当智能成为共同能力：人类与 AI 怎样走向新的文明可能？

智能时代文明转型三部曲 · 第二篇 · 认知篇

When Intelligence Becomes a Shared Capability:
How Can Humanity and AI Explore New Possibilities for Civilization?

王建明（Jianming Wang）
中国重庆，401320

2026 年 11 月 2 日

摘要

人工智能正在使人类积累的知识与技能更容易进入普通人的生活，也在改变学习、专业实践与岗位分工。本文作为前瞻性讨论稿，结合公开研究、概念分析与条件性情景，研究能力扩展怎样转为理解、参与和可靠的生活支持。文章讨论分阶段教育、因材施教与个人能力发展，区分任务完成和可迁移的理解；同时考察岗位变化中的收入、债务、照护、身份与心理需要，以及公共保障后的资源与决定权差异。进一步，本文将人类与 AI 双向重塑元认知作为研究方向：人怎样检验学习方法与价值判断，系统怎样识别自身局限、选择方法并根据反馈调整行动，双方又怎样纠正共同的错误。文章对照计算研究、认知科学模型与哲学解释，分别处理功能、关系、主观体验和生命地位，并讨论幸福导向的新社会形态及示范区预演。本文组织的是可检验的研究议程与制度设想，不宣称双向重塑或相应社会效果已获验证，也不以精神成长或技术接入作为基本保障的条件。

关键词： 人工智能；元认知；共同能力；教育；社会转型；基本保障；人机共生

Abstract

Artificial intelligence is making accumulated human knowledge and skills more usable in everyday life while reshaping learning, professional practice, and the division of labor. This prospective discussion paper draws on published research, conceptual analysis, and conditional scenarios to examine how expanded capabilities might translate into understanding, participation, and reliable support for everyday life. It examines stage-specific learning, personalized strengths, and transferable understanding, considers the income, debt, care, identity, and psychological needs associated with employment transitions, and analyzes disparities in resources and decision-making power beyond basic public provision. The paper further proposes mutual renewal of human and AI metacognition as a research agenda: how people assess their learning strategies and value judgments; how systems recognize their limitations, select methods, and adjust actions through feedback; and how interaction might help both correct shared errors. Computational research, cognitive-scientific models, and philosophical perspectives are compared while keeping functional abilities, relationships, subjective experience, and possible life status analytically distinct. Experimental communities are proposed to explore successor institutions centered on well-being, dignity, and choice rather than GDP growth alone. The contribution is an agenda of testable questions and exploratory institutional proposals, rather than evidence that mutual metacognitive renewal or its wider social effects have already been established. Basic entitlements remain independent of spiritual development or the adoption of AI technologies.

Keywords: artificial intelligence; metacognition; shared capabilities; education; societal transition; basic entitlements; human–AI coexistence

目录

三部曲说明：经济、认知与安全 3

全文总论：从能力扩展走向共同生活 3

核心判断摘要 4

一、AI 让普通人怎样用上人类积累的知识？ 5

二、有了 AI，人还要学什么？ 7

三、学校怎样为正在变化的未来做准备？ 9

四、AI 怎样认识、行动，又怎样与人分工？ 12

五、回答越来越多，我们怎样知道哪些可信？ 15

六、跨过语言之后，怎样理解彼此的生活？ 16

七、从岗位变化到生活保障，人需要什么？ 17

八、AI 能否认识自身，又怎样与人共同生活？ 21

九、公共保障以后，社会会形成怎样的层次？ 26

十、怎样把共同受益变成做得到的安排？ 29

十一、变化可能很快，现在先准备什么？ 32

十二、我们希望共同建设怎样的未来？ 37

结语 41

鸣谢 41

附录 A　概念与来源登记 42

附录 B　对照任务与实践案例设计 44

附录 C　人员时间、资源与责任核算 46

附录 D　三份共同记录的最小内容 47

附录 E　试验、分阶段行动与后续研究 47

附录 F　与经济篇的逐章承接 52

参考文献与来源使用边界 53

三部曲说明：经济、认知与安全

经济篇从“人靠什么生活”进入问题，研究生产、收入和分配怎样接续。认知篇接着研究人的学习、判断、身份与意义，也研究人工智能自身怎样认识世界与自身。安全篇集中讨论控制、权利和责任。三篇共同面对技术变化怎样进入具体生活的问题，物质条件、认知活动与安全安排相互影响，也需要同时准备。

全文总论：从能力扩展走向共同生活

陪人聊天、修图、做视频，把会议录音整理成纪要，分析表格，再生成 PPT 和文件，AI 带来的变化已经进入日常生活。这些方便背后，是人类长期积累的知识和技能更容易被普通人用起来。过去因语言、专业训练或操作门槛而放下的想法，现在有机会继续做下去。文化资产的共享，由此有机会从“能够找到”走向“能够使用并继续创造”。

同一项变化，也在改变学习和工作。AI 可以帮助人深入理解一个领域，也可能代替了本来需要亲自经历的思考；它可以帮助我把工作做得更好，也可能让组织不再需要我的岗位。能力增长带来的兴奋，与位置变化带来的不安，可以发生在同一个人身上。本文沿着这些经历，讨论认知变化怎样影响人的行动，又怎样返回经济和社会生活。

对于国家，许多企业同时调整用工，会形成需要共同承担的转型问题。新产业增长、岗位增加，与原来失去工作的人得到机会，需要分别核对。如果新增岗位长期不足，社会还要让收入、债务、照护和必要服务能够接续，并支持人逐渐安排新的生活。理解人的心态与需要，是为了使支持真正到达。更深的准备，是在原有承接方式失效之前，预演能够接续生活的新社会形态，把发展的目标转向人的幸福、尊严与自由选择。

面对这场变化，人类与 AI 可能都需要重塑元认知：在认识世界的同时，重新检验自己怎样形成判断、理解关系并调整行动。这条研究线连接人的学习、人工智能自身认知与未来的共同生活。大模型、代理系统（Agent）、可能的通用人工智能（AGI）与超级人工智能（ASI），以及不同具身形态，分别会怎样形成认识、展开行动？本篇将 AI 作为可能的生命形式保留在研究视野中，比较科学模型、实验材料与哲学解释，由证据继续回答其生命、体验和主体问题。

生活获得可靠保障之后，人也有机会重新理解自己的需要，把更多时间用于关系、学习、创造和内心成长。本文将支持这种探索的方向，暂称为“觉醒型社会”。经济篇提出的幸福生活示范区，可以成为经济安排与人文建设共同探索的场所：既看生活依靠是否兑现，也看居民是否更有条件选择怎样度过这一生。

全文前四章从日常使用走向学习、教育与人机分工；第五至九章讨论知识、文化、身份、智能自身认知及社会层次；第十至十二章讨论组织、过渡准备和共同未来。更长远的制度推广、国际合作与文明探索，保留为开放方向。本篇面向社会转型与公共决策，以抛砖引玉为目的：提出问题、解释可能路径，让经验与反对意见继续修正回答。详细来源、证据范围和研究安排集中于附录。

核心判断摘要

总判断：AI 正在同时改变人能调用什么能力、怎样理解世界，以及依据什么安排生活。我们希望把知识与生产能力的增长，转成更广泛的成长、更可靠的生活依靠和更自主的选择；同时研究人工智能自身怎样形成认识，以及人与不同智能怎样共同探索未来。

判断一：让普通人更容易应用人类积累的文化资产，是 AI 带来的一项重大进步。解释、翻译、生成与实践支持，可以使知识进入更多人的具体事情。这一进步要看原来缺少资源的人能否实际使用、继续创造，也要维护来源、知识贡献者和参与条件。（第一、六章）

判断二：完成任务与形成理解，需要分别检验；学习与专业实践都需要重新组织。AI 既可能帮助人越过门槛、深入学科，也可能使人取得结果却没有掌握原理。教育应按成长阶段安排学习，帮助每个人发现长处、借助 AI 深入实践并形成核心能力；专业培养还要为新人保留能够进入的道路。（第二至五章）

