文明级假说	伪装成单一学科已经证明
F 哲学/规范承诺	三目标、人的主体位置、不可用效率取消基本权利	公开说明本文选择什么方向以及为什么	替代经验研究、法律程序或公共授权

本文使用“AGI 时代”指高级 AI 持续进入知识生产、组织行动、资源配置与高后果决策链，使社会开始按“更强智能普遍可调用”的条件重新组织。它不是对一个已经完成且具有统一定义的 AGI 实体的事实预设。当前系统使用“当前高级 AI”“生成式 AI”“智能体系统”等具体称谓；未来 AGI 作为前瞻对象。

当前公开证据仍不足以确立高级 AI 具有主观体验、不可转移的主体连续性或自身后果意识。能力、体验、主体连续性、保护、法律资格、政治成员资格、公共职权与主权必须分别证明，不能自动串联。未来若出现足够强且多方法收敛的主体性证据，应重新讨论其伦理与制度地位；本文不以今天的证据不足永久封闭未来，也不以未来可能性提前授权当前系统。

目录

导论 | 为什么今天第一次出现真正的文明选择窗口

一 | 已经可见：AI 为什么正在成为所有学科共同面对的文明级条件

- 二 | 高概率可能：如果当前能力、成本与组织采纳趋势继续，什么更可能发生
- 三 | 真正开放的未来：哪些重大问题仍不能被提前写成必然
- 四 | 机遇：从更强工具到重新安排人类生活条件
- 五 | 什么不能自动发生：技术进步不会自己生成好文明
- 六 | 方向：存续底线与生存无忧、生活幸福、生命更有意义
- 七 | 人的位置：文明跃迁不是让人“还有事做”，而是让人有真实生活与多种意义来源
- 八 | 共同问题：不同学科怎样一起面对文明级事件，而不取消专业边界
- 九 | 方法一：从现实问题和真实受影响者出发
- 十 | 方法二：新旧并行、比较验证、允许失败与退出
- 十一 | 方法三：AI 扩展人的能力，保留判断、责任、成长与方向
- 十二 | 方法四：按阶段改变重点，不把紧急安排永久化
- 十三 | 目标实现：怎样观察“生存无忧”正在成为现实
- 十四 | 目标实现：怎样观察“生活幸福”正在成为现实
- 十五 | 目标实现：怎样观察“生命更有意义”的条件正在扩大
- 十六 | 从方向到生活世界：愿景、制度、工程与现实系统各自承担什么
- 十七 | 现实如何否定我们：失败、反例与文明自我修正
- 结语 | 技术打开可能，思想决定方向，现实检验道路
- 附录 A | 文明级事件—学科—方向矩阵
- 附录 B | 已经可见／高概率延续／真正开放：未来判断检查表
- 续读与共同参与 | 不是为了读完理论，而是让理论继续被现实改变
- 生成式人工智能使用声明
- 参考文献

导论 | 为什么今天第一次出现真正的文明选择窗口

很多技术都改变过文明。农业改变定居、土地与权力；工业化重写劳动、城市、教育与时间；互联网改变知识、传播和组织。AI 的特殊之处，不是“机器第一次帮助人”，而是越来越多过去必须由人完成的认知、生成、分析、判断准备、协调和执行环节，可以被同一种通用技术基础设施跨领域调用。2026 年《AI Index》记录到，组织采用与个人使用都继续快速上升，AI 进入科学、教育、医疗和政府的范围也在扩大[1]。这一变化仍远未完成，但它已经改变了很多领域可以彼此独立处理问题的条件。

只看单项应用，很容易把今天理解成“又一次效率革命”。但当 AI 同时进入工作、知识、教育、医疗、治理、关系、文化和国家竞争，同一个技术条件会穿过经济、法律、社会、政治、心理、生命与意义结构。一个企业选择自动化可能是局部合理的；大量企业同时自动化会改变劳动议价、教育选择和公共财政。一个学校采用 AI 辅导可能提高学习效率；整个教育体系长期围绕 AI 重构，会改变下一代的判断能力与成长路径。一个政府采用 AI 可以提高服务速度；多个公共系统深度依赖少数基础设施后，会出现事实权力、退出能力和系统冗余问题。

这就是“文明选择窗口”的含义。它不是说人类今天突然拥有了一张空白蓝图，也不是说旧制度会自动崩塌，而是说：一批过去很难同时改变的条件——生产能力、知识成本、组织协调、教育供给、科学发现、医疗辅助、跨语言沟通——正在被同一种技术同时松动。技术越强，文明越不能只问“能不能做”，而必须重新问“为了什么做、由谁决定、收益怎样进入人的生活、哪些过程不能被静默交出去”。

UNDP《2025 年人类发展报告》把 AI 时代概括为“选择的问题”，强调人类发展的重点不在于机器看起来多像人，而在于人能够实际拥有和实现怎样的选择与能力[2]。本文与这一国际发展框架在问题意识上相通，但三目标及文明跃迁的规范方向由作者自行承担，并不借 UNDP 或任何国际机构为其提供规范授权。

技术给出机会，文明决定机会最终变成什么。

本文接下来先把未来判断分成三层。第一层是已经可见的现实，不需要等待 AGI。第二层是在当前能力、成本和组织采纳趋势延续的条件下，具有较强现实依据、但仍需被反证的高概率变化。第三层是真正开放的未来：AGI 何时达到某种统一定义、机器是否形成主体性、超级智能是否出现、极端丰裕是否实现等。这种区分的意义是让文明方向既不被技术决定论绑架，也不因未来不确定而失去今天行动的理由。

一 | 已经可见：AI 为什么正在成为所有学科共同面对的文明级条件

1.1 文明级事件不是“重大新闻”，而是改变多个领域输入条件的共同变化

本文所说的“文明级事件”，不是指某一次产品发布或某一家公司领先，而是指一种变化同时改变多个学科和社会系统的输入条件，并在反馈后产生单一领域无法独立承担的后果。AI 进入知识工作就是这样的事件：它先改变任务成本和速度，随后改变岗位设计、教育信号、基础设施投入、组织权力、法律责任和人的成长路径。

这种事件对所有学科的重要性并不相同，但任何一个专业都不能把它当作“外围背景”。现实不会等经济学、政治学、教育学和哲学依次完成研究后再发生；这些变化在同一时间相互作用。专业边界仍然保护求真的质量，但不能再被当成现实因果的终点。

1.2 已经出现的八类共同变化

文明级变化	已经可见的现实	必须共同进入的专业
智能能力大规模社会化	越来越多个人与组织可直接调用生成、分析、代码、翻译与规划能力	AI/计算机、经济、教育、组织、社会文化
工作任务与组织流程重组	任务暴露和流程自动化扩展，现阶段更可能是岗位内容转型而非统一消失[3]	经济、组织、教育、法政、社会
知识生产方式变化	AI 进入科研、文献、代码、实验设计和科学基础设施[1][19]	认识论、科学哲学、各自然科学、教育、政治经济
判断能力逐步外部化	推荐、辅助决策、自动化与智能体使更多判断环节先经过系统	认知、系统科学、伦理法学、政治、教育
算力与基础设施扩张	AI 计算能力、数据中心、能源和云基础设施成为新的物理与资本条件[9][10]	能源/地球系统、经济、国家安全、竞争政策
公共治理开始使用 AI	公共服务、政策分析和行政流程采用 AI 的制度实践增加[20][36]	政治、法学、公共行政、认识论、AI
关系、文化与教育被技术中介	AI 进入学习、表达、陪伴、翻译和文化生成，改变人如何接触知识与彼此	教育、社会人类学、认知、宗教文化、伦理
国家竞争与共同风险同步上升	AI 能力越来越与经济、安全和技术主权联系，同时风险跨境扩散[11][16]	国际关系/战略安全、政治、经济、系统科学、法学

1.3 “所有学科都该看到”不等于“所有学科都要研究同一件事”

计算机科学没有义务替经济学测量收入分配，经济学也不能替教育研究判断长期学习，哲学不能替医学决定治疗事实，文明理论更不能替国际关系判断战略稳定。共同面对的真正含义是：每门学科都要知道自己的结论会继续进入谁的因果链，也要知道别处的变化何时会反过来改变自己的研究对象。

共同面对，不等于共同裁决；融合的是现实问题与因果链，不是各学科的证据标准。

二 | 高概率可能：如果当前能力、成本与组织采纳趋势继续，什么更可能发生

“高概率”在本文中不是统计置信度，也不是时间表预测。它是一种 C 类条件判断：如果当前 AI 能力持续提升、使用成本继续下降、组织采用继续扩展，且没有出现足以逆转这些趋势的重大技术、法律、经济或社会阻力，那么以下方向具有较强现实理由继续发展。每一项都列出可能改变判断的条件。

2.1 知识成本继续下降，而判断的重要性上升

生成、检索、翻译、摘要、代码和初步分析的成本已经显著下降。科学研究也在形成新的 AI 基础设施 [1][19]。如果这一趋势延续，未来更稀缺的可能不再只是“拥有答案”，而是提出值得问的问题、辨别证据强弱、处理冲突目标、识别异常并对结果负责。

这不是“人类判断永远不可替代”的本体宣言。若未来系统在开放环境下稳定完成问题重定义、证据评估、价值冲突处理和后果承接，并且社会愿意让其承担相应资格与责任，这一判断必须重开。当前它只提醒教育和组织：答案变便宜时，不要顺手把形成判断的过程也取消。

2.2 很多职业更可能先被重构，而不是整齐地消失

ILO 2025 年的全球指数估计约四分之一就业处于某种生成式 AI 暴露范围，并明确指出，因大多数职业仍包含需要人类投入的任务，转型比全面冗余更可能 [3]。实验与现场研究也已在部分知识工作中观察到任务级生产率增益 [4][5]。

因此，高概率问题不是简单的“还有没有工作”，而是岗位内部还剩什么：人是否仍经历完整问题、积累经验、做判断、承担责任并从浅层任务走向深角色。岗位存在而成长路径消失，仍然可能构成文明层面的失位。

2.3 组织依赖会比个人偏好更快地改变生活环境

一旦 AI 被写入企业、学校、医院、政府和平台的标准流程，个人是否主动选择使用 AI 就不再是唯一变量。招聘、评价、教育、服务、信贷、公共行政和信息环境都可能被组织层的 AI 采用改变。OECD 与 UNESCO 的公共部门工具包、OECD 政府 AI 研究以及 NIST 风险框架都把 AI 看作完整组织生命周期和治理问题，而非个人工具问题 [17][18][20][36]。

所以未来最重要的决定之一会发生在组织层：节省下来的时间给谁，AI 输出是否成为默认，谁拥有否决与申诉，人的训练和冗余是否保留。

2.4 生产率可能提高，但收益分配不会自动改善

已有实验和现场研究支持某些任务和场景的生产率增益 [4][5]；OECD 的宏观研究同时提醒，微观效率到宏观生产率、工资和福利之间仍要经过扩散、竞争、互补与分配机制 [6]。自动化研究也长期说明，技术既可能替代任务，也可能创造新任务 [7][8]。

因此，“AI 提高生产率”与“人获得更多自由时间”之间不存在自动等号。如果所有权、市场结构和议价关系使收益高度集中，生产率提高也可能与劳动不安全、教育竞争和权力集中同时发生。

2.5 教育的目的会被迫改变

当解释、练习、写作、代码和答案更容易获得，教育若继续主要围绕知识搬运和标准答案组织，将越来越难说明自己的价值。UNESCO 已经要求各国以人本、年龄适切、数据保护和能力发展为中心处理生成式 AI 进入教育 [14]。教育研究也显示，AI 辅导可以帮助学习，但缺少护栏时也可能出现短期表现提高而独立学习受损的情形 [29][30]。

高概率变化因此不是“教师消失”，而是教育必须重新回答：哪些知识仍要亲自掌握，哪些过程必须亲自经历，怎样在 AI 辅助下形成更强的理解、判断、责任、关系和自我成长。

2.6 技术基础设施会进一步成为事实权力基础

前沿 AI 需要算力、芯片、数据中心、能源、网络与云服务。OECD 已把 AI 基础设施竞争单独作为竞争政策问题[10]；IEA 预计数据中心电力需求在基准情景下到 2030 年约达 945 TWh，AI 是主要增长驱动之一[9]。技术能力越成为公共和经济活动的基础，控制这些基础设施的主体就越可能获得超出传统产品供应商的现实影响力。

事实权力不等于合法权力。基础设施重要性不能自动生成政治授权、公共职权或主权，但它会迫使政治学、法学、竞争政策、国际关系和能源系统共同进入。

2.7 人类意义来源会从职业单中心向复合来源重组

现代社会中，职业长期同时承担收入、身份、日常时间、同事关系、成长、贡献和被需要。AI 若持续重组岗位，这一“意义打包结构”会松动。风险是旧中心先变薄而新的承载尚未形成；机会是创造、学习、照护、家庭、社群、公共参与、文化、信仰、探索等路径获得更大现实空间。

这里的“意义来源重组”是 C 类结构判断，而不是对每个人心理状态的预测。若长期研究显示职业结构即使大幅自动化仍能稳定承载身份、成长与被需要，或新的非职业路径难以形成持续角色，本判断应收缩。

2.8 国家竞争会加速，而共同风险也会同步增加

当 AI 具有明显经济、科研与安全价值，国家和企业都有理由争取领先。2026 年《AI Index》已经显示国家间模型能力竞争、AI 主权和基础设施竞争的重要性上升[1]。联合国《全球数字契约》则把 AI 治理明确置于国际合作、人权、发展、能力建设和全球风险框架中[11]。

安全领域存在典型的局部理性—整体风险：一方为避免落后而加速，其他方必须跟随；决策时间越短，自动化和预授权的诱因越强，沟通与纠错窗口可能进一步压缩。International AI Safety Report 2026 也指出，通用 AI 的误用、故障和系统性风险并存，现有保障仍有明显局限[16]。ICRC 针对自主武器的立场进一步强调，需要在人类控制、法律责任与不可接受风险之间设置明确边界[35]。完整战略稳定判断仍需国际关系和战略安全专业独立承担。

三 | 真正开放的未来：哪些重大问题仍不能被提前写成必然

文明方向需要远见，但不能靠把未知写成已知来获得力量。以下问题都可能深刻改变文明，却仍不适合作为今天所有制度选择的必然前提。

真正开放的问题	为什么不能提前结算	今天仍可做什么
统一意义上的 AGI 何时出现	能力定义、评测与系统边界仍在变化，不能把单一日期当作事实	按已出现的能力和现实部署治理，保留前瞻压力测试
AI 是否形成主观体验或主体连续性	当前公开证据不足，相关科学理论本	保持证据开放，同时不让功能角色自