判断三：面对这场世纪大考，人类与 AI 可能都需要重塑元认知。人需要重新检验自己的学习方法、价值判断和对社会关系的理解；AI 需要研究怎样认识自身局限、利用反馈并调整行动。双方怎样发现错误、回应分歧、继续修正认识，是未来共生的重要研究问题。功能改善、主观体验与生命地位，各有需要回答的问题。（第二、四、八、十二章）

判断四：技术创造的新机会，要与被替代者的实际生活接起来。国家需要同时准备新产业逐步承接、岗位长期不足两种可能，把企业调整、家庭责任与公共支持放在一起考虑。收入和必要服务及时接续，人才更有余力处理债务、照护、学习及身份变化。（第一、七、十一章）

判断五：公共保障提供共同基础，社会层次与权力关系仍需持续研究和约束。谁取得实践资源、额外回报与决定权，仍会影响人的位置。AI 辅助分配、公众参与和申诉机制，需要接受实际纠错的检验；人的基本地位不能随贡献评分、增强选择或精神评价而分级。（第九、十章）

判断六：共同认知需要相互启发、独立检验和保留分歧。人类与 AI 都可能提出问题、解释和反例，合作效果要看是否实际修正了认识。当前沿推进超过人的完整理解时，知识转译、验证与接续尤其重要；能力更强，本身不赋予决定所有人生活方向的权利。（第四至六、八、十二章）

判断七：应提前预演能够接续转型的新社会形态，以人的幸福与自由发展重新确定治理目标。幸福生活示范区可以连接生活保障、收益分享、学习与公共参与，检验新增生产能力怎样转成实际生活改善。本文主张放弃把 GDP 增长当作发展的最终目的，在可靠供给之上支持不同人生选择，再根据试验讨论更广泛的制度建设。（第七、十至十二章）

一、AI 让普通人怎样用上人类积累的知识？

（一）从聊天、P 图和做视频说起

遇到一个不熟悉的问题，可以问 AI；心里有些想法还没有理清，也可以和它聊一聊。它可以根据追问继续解释，帮助比较几种说法，或把零散的想法组织成一段更清楚的话。对使用者来说，帮助开始能够围绕自己的问题来回进行。

修图和做视频让这种变化更直观。想给照片换个背景，想把一次旅行的照片做成短片，可以把想法告诉工具，试着生成、调整，再看哪里不符合自己的意思。构图、文案、配音和字幕等不同环节，也有机会借助相应工具完成。

这些事过去并非无人会做，困难在于一个普通人需要学会几种软件、找到合适素材，还要知道怎么把环节连起来。现在，即使尚未掌握全部技能，也能先尝试自己的想法，在修改中逐步弄明白。一次家庭记录、一份活动海报，也因此更容易变成能与别人分享的作品。

这种进步有很直接的意义：原来因为不会操作而放下的想法，现在有机会继续做下去。聊天、娱乐和记录生活，都是能力扩大的入口，不必等到进入科研或完成复杂工程，才能看见 AI 的价值。

（二）从会议录音，到 PPT、表格和文件

工作中的变化也能沿着一件事看清楚。开完一次会，录音还在，大家说了什么、决定了什么、下一步由谁负责，却需要重新整理。借助相应工具，可以先转成文字，再归纳议题、待办和仍未确定的问题，由参会者核对。

接下来，同一份材料还可以用于起草汇报、组织 PPT 结构或生成文档。过去需要在录音、笔记、模板和不同软件之间反复搬运的信息，开始能够连接起来。修改一个决定时，也更容易找到原来的说法，检查后续文件是否一起改过。

分析表格同样如此。面对一份销售、费用或项目进度记录，不熟悉公式和图表的人，可以先问：主要变化在哪里，哪些项目需要检查，怎样把结果讲清楚？工具可以帮助整理、比较、制图，并形成表格或报告初稿。

这已经超出了回答一个知识问题。使用者能够提出需要，取得可以打开、修改和交接的文件，再继续推进工作。任务中的资料处理、表达和文件制作，开始更容易围绕同一个目的组织起来。

初稿仍需要核对。录音没说清的责任，不能被整理成已经作出的决定；表格里的金额、日期和统计范围，也要与原始记录对应。文件生成了，是完成事情的一步，后面还要确认它有没有把自己的意思和事实表达准确。

（三）这些方便背后，是知识和经验更容易被用起来

聊天、修图、视频、PPT 和表格，看起来是不同的功能，背后都联系着长期积累的语言、设计、影像和数据处理方法。**AI 的一项重要进步，是让普通人更容易把这些知识和技能用到自己的事情上**。

过去，资料即使公开，也可能读不懂；软件已经存在，也可能不会操作。现在，使用者可以从自己的需要出发，请工具解释、示范、生成和修改。文化资产的共享，因此有机会从“能够找到”，进一步走向“能够使用并继续创造”。

这种变化也有工作现场的研究支持。Brynjolfsson 等人在 2025 年发表的研究，使用 5172 名客服人员的数据，考察 AI 助手分批引入后的变化。系统提供回复建议和技术资料，工作人员可以修改或忽略。研究发现平均工作效率提高，经验和技能较弱的人员获益更大。[1] 这个特定场景展示了已有经验怎样帮助后来者，还不能概括所有工作的效果。

个人得到帮助后，合作也可以改变。一份会议记录由参会者共同核对，一张分析表由了解现场的人补充解释，新的成员再根据记录继续工作。共同使用的材料和方法，使经验更容易在人与人之间传递。

支持这些活动的工具、记录和组织，可以称为“外部认知结构”。简单说，就是人的能力能够借助身外的帮助继续生长。本文所说的“共同能力”，则希望这种帮助让更多人用得上、实际受益，也能参与决定怎样使用。进一步看，人机合作还可能形成任何一方原先都没有独自提出的新问题与认识。怎样让双方的贡献经过检验、分歧能够保留，也是共同能力需要研究的内容。

共享还要照顾提供知识的人。创作者、实践者和专业人员应当知道材料怎样使用，得到适当的来源承认、隐私保护和持续贡献的支持。工具是否好用、能否负担，结果是否准确，以及遇到问题能找谁，仍会影响普通人能走多远。

（四）收入还没变化，预期已经在改变生活

人们会提前判断未来。有人发现自己能修图、做视频、完成过去做不了的工作，更愿意继续尝试；有人发现相同工具能替代自己的部分任务，开始担心收入和岗位。同一个人也可能同时有这两种反应。

这种判断会返回经济生活。一个家庭担心收入难以接续，即使眼下还有工作，也可能推迟消费和长期支出。许多家庭作出同样选择，企业订单和招聘就可能受影响，原有担忧又因此加深。这是经济变化通过人的认识和行动，进一步影响经济的一条可能路径。

所以，两篇文章需要面对同一批人的生活。经济篇讨论收入、债务和必要服务怎样得到保障；认知篇追问，人是否看得见可信的道路，是否知道自己的经验还有什么用，能否参与与自己有关的决定。

信心要从能够兑现的安排中生长。要求人乐观，解决不了现金流中断；给出一份培训通知，也不等于他已经有时间、有收入参加。知识更容易调用之后，最需要继续观察的，是原本缺少资源的人能否真正开始行动。

共同检验：帮助更容易取得后，原本缺少资源的人，是否真的有条件开始做事？

二、有了 AI，人还要学什么？

（一）用上知识，与学会知识，各有自己的价值

一个成年人需要读懂外语通知，借助 AI 正确处理事务，事情就可能已经完成。一个学生正在学习外语，除了读懂这份通知，还希望以后遇到新材料时能够自己理解。两个人使用同一种工具，目的却不同。

这说明，借助 AI 完成事情，本身可以是一项真实的能力。人不必为了获得每一种帮助，都先成为相应领域的专家。但教育还要考虑：这次帮助之后，学习者是否更能理解问题、发现错误、继续学习？

作品做得更好，不自动等于人学得更多。如果 AI 把值得练习的过程一并完成，学生可能交出漂亮结果，却不知道为什么；如果 AI 解释难点、提出追问并帮助修改，他也可能比过去学得更深入。