真正开放的问题	为什么不能提前结算	今天仍可做什么
	身也未收敛	动变成主体资格
是否出现可持续的超级智能或失控系统	能力进展快但开放世界可靠性、可控性和安全保障仍存在重大不确定[16]	做好分层安全、独立评估、社会韧性与可停止设计
是否进入极端物质丰裕	生产力、能源、资源、所有权、分配和生态约束都可能成为瓶颈	先把已有生产率收益转化为更广泛生活能力
意识上传、数字永生、后生物文明	涉及尚未解决的意识、身份、生命与物理问题	作为长期研究开放，不用于替代当前有限生命的权利与现实

这一区分还有一个重要后果：文明跃迁不需要等到所有开放问题被解决才开始。只要“AI 能力越来越强、组织越来越依赖、生产和知识条件正在改变”已经成立，人类就必须决定这些变化怎样进入分配、教育、公共权力和生活。等待“真正 AGI”才讨论文明方向，反而会让大量方向选择在组织惯性中先被做完。

很多技术趋势可能难以逆转，但它们怎样进入文明，仍然存在选择。

四 | 机遇：从更强工具到重新安排人类生活条件

4.1 第一种机会：降低维持基本生活所需的人类必要劳动

自动化最深的文明意义，不是让企业少雇多少人，而是人类维持物质生活所需的劳动时间和认知劳动是否可以持续下降。若 AI 与自动化提高生产率、降低教育与公共服务成本，并且这些收益能够被广泛获得，人类可能第一次在更大规模上减弱“必须持续出售大量时间才能获得基本生存”的结构压力。

这并不等于工作会消失，也不等于劳动不再有意义。它意味着工作可以从“生存强制”中部分松绑，人能够更有条件选择值得长期投入的劳动、照护、创造和公共参与。

4.2 第二种机会：让过去属于大机构和少数专家的能力更普及

生成式 AI 和通用工具可以降低翻译、代码、设计、分析、资料组织和专业沟通的门槛。UNDP 2025 强调，应把 AI 发展重心放在人的能力、选择与互补上[2]。如果这种能力扩展不只属于最富机构，而能进入小企业、学校、社区、发展中地区与普通个体，技术进步就可能扩大“谁能够做复杂事情”的社会边界。

4.3 第三种机会：让科学、医疗和地球问题获得更强的处理能力

AI 已经进入药物、蛋白、天气、材料、天文和生物研究，并在部分科学任务上扩大搜索和模拟空间[1][19]。WHO 也将多模态模型视为可能进入临床、患者服务、教育、科研和药物开发的通用能力，同时强调错误、偏差、自动化偏误和治理风险[15]。

如果这些能力被用于疾病、气候、能源、农业、灾害和基础科学，人类可能更快解决过去受限于信息和计算复杂度的问题。这种机会并不只属于技术产业，它可能直接改变健康寿命、环境承载和全球发展。

4.4 第四种机会：教育终于可以从“知识稀缺”更多转向“人的成长”

当高质量解释和练习越来越便宜，教育可以把更多稀缺资源投入人更难被标准化替代的过程：提出问题、深度理解、实践、关系、合作、身体经验、伦理判断、创造和公共参与。Vygotsky 与情境学习传统长期说明，能力是在社会互动与逐步参与中形成的[27][28]；AI 可以扩展这一过程，但不能把“给答案”误作“人成长”。

4.5 第五种机会：一个人的能力和关心可以抵达更多真实的人

AI 可以跨越语言、时间和专业门槛，把一个教师的知识、一位医生的健康教育、一名研究者的解释、一位普通人的关心放大到更多人。这里最值得追求的不是“模拟一个人替他存在”，而是让血肉之人的能力、经验与善意在明确授权和可追溯责任下延展。

4.6 第六种机会：人类被迫重新回答“发展究竟为了什么”

当“做得更快、更便宜、产出更多”越来越容易，效率本身就越来越不足以作为文明的终极回答。一个只会继续扩大生产和优化，却不知道为什么要把人的时间释放出来、为什么要保护关系和选择、为什么值得让更多人进入公共生活的文明，可能拥有强系统，却缺少方向。

AGI 最大的机会，也许不是替人完成世界，而是逼迫人类第一次认真决定：当生存与生产不再占满全部能力以后，我们究竟想把文明用来做什么。

五 | 什么不能自动发生：技术进步不会自己生成好文明

技术乐观主义最容易犯的错误，不是高估某一次模型分数，而是把“能力可能”直接写成“文明结果”。文明方向稿必须把中间的转换说清。

技术变化	常见自动推断	真正还需要发生的转换
AI 能力更强	人人自然更强	可达性、教育、组织设计与人的能力留存
生产率提高	人人自然更自由	所有权、收益分配、工时、基本保障与议价结构
自动化增加	人自然拥有更多时间	节省时间是否还给人，是否被新的绩效要求重新占满
信息更丰富	社会自然更理性	证据判断、媒体生态、教育与公共讨论能力
连接更方便	关系自然更好	真实互惠、承责、在场、退出与不被操纵
预测更准确	治理自然更正当	授权、解释、申诉、责任、权利与民

技术变化	常见自动推断	真正还需要发生的转换
		主程序
系统更稳定	文明自然更好	主体位置、基本权利、可退出、方向是否值得
AI 更聪明	AI 自然更有权决定	能力与法律/政治资格必须分别证成

因此，文明跃迁的第一原则不是“相信技术”，也不是“阻止技术”，而是拒绝自动转换。每一项技术红利都必须追问：谁真正获得了它，人的能力是否同步增长，权力是否静默转移，退出是否仍然存在，生活是否真的变得更好。

OECD 2024 更新的 AI 原则把包容增长、人的能力增强、人权、民主价值、安全、可停止和问责同时纳入人工智能政策[12]；UNESCO 2021 建议同样把人权、尊严、人类监督、环境、教育和文化作为共同边界[13]；欧盟《人工智能法》则以风险分级、特定义务与高风险系统要求建立具有法律约束力的治理框架[37]。这些国际框架并不等于本文的文明目标，但说明现实治理已经不能只用“创新速度”作为单一尺度。

六 | 方向：存续底线与生存无忧、生活幸福、生命更有意义

6.1 存续是不可失守的底线，但不是文明终点

任何文明方向首先都必须保住生命与文明继续存在的条件。AI 军事化、生物风险、关键基础设施失控、生态与能源约束等高后果风险不能因为长期愿景美好而被降低优先级。International AI Safety Report 2026 把误用、故障与系统性风险并列，并指出现有技术保障仍不完全可靠[16]。

但存续只能回答“还能不能继续”。如果一个系统极其稳定，却让绝大多数人长期处于匮乏、被动、失位和无选择之中，它仍不是本文希望抵达的完整文明。

6.2 第一目标：生存无忧

生存无忧不是“所有人永远不遇到困难”，而是基本食物、住所、基础健康、教育、安全和继续生活的条件，越来越少因为一次失业、疾病、技术替代或出身而整体崩塌。它特别意味着：人的基本生存不应长期以“先证明自己足够市场价值”为唯一前提。

AGI 时代之所以让这一目标更现实，是因为生产、知识和服务的边际成本可能下降；之所以不自动实现，是因为生产率收益、所有权、公共财政和资源分配仍由制度决定。

6.3 第二目标：生活幸福

活下来以后，人还需要真实生活。本文所称生活幸福不是一个必须被统一测量的心理分数，而是人拥有更多自主时间、健康、安全、关系、自由、乐趣、文化、自然、选择和退出能力。幸福不能用高消费或高平台活跃度替代。

自由时间也不是自动的幸福。如果自动化释放的时间只是被新的注意力平台、绩效要求或孤立占满，文明并没有真正把能力还给人。真正的方向是让人有时间也有现实条件去爱、照护、学习、玩、创造、休息、参与公共生活和与他人共同生活。

6.4 第三目标：生命更有意义

“生命更有意义”不是文明给每个人分配一个统一答案。意义来源可以不同：关系、创造、照护、知识、责任、公共参与、信仰、探索、经历和存在本身都可能生成意义。文明能做的不是规定你为什么活，而是扩大你能够真正进入这些意义来源的条件。

文明跃迁不是给人分配意义，而是让更多人拥有生成意义的条件。

6.5 三目标递进，但不能互相兑换

生存无忧→生活幸福→生命更有意义是方向递进，不是三项可以互相结算的 KPI。不能因为总体安全提高就持续牺牲某些人的基本权利；不能因为平均幸福上升就取消人的选择；也不能以“更有意义”为名要求个人接受匮乏和强制。

三目标之所以在 AGI 时代重新具有现实重量，是因为技术正在同时松动生存、劳动、知识和时间条件。过去很多社会必须先把大量文明能力投入维持基本生产；未来如果必要劳动和知识稀缺下降，人类就有理由把文明目标继续向生活和意义展开。

七 | 人的位置：文明跃迁不是让人“还有事做”，而是让人有真实生活与多种意义来源

7.1 “给人找任务”不是文明答案

面对自动化，最直接的反应往往是“未来怎样让人还有工作”。工作当然重要，但如果文明只是人为制造低价值任务来证明人仍有用途，就把人的价值重新锁回“是否能被系统使用”。真正的问题不是系统还用不用得上人，而是人是否仍有真实判断、责任、成长、关系、创造和共同方向的位置。

7.2 结构意义：意义进入现实生活的中层接口

本文沿用《意义动力学》的中层含义使用“结构意义”：关注能力、需求和连接怎样在关系、路径、角色、贡献、参与和制度中形成可持续行动关联。它不回答意义本体，也不把私人体验压缩成结构变量。其作用只是让我们看见：一个人的人生意义虽然不能被制度定义，但文明可以扩大或压缩他进入真实关系、承担责任、贡献、成长和被需要的位置。

7.3 从职业单中心走向复合意义来源

职业在现代文明中承担了过多功能：收入、身份、社交、时间、成长、贡献和价值确认常被打包在一个岗位里。因此自动化冲击职业时，问题可能同时表现为经济压力、身份失稳、关系减少和意义来源收缩。未来更稳健的方向不是让“工作”消失，而是让一个人的意义世界不再只能依赖一条单一职业路径。

意义来源	可能的现实入口	AGI 时代需要避免的压缩
关系与爱	家庭、伴侣、朋友、共同体	把关系完全平台化、把陪伴替代成无责任模拟
创造	艺术、设计、写作、发明、手工、创业	只保留末端修饰，让人失去形成作品和判断的过程
知识与成长	学习、研究、师承、探索	答案替代理解，系统替代长期能力形成
照护与被需要	育儿、养老、医疗、邻里支持、志愿服务	把照护只写成低效率成本，忽视关系与责任
劳动与贡献	职业、项目、维护、服务、生产	岗位切成浅任务，贡献无法沉淀为角色与信用
公共参与	公民讨论、社区、治理、公共服务	公共决定被不透明系统和少数接口垄断
信仰与精神生活	宗教、传统、纪念、终极关怀	让 AI 功能角色冒充传统代表、圣职或逝者本人
探索与存在	旅行、自然、身体经验、游戏、未知	把自由时间重新占满为消费与算法注意力

结构意义来源越多元，一个人的某一条路径失稳时，整个人生越不容易同时坍塌。但多元不等于路径越多越好；如果没有时间、资源、关系和制度承载，大量“选择”只会变成新的过载。文明真正要扩大的是可进入、可持续、能够形成责任与成长的路径。

八 | 共同问题：不同学科怎样一起面对文明级事件，而不取消专业边界

AGI 时代的文明级事件之所以重要，是因为它们同时改变多个学科的对象。各学科不是为了证明“文明跃迁理论正确”而进入，而是因为现实问题已经越过了它们过去可以独立处理的边界。

共同事件	最先能够判断的学科	必须重新进入的学科	最终不能遗漏的文明问题
AI 重组劳动	AI、经济、组织研究	教育、社会、法政、哲学	生产率提高后，基本生活、身份、成长与被需要怎样重组？
AI 进入教育	教育、认知、AI	伦理法学、社会文化、经济、政治	答案更便宜后，人还怎样形成判断、关系、公民能力与自我？

共同事件	最先能够判断的学科	必须重新进入的学科	最终不能遗漏的文明问题
AI 进入医疗与生命决定	医学/生命科学、AI	伦理法学、经济、社会、哲学	准确率提高后，谁决定、谁负责、谁被照护？
AI 进入政府	政治、法学、公共行政、AI	认识论、经济、社会、系统科学	预测更准是否形成事实权力？公民还能否理解、质疑和申诉？
AI 进入关系文化	社会人类学、认知、AI	宗教、伦理法学、经济、哲学	可达性提高是否同时保留真实关系、代表边界与文化多样性？
AI 军事化	AI、战略安全/国际关系	系统科学、伦理法学、认识论、经济	局部安全选择是否聚合成更短的纠错窗口和共同存续风险？
算力与资源扩张	能源/地球系统、工程、经济	政治、国际关系、社会	AI 增长怎样进入电力、材料、水、空间和地球承载边界？

8.1 一个领域的好指标，进入另一个领域可能不是好指标

效率在计算机科学和企业管理中可能是重要指标，在教育中却不能替代理解与成长，在关系和哀悼中甚至不是主要目标。准确率在医疗中重要，但不能独自决定患者权利和责任。预测在治理中价值，但不能替代合法性和申诉。跨学科升维的核心不是让所有学科采用同一指标，而是知道本学科的“好”进入别处以后还要继续接受什么判断。

8.2 聚合比简单“融合”更重要

文明风险常常不来自一项应用本身，而来自许多局部合理选择同时采用同类基础设施和同类优化逻辑。AI 教育、AI 医疗、AI 办公、AI 政府和 AI 科研都可能分别提高效率；如果它们同时依赖少数模型、云平台和数据通道，聚合后却会出现共同依赖、相关性失败、事实权力和退出困难。这是系统科学、经济、政治与工程必须共同研究的新对象。

类似的聚合还可能出现在文化与模型之间。实验研究在特定创作任务中观察到，生成式 AI 可以提高个体创作表现，却可能降低群体内容多样性[21]；另有研究表明，在特定递归训练条件下，模型持续从模型生成数据学习可能发生分布退化甚至模型坍缩[22]。两者并不能直接证明现实世界已经形成一个完整闭环，但足以提出一个可反驳的 C 类问题：如果人越来越从相似模型获得表达，而模型又越来越从这些表达学习，文化与认知的可探索空间会不会共同收窄？