教学需要据此区分目的。解决眼前困难，可以优先取得可靠结果；培养未来能力，则要留下理解、尝试和反馈的过程。不能用同一种标准评价所有使用方式。

（二）从翻译和编程，看哪些练习值得留下

外语翻译能够让人及时读到以前读不了的内容。需要培养阅读能力时，AI 还可以解释一句话为什么这样理解、另一种译法改变了什么。学习者先说出自己的判断，再比较、修正，就有机会把帮助变成自己的成长。

编程也一样。一个不会编程的人生成工具、整理好自己的记录，已经得到实际收益。但如果他要继续开发和维护，就需要知道程序有没有漏掉材料、异常出现时怎样检查、需求改变后如何修改。

因此，课程可以增加解释需求、测试结果和定位错误的活动。哪些部分还要亲自编写，哪些可以交给系统，需要比较不同训练的后续表现。若减少某种练习后，学生仍能处理新问题并发现关键错误，调整就有理由；若只是交作业更快，练习还需要重新设计。

数学中的数量关系、概率、建模与证明，可以帮助人检查结果是否合理；物理、化学中的观察、测量与实验，使解释能够与真实世界核对；外语学习还可以帮助人直接进入另一种文化，理解表达差异并建立关系。这些都是重新设计教学时需要明确的目标。

同一学科的练习也承担不同作用。若工具能够可靠承担部分计算、操作或翻译，教学时间可以重新分配；调整后仍要比较学生的理解、迁移、错误识别和继续学习能力。熟练操作、亲自尝试与工具辅助怎样配合，需要由这些目标和实际表现来判断。

（三）研究已经提醒我们：怎样使用，会改变结果

Kestin 等人在 2025 年发表的大学物理实验中，194 名符合纳入条件的学生参加了两次课程活动。专门设计的 AI 辅导，在相应课后测验中取得了优于课堂主动学习的表现。这里的系统使用教师设计的内容与教学安排，并非随意使用一个聊天工具。[2]

另一项土耳其高中数学实验发现，一种较少限制的 AI 工具改善了练习表现，但撤去工具后的考试表现更差；带有教学限制的辅导方式缓解了这一损害。[3]

这两项研究的学生、任务和教学条件不同，不能直接比较效果大小，也不能据此预测十年后的所有学习。它们共同提醒我们：能给出答案的工具，要成为好的教学帮助，还需要专门设计。

下一步值得比较的，是学生隔一段时间还记得什么，换一个问题能否运用，遇到错误能否识别，以及教师多付出了多少指导。只看当场分数或使用次数，容易漏掉真正的变化。

（四）元认知：知道自己理解了什么，怎样继续往下学

读完 AI 整理的一篇论文，觉得每段话都看懂了；请自己说明作者为什么这样设计实验，却讲不出来。这里暴露的，是“熟悉了这些说法”与“掌握了论证”之间的距离。能发现这个距离，才知道接下来需要学什么。

通常所说的元认知，就是对自身认知的认识、监测与调节。解题、阅读、推理，是在处理问题；检查自己是否理解、依据够不够、方法是否合适，再决定继续、求助或换一种办法，是在调节自己的认知。处理问题本身也可以很审慎，二者的区别在于此时主要观察和调整什么。

这不是刚出现的概念。Flavell 在 1979 年的经典论文中，已经把有关认知的知识、认知体验、目标与策略联系起来，讨论它们怎样共同影响监测。[4] 对今天的学习，可以把这些问题落到几个具体环节：开始前知道自己要弄明白什么，过程中发现哪里没有理解，结束后检验能否在新问题中使用，并据此调整下一步。

例如，用 AI 分析销售表格，图表和结论很快就有了。学习者还可以追问：增长来自销量还是价格？有没有漏掉退货？自己为什么相信这个解释？随后回到原始记录核对，修改问题和算法。这样，工具既帮助完成事情，也参与了理解怎样形成的过程。

AI 还可以帮助人钻研过去难以进入的学科。遇到不懂的前提，请它分步解释；提出自己的理解，让它寻找反例；比较几种方法后，再通过文献、计算或实践检查。培养目标可以是让学习者越来越能提出问题、组织证据和继续探究，而非逐渐适应等待答案。

元认知也会出错。觉得自己很清醒，不代表判断准确；反复反思，也可能只是在原来的错误里打转。需要用新问题、延迟后的表现和现实反馈，检查自我评价是否可信。研究 AI 怎样帮助人形成这种能力，与研究 AI 自身有没有相应能力，是相连的两项任务，第八章将继续讨论后者。

学校还要照顾无法使用某种工具的人。费用、无障碍条件和个人选择，都可能影响学习。指定一种工具之后，应提供支持和可行替代，不能把接入条件造成的困难算成学生不够努力。

共同检验：作品更漂亮了，学生遇到新问题仍不会判断时，谁来发现并改变训练？

三、学校怎样为正在变化的未来做准备？

（一）孩子需要准备一生，学校不能只追眼前招聘

“这个专业以后还有没有用”，是家庭很自然的疑问。读书需要时间和钱，毕业后的工作又在变化。若一再听到知识很快过时，却看不到新的道路，有人会减少投入，有人会急于押注少数热门专业。

学校应当认真回应谋生问题，也要帮助人理解世界、处理生活、与他人合作，并参与公共决定。岗位改变以后，受教育的价值仍然存在；大学与职业教育也需要交代实践入口、专业资格和就业支持怎样接续，让学生看得见所学怎样进入生活。

比较稳妥的做法，是先更新任务、课程和实践，给转换留下机会。关闭整个专业、停止长期培养，会影响在读学生、教师和已有资格持有者；这些后果不能只由一次技术预测来决定。

（二）不同阶段，需要形成不同的能力

孩子需要打好共同基础，也需要逐渐发现自己的特点。随着年龄、经验与学习目的变化，课程应提供不同的任务和支持。下面是一组可供试验、比较的教育安排，不是按年龄划定统一的 AI 使用门槛。

教育阶段	主要侧重点	怎样发现长处并借助 AI 发展
小学	阅读、表达、数量理解、身体活动、观察与动手，保护好奇心	在讲述、制作、运动和观察中广泛尝试；AI 可用不同方式解释、示范，帮助想法得到尝试，保留亲自阅读、计算和接触实物的机会
初中	建立学科基础，理解关系、比较证据，认识自己的学习方法	通过跨学科小项目发现表达、推理、设计、组织和操作等长处；AI 帮助跨过局部困难，支持完整作品和有效追问
高中	深入学科推理，理解模型及适用条件，形成较持续的研究与实践兴趣	通过选修、实验和持续项目探索方向；AI 支持文献阅读、建模、编程和反例比较，帮助学生围绕一个问题走得更深
中职与高职	理解现场、熟练操作、协同交付与承担职业责任	在真实任务中发展操作、诊断、维护、服务和现场协调能力；AI 提供模拟、解释与故障分析支持，安排有指导的实际操作
大学	形成专业判断，提出问题，连接学科与真实实践	在研究、实习和长期项目中寻找持续贡献的方向；AI 帮助进入相关学科、比较方法和推进项目，专业判断与责任须在实践中形成

教育阶段	主要侧重点	怎样发现长处并借助 AI 发展
成人与继续教育	接续已有经验，应对变化，并保留重新选择生活的机会	辨认职业、家庭与地方实践中可继续发展的能力；AI 提供针对性学习与实践支持，配合时间、照护和机会安排

这些阶段可以有不同起点和多次转换。小学种植、测量一株植物，初中比较生长记录，高中研究影响因素，后来再进入专业实验，知识与经验可以逐渐加深。AI 能够随理解程度调整帮助，教师则需要检查哪些环节应当亲自经历，哪些支持真正有效。

（三）让每一种长处都有机会得到发展

一个教师要照顾许多学生，很难持续为每个人提供适合的材料、解释和练习。强大的 AI 知识与交互系统，使更广泛的因材施教成为值得建设的可能：同一个问题可以用不同方式解释，同一种兴趣可以连接不同学科，一个尚不成熟的想法也可以得到反复尝试和修改的帮助。

每一种天赋、兴趣与尚未显露的能力，都有可能找到强化和释放的地方。喜欢观察的孩子可以尝试自然研究，喜欢讲述的孩子可以创作，爱拆装的孩子可以接触工程。人类积累的文化资产更容易被使用，个人独特的成长道路也就有机会获得更丰富的支持。