学科分别求真，文明共同承担后果。

九 | 方法一：从现实问题和真实受影响者出发

9.1 不从完整蓝图开始

文明跃迁不是先画出一个完美社会，然后寻找现实来证明它。真正可行的方法是从已经发生的问题开始：某类劳动被重组、某个学校出现学习依赖、某个公共系统出现算法事实权力、某个社区缺少照护、某个资源瓶颈限制 AI 扩张。问题越具体，受影响者越真实，专业判断和后果越可检验。

9.2 真实受影响者不是“用户画像”

谁承担风险、谁得到收益、谁失去路径、谁无法退出，不能只由设计者想象。WHO 对医疗 AI 和 UNESCO 对教育 AI 都强调利益相关者参与、具体使用情境与人的监督[14][15]。文明方法要把患者、学生、教师、劳动者、家庭、弱势群体、社区和未来世代从抽象“受益者”变成可以改变方案的人。

9.3 先解决最真实的困难，再讨论是否具有普遍性

一个机制在一座城市、一个学校、一个组织里有效，不等于它已经成为普遍文明答案。反过来，小规模失败也不必否定方向本身。现实试验的功能是把抽象因果变成可观察后果：谁实际受益，哪些成本被遗漏，什么权利冲突出现，是否产生新的依赖，退出是否真的可行。

十 | 方法二：新旧并行、比较验证、允许失败与退出

10.1 新文明不是“宣布旧世界结束”

旧制度承载着权利、生活、信用、资产、家庭责任和社会连续性。即使新机制在理论上更好，也不能仅凭愿景就要求所有人立即迁移。更合理的路径是让新旧机制在可承受范围内并行，允许人比较真实效果，并让更好的现实逐步获得信任。

10.2 “先证明更好”比“先证明正确”更重要

文明方向无法在纸面上被完全证明。一个新的资源、教育、协作或公共机制，真正需要证明的是：它在相同或更低成本下是否让更多人获得更好的生活条件，是否减少了强制和不可逆损害，是否保留更多选择和修正空间。

10.3 失败权是文明创新的基础设施

如果每一次试验失败都意味着永久资格损失、巨大生活损失或无法恢复的身份惩罚，人们就不会真正尝试新路径。文明跃迁要降低可控失败的代价，让人和组织有机会退出、重来、修正和回到旧路径。

10.4 可退出不是产品按钮，而是现实能力

真正退出意味着：停止使用后仍有替代服务，数据和身份不被永久锁定，基本生活和权利不因退出受损，组织仍保留人工接管和备份能力。OECD AI 原则强调在可能造成过度伤害或出现不良行为时，应有覆

盖、修复和安全停用机制[12]；NIST 风险框架也强调生命周期治理[17][18]。文明层面还要继续问：个人、组织和国家是否真正有替代能力，而不是名义上允许退出。

十一 | 方法三：AI 扩展人的能力，保留判断、责任、成长与方向

11.1 人在环中，不等于人在主体位置

一个系统可以保留“人工确认”按钮，却把真正的问题定义、证据组织、方案比较和责任判断全部提前交给 AI。此时人仍在流程里，但只剩末端签字。自动化研究早就说明，过度依赖会带来技能衰退、监督负担和接管困难[33][34]。

11.2 判断能力需要在真实判断中成长

人类的判断、创造和责任并不是储存在脑中的静态资产，而是在长期处理真实问题、犯错、解释、承担后果和与他人协作中形成。教育和组织若为了效率取消这些过程，即使短期结果更好，也可能减少未来厚主体的再生产。

11.3 AI 增强的最低方向

要保留的人的位置	AI 适合承担什么	不能静默发生什么
理解	整理信息、解释不同材料、提供比较	让系统摘要变成唯一事实入口
判断	提供候选方案、风险和反例	把默认推荐变成事实决定
创造	扩展素材、模拟、原型与表达能力	把人长期留在后修饰接口
责任	记录过程、提示风险、支持追踪	让责任在“模型/平台/组织”之间漂浮
成长	个性化反馈、练习、导师辅助	取消困难、失败、带教和独立能力形成
关系	翻译、提醒、能力延展	用无责任模拟取代真实互惠与在场
方向	提供推演、方案比较、后果可视化	让能力优势自动获得最终决定权

11.4 组织激励必须把长期主体能力算进成本

现实组织往往更容易衡量 AI 节省了多少工时、成本和交付时间，却较难衡量失去了多少带教、判断训练、经验冗余和长期责任能力。如果这些慢变量长期不进入组织决策，主体替代会拥有结构性的短期优势。文明方向并不是禁止自动化，而是要求组织把未来接管能力、学习、责任和人的成长也当作真实资产。

十二 | 方法四：按阶段改变重点，不把紧急安排永久化

12.1 启动期：先避免失去未来和纠错窗口

当 AI 能力快速扩展、战略竞争加速而制度仍不成熟时，启动期的首要任务是保住共同存续、基本权利和继续试验的空间。复杂系统安全研究长期提醒，高后果事故可能来自部件之间、组织流程与控制结构的相互作用，而不是一个单点错误[31][32]。因此，高后果系统需要更强的监测、责任、停止与接管；社会需要避免因一次技术浪潮造成大规模不可逆失位；新机制要尽快证明能否产生真实好处。

12.2 过渡期：让新旧结构形成可比较的现实

过渡期的重要任务不是扩大临时权力，而是让新机制从少数试验进入可复现、可比较、可申诉的制度环境。此时必须开始清理启动期的例外安排，防止“为了安全和速度”的临时权限沉淀为永久权力。

12.3 稳定期：把权利、退出、平衡与生活质量放回中心

当基本存续与新运行条件已经较稳定，文明不应继续用启动期的紧急逻辑治理日常生活。稳定期更强调常态化权利边界、权力平衡、长期治理、程序、可逆性和生活质量。技术效率越高，越要避免把所有生活重新纳入统一优化。

12.4 自我维持不是“系统自动运行”，而是社会具备持续修正能力

长期成熟的文明不需要一个永远正确的中央设计者，而需要让专业研究、公共讨论、制度纠错、工程维护和真实生活经验不断进入。Ostrom 关于多中心治理的研究说明，复杂公共事务并不只有单一中央控制这一种制度想象[24]；这不能直接证明未来文明应采用某一治理结构，却支持把多中心试验、比较与纠错保留为开放方向。自我维持的核心不是锁死制度，而是让制度有能力发现自己错了并改变。

十三 | 目标实现：怎样观察“生存无忧”正在成为现实

生存无忧不能被压缩成一条收入线，也不能用一个福利制度名称替代。它是一组现实方向：基本生活越来越少因市场波动、技术替代、疾病、家庭背景和一次失败而整体崩塌；人拥有足够的选择，不必为了维持生存接受任何工作、任何平台条件或任何技术依赖。

观察方向	正在接近目标的迹象	需要警惕的假实现
基本资源可达	食物、住所、基础医疗、教育与基本数字接入更加稳定	名义有保障但申请复杂、惩罚性强、无法实际获得
就业冲击韧性	技术替代后仍能维持生活、学习和再进入其他路径	用短期补贴掩盖长期失位与资格锁定
基本权利不随评分变化	退出平台、拒绝某类 AI 使用不影响基本生存与公民权利	把贡献、信用或行为评分变成基本资源门槛
失败可恢复	一次创业、转型、学习失败不会永久	失败导致长期债务、身份和路径封锁