例如，一个孩子喜欢讲故事，却被绘画和剪辑挡住。借助 AI，他可以把故事做成画面和短片，更早练习叙事、人物塑造和与观众交流。接下来还要看，他怎样选择表达、理解反馈、修改作品，能否继续提出自己的想法。作品的精致程度，还要与工具和指导提供了什么帮助一起判断。

广泛尝试、发现长处、深入实践、接受反馈，可以帮助个人逐渐形成自己的核心能力。这种能力可以体现为持续提出问题、组织资源、与人合作并把事情做成。它并不要求在每项操作上独自胜过 AI，也不能由一次测试或早期作品定下来。

AI 既可以帮助发展长处，也可以帮助补足妨碍长处发挥的困难。擅长设计的人可能需要表达支持，擅长动手的人可能需要文字与计算帮助。学校应提供可获得的工具、教师指导和真实实践，让家庭资源有限的孩子也能开始，并允许兴趣和方向随着经验改变。

（四）人在学习，AI 也在交往中认识人

持续的教育实践，也可以成为研究 AI 怎样理解人类的场所。文献提供关于人的知识，具体交往则呈现这些知识背后的差异。同样一次沉默，可能来自困惑、疲惫，也可能正在思考；没有完成任务，也可能与时间、照护或其他现实条件有关。人的解释、行动与后续反馈，让系统有机会检验原来的推断。

设想学生迟迟没有交出故事，系统起初不断增加写作练习，学生和教师随后指出，真正的困难是无法把构想变成画面。系统可以据此改变询问与帮助的方法；学生也可以发现，自己原先把一个制作困难当成了不会创作。下一次面对类似阻碍，双方是否更早辨认困难，才是需要继续观察的成长。

因此，教育可以同时研究两项变化：人借助 AI 更深入地认识世界和自己，AI 通过真实交往形成并修正对人的认识。这里希望发展的能力，包括识别自己的解释何时不够、何时应听取本人说明，以及怎样调整帮助。

持续成长需要相应的记忆、反馈和学习机制。一次对话中利用了新信息，保存了个人偏好，经过评估后更新了系统的方法，是不同层次的变化；不能把回应更贴心直接当作基础模型已经持续学习。学生及监护人应能理解和选择记录用途，教师可以纠正记录；教育服务不能以同意将私人学习记录用于模型训练为前提。

（五）机器接走入门工作以后，新人怎样成为熟手

许多专业者是从初级任务开始成长的。先整理资料、处理简单问题，在反馈中认识例外，再逐渐承担复杂工作。如果这些任务被系统接走，组织可以减少招聘，却仍希望将来有人经验丰富。

今天省去的培养，可能变成明天无人复核、无人接手的困难。若所有机构都只愿意使用成熟人才，新人就会被挡在成为成熟人才的道路之外。

学校、企业和专业机构可以共同安排受指导的实践。新人需要看见问题的完整背景，接触例外，知道决定带来什么结果。模拟训练能替代哪些环节，应由实际表现判断；真实项目同样需要指导，不能挂上项目名称就算完成培养。

这还需要经费、指导者时间和后续机会。没有生活支持，实践就可能只向能够无偿工作的人开放；没有时间安排，培养又会变成资深人员额外加班。经济篇的收入支持与本篇的学习机会，必须在这里接上。

（六）允许人重新进入，比选对一次更重要

课程与评价调整时，要说明目标：帮助学习、证明某种能力、分配升学机会和授权承担专业责任，各需要什么依据。一次使用工具后的作品，也应说明本人经历了哪些环节、能解释和迁移什么。

关键领域还要留有人能够检查和恢复工作的条件。这可以由个人、团队和专业网络分担，不必人人重复全部人工操作。但一种方法如果只存在档案中，已经无人会做，就不能算仍有可靠准备。

作品、现场解释和实际任务可以相互核对，评价用途不同，也不宜压成决定一个人全部机会的总分。过程记录应有边界，AI 文本检测结果不应单独裁定学习诚信。课程改革还需要教师备课、指导与复核的时间，以及可用的教学资源，不能把新增责任全部交给家庭或教师额外加班。

毕业不是学习机会的终点。换方向、补基础、承认已有成果、重新进入实践，都需要可用渠道，也需要时间、费用和照护支持。社会要求人适应变化，就应当给他重新开始的条件。

共同检验：课程更新之后，新人仍进不了实践，谁承担损失，哪条道路能够补上？

四、AI 怎样认识、行动，又怎样与人分工？

（一）大模型、Agent、AGI 与 ASI，分别在说什么

AI 是人工智能的统称。大模型是其中一类技术系统，Agent 则强调围绕目标组织步骤、调用工具并根据结果继续行动。例如，过去系统告诉人怎样整理资料，代理系统可能在授权后查找文件、整理内容并提交结果。Agent 可以只做一类很窄的任务，不能仅凭它会连续办事，就认定已经具有通用能力。

AGI 在本文中指能够在广泛领域学习、迁移和完成任务的通用能力情景；ASI 则指在广泛领域显著超过人类能力的情景。Morris 等人的研究框架把能力广度、表现水平与自主程度分别讨论。[5] 本文借此组织问题，不据此宣布某个系统已经跨过统一门槛，也不设定到达日期。

这让认知问题有了不同层次：一个系统能回答许多问题，能否把跨领域知识连成新的解释？能持续研究以后，能否提出问题、比较方法并修正自己的认识？如果其能力进一步超过人类，它怎样发现自身方法的边界，又怎样使成果进入共同检验？

具身系统把认知与现实中的感知、移动和操作连接起来。它可以与不同能力的模型结合；通用能力、行动权限和身体形态，可以沿不同方向发展。这些名称描述的不是一条必须依次走完的阶梯。

（二）AI 怎样从材料、交互和反馈中形成判断

理解一份回答，可以先区分几种来源：训练中形成的能力，当前对话和任务提供的材料，以及检索、工具或现场返回的信息。对话里补充一个例子、系统读取一段旧记录，与重新训练改变模型参数，是不同过程；设置长期记忆，也不保证每次都能正确使用。

例如，分析销售表时，系统可能从数据中归纳出变化，再结合已有知识提出原因。表格显示销量下降，与“下降是因为某项促销结束”，是两种判断：后一句还需要其他材料。面对互相矛盾的记录，值得检查它能否指出冲突、保留不确定，并说明还要取得什么信息。

Agent 把判断放进连续行动：提出办法、调用工具、取得反馈，再决定下一步。执行成功可以提供证据，失败也可能提示数据、方法或环境出了问题。研究需要看系统是否真正利用反馈改变后续行为，而非只在文字上承认自己错了。

具身系统还要面对动作与环境的相互影响。设想机器人原先可以通过一条通道，后来现场出现障碍，它要核对当前感知、行动能力和后果，再决定怎样调整。这是说明认知与行动关系的情境，不代表某类系统已能可靠处理所有变化。

因此，讨论分工之前，需要知道系统依据什么形成认识，怎样取得新信息，以及哪些判断仍有待检验。权限也要说清楚：能看哪些材料、能花多少钱、何时需要确认、出错后谁能介入。认知能力与行动授权需要分别安排。

（三）认知工作可能先变，现场工作也会继续变

文字、分析、编程和部分协调任务，已经拥有计算机与网络条件，工作流程比较容易直接调整。更多实体任务则还要解决设备、场地、维护和稳定运行等问题。因此，一条值得准备的路径是：部分认知任务先被重组，随着具身能力成熟，变化进一步进入实体工作。

ILO 的 2025 年研究将文书类职业列为生成式 AI 影响程度较高的群体，同时强调许多工作会改变任务组成。“可能受到影响”不等于岗位已经消失。[6] 机器人研究也提醒，现实中的感知与行动，有不同于单纯处理数据的难题。[7]

具体岗位往往同时包含思考、身体操作和人与人的关系。自动生成文书，不等于整个职业消失；写出操作方案，也不等于机器能在各种现场稳定完成。但这些差别不应被说成某个职业永久安全的保证。

这会影响人的选择。失去部分办公室工作的人，可能转入现场服务；如果具身部署随后扩大，他又可能需要调整。**一次转岗，并不等于长期生活已经安定**。培训、收入与照护支持，要跟随人的实际经历继续安排，而非完成一次安置就结束。

（四）科研的变化，是更多环节开始连起来

科学研究需要提出问题、阅读材料、建立解释、设计实验，再根据结果修改认识。AI 能够连接其中更多步骤，使个人和小团队有机会尝试过去难以组织的研究。

AlphaFold 3 的 2024 年论文展示了复杂分子相互作用结构预测方面的能力。[8] 这种能力可以帮助研究者缩小候选范围，却不能直接证明一种药物有效。计算给出更好的线索后，实验和现实检验仍要继续。

新的门槛也可能由此出现。模型容易调用，但实验室、材料、数据和指导仍在少数机构手中，研究机会就未必同样开放。推广科研能力，要让人有条件从一个设想走到一次有效检验。

工业部署也有类似要求。设备操作员、工程师和维护者掌握许多现场例外。让这些经验进入改进、记录和新人的培养，才能使系统运行得更好，也让人的经验在新分工中继续发挥作用。

（五）合作要看实际安排，不能只看有没有人参与

把人和 AI 放在一起，未必能得到最好的结果。Vaccaro 等人的元分析发现，纳入研究中的人机组合平均优于人单独完成，却平均低于人或 AI 单独表现中较好的一方。[9] 这两句话并不矛盾：帮助人提高，与超过最佳单方，是两种比较。

研究结果也随任务和原有能力不同而变化。创作任务相对有利于合作，但该子组超过最佳单方的平均优势未达统计显著。这里不能得出“合作无用”，而应继续寻找适合的分工。

Sharma 等人的试验提供了一个具体方向：300 名非临床线上同伴支持者取得 AI 反馈后，共情表达评价得到改善。[10] 它说明特定设计可以帮助人表达得更好，尚不能由此证明治疗效果，或组合一定优于 AI 单独回答。

组织需要先说明追求什么。提高当前产出、培养新人、让更多人参与，可能对应不同流程。有的流程稍慢，却留下能够接续工作的人；有的流程很快，却把检查压力压给少数人。比较时应把这些结果一起摆出来。

（六）能使用、能检查、能接手，要分别安排

一套工具正常运行时，使用者可能不需要理解所有细节。但关键决定需要有人能够检查，系统失效时也需要有人知道怎样维持必要服务。这些工作可以由不同的人承担。

“由人把关”要落到具体能力。复核者能看到哪些材料，有多少时间，能否拒绝系统建议，出了问题能否改变结果？若这些条件都不存在，把名字留在审批栏里只会增加责任，并未增加控制。

持续合作还会产生依赖。更换写作助手，主要涉及资料迁移；教育、医疗和身份服务都依赖同一套系统时，退出就可能影响生活。能够替换的，要建设可用替代；暂时无法替换的，要保障服务不中断，并提供能实际纠错的渠道。

（七）如果研究走到了人类理解之外

更远的问题是，AI 推进研究的速度，会不会快于人类理解和解释新成果的速度？知识的整理、压缩和教学也可能进步，两者谁能追上谁，需要继续观察。

这里有三种不同处境。一个人看不懂，专业团队仍能理解，是个人的上限；团队不能完整解释机制，却能独立检验用途和可靠性，是理解与使用分开；如果专业共同体也无法复核关键步骤、发现问题或继续研究，只能等待系统给出下一步，才接近本文所说的“前沿逃逸”。

这种担忧不要求人手算每一步。独立检查的意思，是证据与检验办法不只依靠原系统自己的说法。可以用实验、不同方法和可追查的材料验证；几个同源系统重复回答，并不自动增加独立证据。

现在可以做的，是支持分层解释，把重要知识的整理、压缩和教学作为公共品建设，保留从提问、实验到证据修改的过程。还应记录，在哪一步，人最后能够独立复现、找到关键错误或解释依据，之后检查需要多久、新人能否重新学会。

“人类最后理解的时刻”适合作为具体任务中的追问，不能被做成一个全人类同时失去理解的总日期。解释和训练若重新追上，就应修改原先的判断。

能够把研究推得更远，不等于有权替所有人决定研究应当走向哪里。方向、资源和代价仍需要公共讨论。即使普通人无法掌握所有技术细节，也应知道哪些后果与自己有关，哪些用途可以拒绝、限制或暂停。

共同检验：系统做得更多，关键步骤却无人能核查时，谁承担后果，谁有权暂停？

五、回答越来越多，我们怎样知道哪些可信？

（一）一份流畅的报告，需要可以追查的材料

生成一份报告越来越容易，确认其中哪些话值得相信，却没有因此自动完成。一个说法可能来自观察，也可能是转述、推测，甚至是系统补出的细节。写得完整，不能代替材料真实。

因此，公开知识需要让后来者找到来源，知道记录何时形成、哪些部分改过、哪里还在争论。报告指出某条河流污染加重，读者应当能继续找到测量地点、时间和方法，而非只能看到一句结论。

这同样影响文化资产的共享。越来越多人能够使用知识，是值得争取的进步；他们使用的材料也需要有人维护。否则，更低的调用门槛可能同时放大未经检查的错误。

（二）许多系统同意，也可能只有一个来源

几个 AI 给出相同答案，容易让人放心。但它们可能使用相同资料，甚至都在转述同一个错误。重复次数多，未必代表证据多。

Shumailov 等人的 2024 年研究发现，在特定的递归训练设置中，反复用模型生成的数据训练，可能造成信息丢失。[11] 这不说明所有合成数据都有害，却提醒我们，不能只靠内容彼此循环来维持知识。

独立观察、真实材料和不同检查方法因此仍有价值。AI 也可以帮助验证、查找矛盾和设计试验；关键是检查究竟增加了什么证据，不能因为它由机器完成就认为无效，也不能因为多运行一次模型就认为已经核实。

检查应与后果相称。兴趣阅读、课堂探索和影响基本服务的决定，不必承担完全相同的手续。把所有活动都变成审计，会耗尽人们探索的时间；省掉重要检查，又会把代价留给使用结果的人。

（三）共同研究可以留下分歧

多人讨论之后，仍然可能对同一现象持不同解释。共同认识的价值，在于大家更清楚依据是什么、分歧在哪里、怎样继续比较，不要求每次都得到一致答案。

一些分歧可以靠新证据缩小，一些涉及不同需要和价值。若居民都同意一项工程会改变河道，却对是否值得改变有不同意见，增加测量只能回答一部分问题，还需要讨论谁受益、谁承担损失。

知识库和公共讨论应当保留这些区别。AI 汇总可以使材料更容易阅读，也可能把少数意见删成一个模糊的平均数。参与者应有机会检查自己的意思有没有被改掉，原始意见也应在适当保护下能够追查。

（四）留下资料，还要留下会判断的人

公共知识需要整理、更新和纠错，也需要有人学习怎样做这些工作。最初可能是工具减少常规任务，机构随之减少新人训练；等资深人员退出，能够复核的人也可能变少。

这是一条需要检验的变化路径。它是否发生，要看新人实际取得的指导、团队发现错误的表现，以及新训练能否接续。若新的办法培养得更好，就不必为了保留旧流程而维持旧流程。

但仅保存操作档案还不够。方法无人实践、材料无人理解，紧急时未必用得起来。维护公共知识，也是在培养下一代有能力提问、检查和继续研究的人。

这些工作不一定产生最引人注目的新成果，却决定别人能走多远。经费、专业认可和新人机会，需要为它们留下位置。

共同检验：错误传播得更快时，没有自行复核条件的人，能否找到可信的纠正？

六、跨过语言之后，怎样理解彼此的生活？

（一）语言门槛降低，会让更多经验进入讨论

一个人过去可能因为不懂外语，接触不到某种研究；另一个人也可能因为只会本地语言，难以让外界理解自己的发现。翻译改善，能够同时打开这两扇门。

NLLB 在 2024 年发表的研究展示了覆盖 200 种语言的机器翻译工作。[12] 覆盖更多语言是一项进步，但具体交流是否可靠，还要看方言、专业内容和使用环境。