观察方向	正在接近目标的迹象	需要警惕的假实现
	掉出社会结构	
地域与群体差距缩小	AI 能力和基本服务向边缘地区与弱势群体扩展	平均指标提高但技术红利只进入少数中心

“生存无忧”不是取消个人努力、市场和差异，也不是承诺资源无限。它只改变一条文明底线：基本生存不应长期建立在“必须先证明自己足够有用”之上。

十四 | 目标实现：怎样观察“生活幸福”正在成为现实

生活幸福比生存复杂得多，不能被一个统一满意度替代。Sen 的能力方法提醒，发展不仅看资源和产出，也要看人实际能够做什么、成为什么[23]；自我决定理论长期强调自主、胜任和关系与人的动机和福祉相关[25]；社会关系研究也持续显示人与人连接对健康和长期福祉的重要性[26]。这些研究不能直接证明本文的文明目标，但说明生活质量不能只用 GDP 或效率理解。

生活维度	方向性观察问题
自主时间	必要劳动和行政负担下降后，人是否实际拥有更多可自己决定的时间？
健康与安全	技术是否提高预防、诊疗和照护可达，同时不增加新的不可解释风险？
真实关系	人是否有更多时间和空间建立家庭、友谊、社区与长期互助，而非只有平台互动？
选择与退出	人是否可以拒绝某种工作、平台、算法推荐和组织安排，而不承担不可承受代价？
文化与乐趣	城市、社区、教育和数字空间是否给游戏、艺术、自然、运动和非功利生活留出位置？
不被持续优化	人的生活是否允许慢、闲、偶然、失败、私人空间和不产生指标的存在？

AGI 可能提高生活幸福，也可能把效率逻辑进一步深入家庭和私人生活。文明方向的判断标准不是“技术使用越多越好”，而是技术最终是否扩大了人的真实选择与生活丰富度。

十五 | 目标实现：怎样观察“生命更有意义”的条件正在扩大

15.1 我们不测量“你有多少意义”

意义最容易被治理体系误用。一旦“有意义”变成统一评分，文明就会开始替个人规定什么生活值得。本文拒绝这种方向。能够被观察的不是一个人“意义值多少”，而是文明是否扩大了生成意义的条件。

我们不测量“你有多少意义”，而是观察文明是否正在扩大人生成意义的条件。

15.2 意义条件的方向性观察

条件	可以观察的现实变化	不能被误写成
真实关系	照护、友谊、家庭与共同体有时间、有空间、有互惠	连接数量、消息频率
被需要与贡献	维护、照护、创造、桥接、公共服务等不同贡献有真实路径	统一贡献总分
成长	人能从浅任务走向深角色，持续积累判断、责任与能力	证书数量、平台等级
创造与探索	有低门槛进入艺术、科学、项目、自然与未知的机会	内容产量、流量
公共参与	更多人能理解、质疑并影响共同生活	被动点赞、形式投票次数
精神与终极关怀	不同宗教、传统与个人答案有自主空间和真实共同体承载	AI 替个人决定信仰或模拟神圣权威
未预定义的人生空间	人仍可改变路径、拒绝既定最优、经历偶然和重新开始	算法为每个人预先规划“最佳人生”

15.3 意义来源必须文化化，而不能全球统一

东亚社会可能更重关系与责任，某些宗教共同体把日常生活与超越性秩序紧密连接，现代自由主义社会可能更强调个人选择和公民权利，地方共同体可能依赖非正式互助和口述传统。结构意义的共同问题可以比较，但它的承载形式必然文化化。

因此，文明跃迁不能把某一种家庭、职业、宗教或公共参与模式变成人类统一模板。真正共同的方向只是：让不同的人有更多真实、可持续、非强制的路径去形成属于自己的关系、责任、创造、成长与回答。

十六 | 从方向到生活世界：愿景、制度、工程与现实系统各自承担什么

16.1 思想引导回答“为什么”和“往哪里”

本文的主要责任到这里已经完成：说明为什么 AGI 打开文明选择窗口、为什么技术红利不自动成为人的红利、为什么三目标构成值得争论和检验的方向、以及怎样以开放方法逐步接近。思想不能被工程替代。

16.2 《活的文明》负责让读者看见“生活在其中是什么样”

学术论证能够说明方向和因果，却很难仅靠概念让人判断那个未来是否值得生活。《活的文明》三部曲承担的是未来世界的生活显影：工作、家庭、教育、关系、社区、公共生活和技术怎样被重新组织。它是愿景与续读入口，不是本文经验事实的证据。

16.3 星火、星云、AGI-COS、Human-COS 等只作为候选承载

制度与工程系统的任务是回答“怎样承载”。本文所称星火，是个体能力与关心延展的候选系统；星云，是文明协作与运行的候选承载；AGI-COS，是机器侧能力、权限、审计、停止与退出的工程治理候选框架；Human-COS，是高后果选择在不可逆之前进行多主体、多视角后果推演与判断支持的候选系统。它们仍处于研究与验证轨道：可以承载、试验、失败和退出，不能因为属于文明跃迁体系就自动获得正确性，更不能反向成为本文方向判断的证据。

16.4 从方向到工程的因果顺序不能倒置

思想定方向，学科求真，制度定边界，工程承载现实；现实反过来修正一切。

如果一个机制需要先让读者学习大量内部编号才能理解它为什么存在，通常说明表达顺序出了问题。能用“让人可以退出”“让责任落回真实主体”“让新的生活先证明更好”说清，就不先用复杂协议替代思想。

十七 | 现实如何否定我们：失败、反例与文明自我修正

一篇文明方向文最危险的状态，是把“值得追求”写成“因此一定有效”。本文必须给出明确的反向修改入口。

本文判断	什么事实会迫使本文修改
AI 长期可能压缩人的判断与成长	长期、多场景研究稳定显示 AI 撤除后人的独立能力、接管能力和责任能力普遍增强，则应下调“主体变薄”风险判断
生产率红利不会自动转化为自由	若不同制度下 AI 生产率提高普遍、持续带来工时下降、收入安全提高且分配差异不扩大，应缩小分配中介的担忧
职业单中心会松动	若大规模自动化后职业仍稳定承载身份、关系、成长和意义，

本文判断	什么事实会迫使本文修改
	且非职业路径不形成持续承载，应调整意义来源重组判断
基础设施集中会形成系统性依赖	若开放、竞争、互操作与公共基础设施长期维持真正可替代性和独立退出，聚合依赖判断应收缩
三目标具有较强跨文化吸引力	若不同文明、群体和长期公共讨论反复显示三目标遗漏了不可替代的文明目标，作者应重开规范框架，而不是把异议解释为“尚未理解理论”
新旧并行与可退出是较优方法	若现实反复显示某些高后果条件下并行显著增加不可逆风险，或退出机制制造更大伤害，应按领域和阶段重新设定方法

17.1 自我修正不是第四目标

存续、生存无忧、生活幸福、生命更有意义回答文明想达到什么状态；自我修正回答文明怎样不被自己的旧答案困住。它是运行能力，不应再被机械增加为第四目标。

17.2 方向也必须允许被不同文明经验改变

规范承诺不能像实验假说那样被单次数据“证伪”，但它可以被更强论证、长期后果、受影响者经验和跨文化反例迫使修订。如果一种所谓“有意义文明”不断要求人接受强制、如果一种“幸福治理”持续削弱自由，如果一种“生存安全”永久取消退出，那么失败的不是公众，而是我们的方向表达或制度承载。

17.3 最有价值的参与不是赞同

最有价值的参与，不是赞同本文，而是足以改变本文的专业工作。

研究者可以用长期数据检验能力增强与主体变薄；经济学者可以比较不同分配制度；教育者可以测量 AI 撤除后的独立能力；社会学与人类学可以提供不同文明的意义承载反例；战略研究者可以检验 AI 竞速与危机稳定；生态与能源研究可以重新计算 AI 扩张的物理边界；工程师可以证明某些可退出和冗余设计是否真正可行。

结语 | 技术打开可能，思想决定方向，现实检验道路

AGI 时代真正改变人类的，不只是机器会不会更聪明，而是越来越多过去必须依赖人的知识、生产、分析、组织和执行能力可以被技术外延。这样的能力增长会自然进入企业、学校、医院、政府、家庭与国家竞争，不需要等某一天全世界共同宣布“AGI 已经到来”。

这也是为什么文明方向不能推迟。技术和组织惯性本身就会做选择：把效率放在哪里，把成本算给谁，把判断交给谁，把时间还还不还给人，把基础设施掌握在谁手里，把教育培养成什么，把关系变成什么。如果人类不主动回答，很多文明方向仍然会被选择，只是选择者变成市场激励、平台默认、组织 KPI 和竞争压力。

本文选择公开承担另一种方向：先守住存续，在技术能力能够降低匮乏时推动生存无忧；在生存被更好承载后，把文明能力继续用于真实、自由、丰富、有关系、有乐趣的生活幸福；再进一步，不替人规定意义，而扩大关系、创造、照护、学习、责任、公共参与、信仰、探索和存在本身能够真实进入人生的条件。

这不是对未来的预测，也不是要求所有文明接受同一种制度。它只是提出一个可以被事实、专业、不同文化与现实失败持续改变的方向判断。

AGI 最重要的意义，不是机器最终能有多像人，而是当机器可以承担越来越多过去必须由人承担的事情以后，人类终于必须重新回答：我们希望文明为什么而存在。

技术打开可能，思想决定方向，现实检验道路。

附录 A | 文明级事件—学科—方向矩阵

文明级事件	共同进入的主要学科	若不继续追踪可能遗漏什么	文明方向问题
智能能力普及	AI、经济、教育、组织、社会	能力可达与能力集中并存	更多人真的变强了吗？
劳动重组	经济、组织、教育、社会、法政、哲学	只看就业数量，遗漏岗位内部成长与被需要	生产率能否变成时间、安全和选择？
知识生产 AI 化	认识论、科学、AI、教育、政治经济	只看产出，遗漏求真、基础设施和方向权	知识更快以后，谁还能独立判断？
教育 AI 化	教育、认知、AI、伦理法学、社会政治	只看成绩，遗漏关系、主体成长、公民能力	答案更便宜以后，人应该成长为什么？
医疗 AI 化	医学、生命科学、AI、伦理法学、社会经济	只看准确率，遗漏生命决定、照护和责任	能力提高后，谁决定、谁负责、谁被照护？
公共治理 AI 化	政治、法学、公共行政、AI、认识论	只看行政效率，遗漏合法性、申诉与事实权力	政府更会预测后，公民还有多少理解和质疑能力？
关系与文化 AI 化	社会人类学、认知、宗教文化、伦理、AI	只看互动和满意，遗漏真实关系、代表资格和文化多样性	技术应放大关系还是替代关系？
AI 军事化	战略安全、国际关系、AI、系统、伦理法学	只看单方能力，遗漏竞速、误判和共同存续	国家局部安全怎样避免聚合成共同不安全？
算力/能源扩张	能源、地球系统、经济、工程、政治	只看模型能力，遗漏物理承载和资源分布	智能扩张怎样不越过地球与社会承载？

这个矩阵不构成新的学科目录。此前的联合现实压力反查已经确认三类事实责任必须由独立专业重新进入，当前专业求真层因此由原十三篇扩展为十六篇：国际关系／战略安全承担威慑、危机稳定与战略互动；生态／地球系统／资源环境承担能源、水、材料、土地与物质承载；医学／公共卫生／健康系统承担临床有

效性、患者安全与人群健康。“第一／第二／第三压力”只保留为现实不可逆压力顺序，不代表学科价值或文稿质量，也不再承担增篇决策功能。

附录 B | 已经可见／高概率延续／真正开放：未来判断检查表

层次	写作规则	典型例子	错误写法
已经可见	必须有时间、对象和来源，可直接被外部复核	组织 AI 采用扩大、就业任务暴露、数据中心用电增长	“所有行业已经被 AGI 重构”
高概率延续	明确条件与反证入口，通常属于 C 类结构判断	工作重构、组织依赖、教育目的变化、意义来源重组	“大规模失业必然发生”
真正开放	明确承认不知道，不作为当前制度必然前提	AGI 日期、机器意识、超级智能、极端丰裕、上传意识	“超级智能必将实现，所以今天必须……”

外发前最低检查：如果删去所有关于超级智能、机器意识和具体 AGI 年份的假设，本文的主要方向论证仍应成立。若不成立，说明文章把文明选择错误地建立在开放未来上。

续读与共同参与 | 不是为了读完理论，而是让理论继续被现实改变

本文已经回答到这里：为什么 AGI 时代出现文明选择窗口；哪些现实与高概率变化值得今天进入；为什么技术机会不会自动成为人的红利；本文选择怎样的三目标方向；怎样用开放、可试、可退、可改的方法逐步接近；以及什么事实能够迫使本文重开。接下来不按文件顺序堆叠，而按问题继续。

向上 | 如果继续追问：为什么这些局部变化最终成为文明问题

进入《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》V2.4 及文明跃迁五卷书。前者处理学科怎样重新接上因果链，后者进一步展开文明为什么需要改变、制度和人的位置怎样重新理解。

五卷书公开入口：<https://www.civitas.top/documents/p02-five-001/>[文明跃迁五卷书](#)

向下 | 如果关心：这些方向怎样进入可验证的制度、工程和现实系统

可继续进入第 16.3 节所说明的星火、星云、AGI-COS、Human-COS 及其它制度／实验文稿。它们是候选承载，不是本文方向的证明；真正价值在于让方向接受实现、失败、退出和现实后果的检验。

[星火](#)

[星云](#)

[AGI-COS](#)

[文明跃迁文库](#)

向左 | 如果继续追问：现实到底怎样传导到这些文明判断

回到第一篇《AGI时代的人工智能应用前景与风险研究——从能力扩展到文明后果》V1.2，并进入相应专业深稿。第一篇负责现实应用、组织中介、跨域反馈与聚合后果；十六篇专业稿负责各自事实、理论、证据止步与竞争解释。本文不能替它们求真。

向右 | 如果不同意本文：最值得进入的相反方向是什么

可以直接从以下问题反驳：AI是否会比本文预期更稳定地提高人的独立能力？生产率红利是否可能在现有制度下自然扩散？职业是否仍将长期保持意义中心？基础设施集中是否可以通过竞争和开源充分化解？三目标是否遗漏了更重要的文明目标？“可退出、新旧并行”是否在某些高后果领域反而增加风险？任何能提高这些问题事实质量的反例，都比概念上的赞同更重要。