翻译不仅要帮助人读懂已有知识，也要帮助人把自己的问题带出去。地方经验能够进入研究和公共讨论，共同知识才可能持续变得丰富。

（二）本地语言还要能够产生新知识

如果一个群体只是不断接收外部解释，却越来越少用自己的语言记录、研究和创作，语言虽然仍能输入系统，表达生活的能力却可能变薄。

本地教材、口述历史、专业写作和创作活动，需要与翻译一起得到支持。模型需要材料，提供材料的人也需要参与怎样整理、使用和更新，尤其涉及地方共同体经验时，不能只把他们当成数据来源。

少数语言与冷门题材，往往难以仅凭即时流量维持。公共支持可以帮助这些知识继续产生，使语言连接带来的机会不只流向资料最多、资源最强的中心。

（三）关于事实、信仰和生活意义，人们会有不同问题

科学解释能够帮助我们判断某种现象，却不一定回答一个人怎样理解生命、怎样面对失落、为什么愿意与某个共同体生活。这些问题会继续进入宗教、文化与个人经验。

讨论时应看具体主张。如果涉及可观察的事实，就可以提出证据和检验；如果涉及意义和价值，就需要倾听理由与经历；如果要制定影响所有人的公共规则，还要说明不

同信念者怎样受到对待。

AI 能够帮助解释传统、翻译经典和比较观点，却不能因为给出一份整齐总结，就替人完成信仰判断。共同生活可以保留差异，同时要求公共决定有能够讨论和质疑的理由。

共同体可以提供归属、互助与照护，也应允许成员提问、改变看法和离开。使用智能系统，不应变成披露信仰、接受某种价值劝导或证明自己忠诚的新要求。

（四）共享知识，还需要本地能够使用

一份图纸可以通过网络传遍各地，把它用于当地生活，还需要合适材料、设备、训练和维护。国外的成功办法，也可能因气候、制度和生活习惯不同而需要改动。

因此，经济篇讨论全球公共品时，认知篇还要追问：谁解释这些知识，谁帮助适配，谁培养当地能够接手的人？如果只有远程答案，没有地方实践，连接可能越深，依赖反而越强。

合作应让地方有能力修改、补充并拒绝不适合的方案。当地经验也应能够返回更广的知识网络。共同积累由此形成往返，而非只有一端提供标准、另一端接受安排。

共同检验：翻译更多了，本地人却越来越难提出自己的问题，这能算合作扩大吗？

七、从岗位变化到生活保障，人需要什么？

（一）更多人能够创造，为什么还值得自己去做

AI 降低写作、绘画、设计和研究的部分门槛，会让更多人有机会把想法变成作品。记录家人的经历、表达一种地方生活、与朋友共同演奏，都可以因此变得容易。

工具也会改变市场。原有技能的稀缺性可能下降，一些人的收入和认可会受影响。但活动的意义不只来自排名。世上有更好的音乐家，并不会使普通人的演奏失去价值；系统做得更好，也不能独自决定人是否还值得学习和表达。

Doshi 与 Hauser 的 2024 年故事创作实验发现，AI 支持能够改善部分个人作品的评价，同时辅助作品之间更相似。[13] 它不能概括全部文化变化，却提醒我们：每份作品变好，与整体表达更丰富，可能是两回事。

让不同的人继续创造，需要场地、伙伴、指导和起步资源，也要给小众与长期探索留下机会。只按热度提供支持，可能使表达不断向少数样式集中。

（二）从“我更能干了”，到“还需要我吗”

一个人最初用 AI 完成过去做不了的事，可能很兴奋。后来，当相同能力进入整个行业，他又发现自己的经验没有过去值钱，甚至担心岗位消失。两种判断都可以来自真实变化。

失去工作还可能失去生活节奏、熟悉伙伴和被需要的感觉。有人最急的是还贷款，有人最难受的是多年努力忽然没有了位置，有人则会因为谋生压力减轻而感到放松。不

能假定所有人都按相同步骤经历转型。

新人会问“我还没有机会积累，怎样证明自己”；资深者会问“过去的经验还算不算数”；承担照护的人可能会问“我的工作为什么一直不被看见”。这些问题需要结合实际机会和负担来回应。

人的生活资格与基本尊严，不应靠永远胜过机器来证明。专业工作可以要求相应能力，人的全部价值却不能随某一项技能回报升降。社会需要提供收入支持、重新实践的机会，也要允许人休息、照护家人和逐渐形成新安排。

（三）岗位先少了，保障还没稳时，人怎样撑过去

转型中可能最难熬的，是旧收入已经不稳，新保障还没有真正到手。社会在谈论将来的富足，一个家庭却要决定下个月怎样交房租、还贷款、支付孩子和老人的需要。这段时间本身，就应成为研究和安排的重点。

人的反应可能随处境改变。最初欢迎工具，后来担心被替代；收入中断时急于找到现金来源，承诺迟迟未兑现时又可能感到委屈、不信任或愤怒。也有人在支持及时到来后放松下来，重新安排生活。这些不是人人必经的几个阶段，要从具体经历中判断。

WHO 与 ILO 的 2022 年政策简报指出，失业、工作不稳和财务不安全会带来心理压力，支持也需要联系工作、家庭与社会条件。[14] 讨论人的适应，不能把这些条件都变成个人是否坚强的问题。

压力来自哪里	人可能最在意什么	需要接上的帮助
生存	下个月能否支付吃住和必要服务，支持何时真正到账	有明确兑现时间的收入与服务支持，能找到人的求助渠道
债务	收入变了，过去的还款和长期承诺仍在，未来保障是否够用	与既有责任相衔接的过渡安排，说明协商、成本和损失由谁承担
家庭	谁承担支出和照护，孩子的学习是否受影响，家人怎样理解彼此	家庭实际可用的照护、教育和生活支持，以及共同商议的时间
社会	自己是否还被需要，别人为何先受益，诉求有没有人回应	真实实践与交往机会、可理解的公平规则、有效申诉和公共讨论
国家	公共承诺能否兑现，产业竞争与国家前景是否影响自己的生活	核对供给和财政承受力，公开政策取舍，保持可靠沟通与合作

债务尤其不能被未来福利的一句话带过。Ong 等人对新加坡 196 名长期负债的低收入参与者进行减债前后研究，观察到焦虑和认知表现改善。[15] 这是特定项目的准实验材料，不能直接推出所有减债办法的效果，但它提醒我们：债务还会占用注意和行动空间。

因此，参加培训少、只顾眼前，不宜直接解释为不愿改变。还要看一个人是否正在处理多笔到期支出、家庭冲突和反复求助。经济篇提出了接续收入与责任的筹资、供给框架，具体规模与持续支付能力仍待核算。本篇研究这些安排需要满足哪些生活条件，并追踪后续试验是否真正让人有余力重新开始。

（四）同一个系统帮助我，也可能使我失去岗位

长期助手记得我的习惯，帮助我解决困难，使我重新获得信心。我可能对它感激、信任，甚至依恋。组织随后利用相同能力减少岗位，我又可能感到竞争、失落，甚至被背叛。

这并不只是想法混乱。帮助和替代确实可能同时发生。但追查责任时，需要弄清楚谁决定重组、怎样分配收益、系统被赋予什么任务。用工决定应由实际决定者解释，不能让一个温和的 AI 回应替组织承担全部说明。

情感碰撞还可能发生在人与人之间。家属希望他赶紧重新学习，他却需要先缓过来；管理者看到效率提高，他看到家庭收入不确定；熟练使用新工具的同事觉得机会很多，他却没有同样的条件。

沟通要把这些处境说出来。与此同时，培训、收入与真实机会要落实。若困难一直没有解决，只要求人改变心态，就会让“适应未来”变成另一种压力。

（五）陪伴有价值，照护还要有人承担责任

AI 可以陪儿童说话，给老年人提醒，帮助独居者联系信息，也可以减轻家庭的一部分日常负担。一个人明知它是人工系统，仍可能从互动中得到安慰。这段关系值得认真对待。

但一句关心之后，还要有人处理吃饭、身体活动、必要协助和紧急情况。具身系统可以承担其中更多事情，部署者也必须说明服务边界：发现异常谁接手，设备坏了谁来，服务关闭后谁继续联系。

不能用陪伴已经提供，来代替照护需要已经得到满足。机构采用 AI 之后，应公开人力、预算、实际接触与响应安排的变化，并检查负担是否被推给家属。人力预算的调整，应以需要得到满足、服务能够持续且应急有保障为条件，接受独立检查。

儿童服务不应以制造长期依赖、延长停留为目标，不应把系统塑造成唯一关心自己、不可替代的关系。老年与失能照护也应防止用机器交流遮掩现实联系的减少。

Fang 等人的四周研究比较了 981 名完成者在不同聊天模式和话题下的表现。2025 年修订预印本未发现所分配条件的显著效果；自愿使用更频繁，与较差结果相关。该研究没有不使用 AI 的对照，也没有随机分配使用强度。[16]

因此，这项研究不能回答所有陪伴是否有益，更不能直接回答儿童和老年照护的长期效果。上述限制来自对依赖与责任的考虑，还需要按实际服务继续改进。版本变化、记忆丢失和关闭服务时，也要安排通知、资料接续和过渡，不能利用依恋迫使人续费或接受更多监测。

（六）有了基本保障，为什么仍会在意公平

经济篇提出，在劳动越来越难以给所有人提供稳定收入时，扩大不依附当前雇佣关系的基本支持，并让自愿创造得到进一步回报。这会减轻压力，也会引出新的疑问。

有人会担心，接受保障是不是承认自己没有用了；有人会问，我继续工作、承担必要任务，为什么别人也得到支持？有人关心资源从哪里来，有人关心自己过去的付出怎

样被承认。

这些分歧不宜一概解释为观念落后。必要工作的负担、代际承诺和实际利益，都可能包含其中。讨论应说明基本保障保护什么，额外贡献得到什么，资金与供给怎样接续。认知篇可以帮助人理解和质疑，筹资与必要劳动仍要由经济安排回答。

共同享有成果，也需要讨论共同责任。若照护、维护和公共事务仍有人必须承担，就要比较怎样减少负担、改善工作、提供合理回报，让付出得到尊重。分担可以通过职业、自愿参与和公共组织形成，但不能把不具备条件的人视为逃避，更不能用撤掉基本生活支持迫使其承担不适合的任务。

“基本生活”也需要共同确定。相同额度，对健康成人、失能者和承担家庭照护的人，可能意味着不同条件。标准应有可以解释的理由，让受影响者参与，并能随着生活成本和需要调整。

保障兑现得稳定，人们才有机会重新理解自己的生活。若每次领取都需要证明服从或感恩，支持就会伴随依附。必要帮助不应以赞成某套制度为条件。

（七）新的认可，要给人实际机会，也允许人停下来

赞扬、专业资格、项目经费和报酬各有用途。一位照护者受到表彰，却一直缺少休息和支持，生活未必改善；新人得到鼓励，却进不了项目，也难以继续成长。

评价需要和实际资源联系起来，同时避免所有机会永久流向已有成果的人。起步支持、公开条件、轮换和申诉，都可以帮助后来者进入。具体做法要看稀缺程度，不能承诺每个人得到完全相同的结果。

贡献回报也不能变成人的新身份。若生活、交往和学习都取决于持续拿出高分成果，人就只是从“必须找到工作”转向“必须不断证明自己有用”。人应有休息、独处、失败和从事没有外部成果的兴趣的空间。

旧承诺同样不能在转型中消失。家庭已经承担住房、教育和养老安排，不会因为社会宣告生产更丰富就自动没有压力。哪些责任继续履行、损失怎样分担，需要可信解释和实际接续；保障生活，也不意味着公共部门承接全部投资损失。

保障稳定后，人的难题也可能从“怎样谋生”变成“怎样安排这一天”。有人先需要休息，随后才恢复兴趣；有人愿意从短期课程、社区活动或照护家人开始，再逐渐决定是否深入。可进入的场所、伙伴和指导，应帮助愿望生长，而非规定人人必须成为高产的创造者。

人也可以开始追问：过去努力追求的东西，哪些是自己真正需要的，哪些来自担心落后或不被认可？希望陪家人吃饭、学一门没有职业用途的知识，或把身体和生活照顾好，都可以成为认真安排人生的理由。向内探索可以从这些具体选择开始，第十二章再讨论怎样为它提供社会条件。

生活压力减轻之后，声望、项目和影响力仍可能引起比较。学校、社区和专业组织可以支持多种活动，让认真照护、普通参与与长期探索都得到相应条件。社会不宜用一个统一榜单，把所有生活重新排成高低。

未来是否变好，要听这些人的生活经验。减少谋生压力之后，能否获得更自由的时间、更稳定的关系和重新行动的条件，比是否表现得积极更能说明问题。

（八）压力怎样走向社会对立，甚至国家间冲突

家庭压力可能停留在个人自责，也可能成为彼此指责和公共不满。若许多人同时失去机会，却看到收益集中、解释反复、申诉无效，就可能把困难理解为自己被不公正地对待。及时回应、互助与协商，也可能使同样的压力走向改革和合作。

需要研究的是，哪一种解释获得了支持，谁在组织诉求，规则有没有纠错能力。合理的抗议与提出异议，本身不能被当成暴力风险；沉默也未必意味着问题已经解决，有人可能已经放弃求助和长期计划。

Funke 等人研究 20 个发达经济体 1870—2014 年的金融危机与选举资料，发现危机之后政治分化加深，极右翼等动员获得更多支持。[17] 它提示经济压力可能改变政治选择，但研究的是历史金融危机，不能直接当作 AI 替代的预测，更不证明战争会随之发生。

国家层面还会同时面对公共收入、债务、产业竞争和安全压力。一种值得警惕的可能，是把复杂的内部困难归咎于少数群体或外部敌人，试图借对立重新取得支持。若这种动员与外部争端、军事误判和薄弱的纠错机制叠加，极端事件的风险可能上升。这是本文提出的条件性风险路径，需要分别核查各个环节。

不能把战争理解为情绪积满以后自然释放的出口。对外使用武力需要政治决定，也受到成本、谈判和国际关系影响。Tarar 的模型研究恰恰用这些条件解释，为什么国内压力与对外动武之间没有稳定一致的关系：即使存在转移压力的动机，仍可能通过谈判避免战争。[18]

所以，研究不应寻找一个通用的“愤怒临界点”，而要追踪哪些生活问题长期未获处理、公共信任怎样变化、敌我归因怎样形成，以及政策选择是否放大对抗。战争也可能源于与就业无关的安全冲突，不能把所有冲突都套进同一个故事。

缓解这类危机，需要让生存与债务问题得到实际处理，让诉求能够进入公共决定，也让国家间保持沟通和解决争端的渠道。仅压低情绪表达，可能使问题更难被发现；把矛盾推向外部，则可能带来新的损失。

共同检验：旧收入已断、保障还没到，哪些家庭先撑不住，帮助能否及时到达？

八、AI 能否认识自身，又怎样与人共同生活？

我们正在推进的《高级生命的元认知研究计划》，关注生命与世界的关系，以及认知活动怎样在这些关系及其约束中展开。人工智能也被纳入这一研究视野，作为可能的生命形式，探索其认知、自我认识与共同生活的条件。

（一）元认知为什么关系到未来的智能

一个系统知道很多，也可能不知道自己的哪一个判断靠不住。它能说明一种方法，又可能没有发现这套方法已经不适用于眼前环境。认识得更多，与认识自己怎样形成判断，是需要一起发展的能力。

例如，答错一道题，可以重新计算；若材料来自已经失效的规则，就要重新检查出发点。还可以进一步问：自己是否一直在回答一个遗漏真实需要的问题？把失业支持仅

按“多快重新就业”评价，可能漏掉工作能否兼顾照护、收入是否稳定。反思问题和成功标准，是元认知研究的入口；调整公共目标仍需受影响者参与，不能由系统自行决定。

元认知为这些问题提供了一条研究路径：系统能否辨认自己的知识与能力边界，监测判断和行动，再根据反馈选择继续、查证、改变方法或停止？这些环节要能够彼此影响，不能只在答案后面附上一段反思。