向前 | 从阅读进入研究、制度、工程与共同创造

研究者可以建立长期跨领域数据；教育者可以比较 AI 撤除后的独立能力；经济和组织研究者可以追踪节省时间和收益究竟流向哪里；战略安全研究者可以建立 AI 竞速与危机稳定模型；生态和能源研究者可以校准算力扩张的真实边界；工程师可以验证可退出、接管、冗余和可审计是否真能在高能力系统中成立；公众和受影响者可以提供理论没有看见的生活后果。

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。

未来链 | 如果希望看见：这些原则进入真实生活以后，人类可能生活在怎样的世界

进入《活的文明》三部曲。它把“可以论证”的方向重新还原为工作、家庭、教育、关系、城市和公共生活中可以想象的世界；愿景不是证据，读者也可以拒绝其中的未来并提出替代文明图景。

《活的文明》三部曲：[未来文明生活图景入口](#)

《意义动力学》：[结构意义与意义来源重组的中层理论入口](#)

文明跃迁五卷书：<https://www.civitas.top/documents/p02-five-001/> | 公众总入口：<https://www.civitas.top/documents/p01-entry-006/> | 完整文库：<https://www.civitas.top/library/>

生成式人工智能使用声明

本文在公开资料检索线索、既有文稿版本比对、结构整理、部分文字起草、参考文献核对、表格组织与版式生成中使用了生成式人工智能辅助。生成式人工智能输出不作为经验数据、权威来源、独立证据或同行评审意见。研究问题、文明方向、A/B/C/F 论证身份、材料取舍、事实与引用核验、规范判断、体系关系和最终文本均由作者审定；作者对全部内容及其学术责任负责。人工智能不列为作者。V1.1 在 V1.0 基础上完成跨稿一致性受控修订：同步十六篇专业责任状态、两篇合流稿版本、AI 主体边界用语、候选承载最小释义与稳定公开入口，不改变三层未来判断、三目标、四方法和目标观察主轴。外部动态事实重点复核至 2026 年 8 月 26 日。正式投稿或出版时，应按目标期刊、出版社或机构要求补充工具名称、版本、日期与使用范围。

参考文献

- [1] Stanford Institute for Human-Centered Artificial Intelligence. (2026). AI Index Report 2026. Stanford University. <https://hai.stanford.edu/ai-index/2026-ai-index-report>
- [2] United Nations Development Programme. (2025). Human Development Report 2025: A Matter of Choice—People and Possibilities in the Age of AI. UNDP. <https://hdr.undp.org/content/human-development-report-2025>
- [3] Gmyrek, P., Berg, J., Kamiński, K., et al. (2025). Generative AI and Jobs: A Refined Global Index of Occupational Exposure. ILO Working Paper 140. <https://doi.org/10.54394/HETP0387>
- [4] Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at Work. Quarterly Journal of Economics, 140(2), 889–942. <https://doi.org/10.1093/qje/qjae044>
- [5] Noy, S., & Zhang, W. (2023). Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. Science, 381(6654), 187–192. <https://doi.org/10.1126/science.adh2586>
- [6] Filippucci, F., Gal, P., Jona-Lasinio, C., Leandro, A., & Nicoletti, G. (2024). The Impact of Artificial Intelligence on Productivity, Distribution and Growth. OECD Artificial Intelligence Papers, No. 15. <https://doi.org/10.1787/8d900037-en>
- [7] Acemoglu, D., & Restrepo, P. (2019). Automation and New Tasks: How Technology Displaces and Reinstates Labor. Journal of Economic Perspectives, 33(2), 3–30. <https://doi.org/10.1257/jep.33.2.3>
- [8] Autor, D. H. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. Journal of Economic Perspectives, 29(3), 3–30. <https://doi.org/10.1257/jep.29.3.3>
- [9] International Energy Agency. (2025). Energy and AI. IEA. <https://www.iea.org/reports/energy-and-ai>
- [10] OECD. (2025). Competition in Artificial Intelligence Infrastructure. OECD Roundtables on Competition Policy Papers, No. 330. <https://doi.org/10.1787/623d1874-en>
- [11] United Nations. (2024). Global Digital Compact, Annex I to the Pact for the Future, A/RES/79/1. <https://www.un.org/pact-for-the-future/en/annex-i-global-digital-compact>
- [12] OECD. (2024). OECD AI Principles, updated 2024. OECD.AI. <https://www.oecd.org/en/topics/ai-principles.html>
- [13] UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. Paris: UNESCO.
- [14] UNESCO. (2023). Guidance for Generative AI in Education and Research. Paris: UNESCO. Updated 2026.
- [15] World Health Organization. (2024). Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models. WHO. ISBN 978-92-4-008475-9.
- [16] International AI Safety Report. (2026). International AI Safety Report 2026. <https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026>
- [17] 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>
- [18] 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>
- [19] Wang, H., Fu, T., Du, Y., et al. (2023). Scientific discovery in the age of artificial intelligence. Nature, 620, 47–60. <https://doi.org/10.1038/s41586-023-06221-2>
- [20] OECD. (2025). Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions. OECD Publishing. <https://doi.org/10.1787/795de142-en>
- [21] Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances, 10(28), eadn5290. <https://doi.org/10.1126/sciadv.adn5290>
- [22] Shumailov, I., Shumaylov, Z., Zhao, Y., Gal, Y., Papernot, N., & Anderson, R. (2024). AI models collapse when trained on recursively generated data. Nature, 631, 755–759. <https://doi.org/10.1038/s41586-024-07566-y>
- [23] Sen, A. (1999). Development as Freedom. Alfred A. Knopf.
- [24] 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>
- [25] Ryan, R. M., & Deci, E. L. (2000). Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. American Psychologist, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- [26] Holt-Lunstad, J., Smith, T. B., & Layton, J. B. (2010). Social Relationships and Mortality Risk: A Meta-analytic Review. PLoS Medicine, 7(7), e1000316. <https://doi.org/10.1371/journal.pmed.1000316>
- [27] Lave, J., & Wenger, E. (1991). Situated Learning: Legitimate Peripheral Participation. Cambridge University Press. <https://doi.org/10.1017/CBO9780511815355>
- [28] Vygotsky, L. S. (1978). Mind in Society: The Development of Higher Psychological Processes. Harvard University Press.
- [29] Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakçı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. Proceedings of the National Academy of Sciences, 122(26), e2422633122. <https://doi.org/10.1073/pnas.2422633122>
- [30] Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a research-based design in an authentic educational setting. Scientific Reports, 15, 17458. <https://doi.org/10.1038/s41598-025-97652-6>
- [31] Perrow, C. (1984). Normal Accidents: Living with High-Risk Technologies. Basic Books.
- [32] Leveson, N. G. (2012). Engineering a Safer World: Systems Thinking Applied to Safety. MIT Press. <https://doi.org/10.7551/mitpress/8179.001.0001>
- [33] Bainbridge, L. (1983). Ironies of Automation. Automatica, 19(6), 775–779. [https://doi.org/10.1016/0005-1098\(83\)90046-8](https://doi.org/10.1016/0005-1098(83)90046-8)

- [34] Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors*, 39(2), 230–253.
<https://doi.org/10.1518/001872097778543886>
- [35] International Committee of the Red Cross. (2025). *Autonomous Weapon Systems and International Humanitarian Law: Selected Issues*. ICRC Position Paper.
- [36] OECD & UNESCO. (2024). *G7 Toolkit for Artificial Intelligence in the Public Sector*. OECD Publishing.
<https://doi.org/10.1787/421c1244-en>
- [37] European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence. *Official Journal of the European Union*.