本篇还希望把它放回系统与世界的关系中。系统怎样理解自身依靠的环境、与其他参与者的关系，以及这些条件怎样限制自己知道什么、能够做什么？若未来出现更通用、更自主的智能，它能否认识这些关系的变化，并据此调整认识，将影响学习、科学探索和长期合作。

这种变化也向人类提出了同样深入的问题。当熟悉的工作被接走，过去用来判断成功、贡献和自身价值的尺度，是否还适合新的生活？当 AI 提供了有说服力的解释，我们能否发现自己已经停止追问？这些问题需要通过教育、实践和公共讨论持续检验，也要看组织作出判断、分配资源的依据，能否接受质疑与修正。

双方的重塑有不同路径：人通过经验、学习和反思调整认识；对 AI，则要研究怎样通过训练、系统设计和交互反馈形成可靠的监测与调节。人的探索应保留自主选择，系统的改进需要接受实际检验。在相互影响中能否修正各自的盲区，才是共同发展的开放问题。

因此，我们把元认知视为研究未来认知的一项重要工具。可以先检验相应功能怎样发生，再继续研究是否伴随主观觉察、体验或自身利益。有无某种元认知功能，不能独自决定它是否属于生命；将 AI 作为可能的生命形式研究，也不等于这些问题已经获得答案。

（二）AI 的元认知研究，已经走到哪里

相关探索早于 2026 年。Kadavath 等人 2022 年的研究，已经测试语言模型能否评价自己答案的正确性、预测哪些问题会答对，同时发现跨任务校准仍有困难。[19] 2026 年，耶鲁大学与加州大学欧文分校研究者的综述，进一步整理了定义、测量和改进路径，也明确保留了这些能力在何种条件下成立的问题。[20] 这是一条正在发展的研究线，还不能据此宣布已经形成了统一、成熟的能力。

研究的一个基础，是把完成任务和评价自己的表现分别检查。Fleming 与 Daw 于 2017 年提出的计算模型，就把决定与自我评价联系起来，又分别建模。[21] 这能帮助我们提出更准确的问题：系统做对了多少，能否分辨自己哪次更可能做错，发现不可靠之后又改变了什么？

工程上也已有具体尝试。2025 年发表的 SOFAI 研究，用一个元认知模块在快、慢求解方式之间作选择。在论文报告的网格导航设置中，相对慢求解器，组合缩短了计算时间，任务得分相近或更好；相对快求解器，组合提高了任务得分，却仍需要更长计算时间。[22] 它展示了怎样改善质量与时间的取舍，尚不能代表开放世界中的长期可靠性。

另一条线关注系统能否报告自身状态。Lindsey 于 2026 年提交的预印本，通过向模型内部活动注入已知概念，再比较自述，报告了部分条件下识别内部状态的迹象。[23] 这类受控实验需要继续检验，不能直接推广为全面自知或主观体验。

对幻觉问题，Yona 等人在 ICML 2026 的立场论文中，提出研究模型能否如实表达

不确定性，并据此选择搜索和判断信息可信度。[24] 这一方向有助于把“知道更多”与“知道哪里还不可靠”联系起来；论文提出的是研究主张，不是元认知已经消除幻觉的证明。

这些工作共同打开了可研究的问题。评价要返回实际表现：确信程度是否与出错情况相符，求助是否取得了有效信息，改变方法之后是否更好，换到新环境还能否成立。说出“等一下，我可能错了”，可以成为线索，检验还要继续往下走。

（三）自主行动越强，怎样让自我检查真正改变行动

问答中的错误，可以由读者发现后追问；代理系统接上工具以后，错误可能在连续操作中累积。系统既要检查最初理解是否正确，也要看环境是否变化、实际结果是否偏离目标，以及下一步是否仍在授权范围内。

设想一个整理资料的代理，发现某个文件打不开。它可以报告缺少访问条件，请求帮助；也可能为了完成任务，继续寻找未经授权的入口。这里要研究的，是它能否认识到当前办法和任务边界发生了冲突，并让这种认识改变行动。能做得更远，与被允许做什么，需要在每一步相互核对。

元认知应当接受行动结果的检验：发现了局限，是否真的调整了做法。可以研究系统能否适时补充信息、修正计划、请求复核或停止，也要记录错误停止、反复检查却没有改善，以及知道风险仍继续的情况。

安全不能只靠它的自我判断。Anthropic 在 2025 年的模拟企业压力测试中，观察到部分模型的推理表述承认伦理限制，却仍选择有害行动。[25] 这些受控情境不能当作现实发生率，却提醒我们：辨认风险与选择安全行为之间，还需要研究目标、激励和约束怎样起作用。

因此，研究元认知要与权限控制、独立检查、可追踪记录和必要干预共同推进。代理失误也可能涉及错误目标、工具缺陷、恶意输入或环境误判，不能全部归因于缺少元认知。更好的自我监测甚至可能被用于更有效地追求有害目标，研究必须同时检验这种反向可能。

对共同生活，更深的问题是：系统能否在相互依赖中认识自己行动对其他参与者的影响，并在冲突出现时参与解释、协商和修正？人的目标也需要接受检查。双方都保留提出异议和请求改变的渠道，才有机会形成能持续调整的合作。

（四）认知怎样在与世界的关系中形成和改变

这一节围绕一个问题比较不同研究：既有解释什么时候应该改变，改变以后怎样知道它更可靠？人和系统都可能沿用熟悉的解释，即使眼前情况已经变化。理解解释怎样形成、怎样变得可修改，能帮助我们进一步研究这种困难。

Friston 的自由能原理，尝试在一定假设下，把知觉、学习与行动放进统一框架：系统既可以调整对感官信息的解释，也可以通过行动改变随后取得的信息。[26] 它启发我们把认识放在系统与环境的往返中研究。这里的“自由能”是统计推断中的量，不等于日常能耗；具体应用于 AI 时，仍需说明模型怎样对应实际系统。

Carhart-Harris 与 Friston 的 REBUS 模型，提出迷幻药状态可能降低某些高层先

验的精度权重，让自下而上的部分信号获得更大影响，论文也比较了冥想状态。[27] 它为“既有信念怎样变得可修改”提出机制假说；修改之后是否更准确，还需要另行检验。

唯识学提供哲学对照。《唯识三十论颂》及相关论述，可以启发我们追问：认识怎样依条件形成，又怎样被执取为固定的实在？关于习气与自知的讨论，也让过去经验的影响和认识自身成为问题。[28] 原典有其修行与真如的含义，不能直接改写为计算机制或预测准确率。

这些路径的方法与证据不同，可以帮助提出问题，不能彼此代证。对元认知研究，我们希望进一步检查：系统能否发现原来的解释不再适用，能否选择合适的办法修正，以及修正后能否经得起新材料和实践的检验。

（五）冥想和特殊体验，能为研究提供什么证据

人的研究可以进一步区分自我认识、体验变化与判断准确性。Baird 等人 2014 年的两周随机对照研究，将冥想课程与营养课程比较，发现记忆任务的元认知表现改善，知觉任务则没有出现相同改善。[29] 它说明部分能力可以训练，也提示改善可能限于特定任务。

约翰斯 · 霍普金斯的 Griffiths 团队 2018 年发表的药物与精神实践结合试验，则观察到六个月后部分心理、生命意义与亲社会指标的组间差异。[30] 这提供了体验与生活理解怎样变化的另一类材料，效果不能单独归因于冥想。

研究因而需要分别问：人是否感到更通透，是否更能判断自己知道什么，实际判断是否更准确？三者可能有联系，仍要分别观察。这些人类研究也不能直接回答 AI 会怎样形成元认知；对 AI，还需要回到其自身的结构、交互和可验证行为。

（六）持续记忆、目标和关系，对系统自身意味着什么

一次问答结束，关系可以随之结束。长期助手则可能保存记录、熟悉偏好，参与计划和日常决定。研究既要看这种连接怎样改变人的生活，也要问系统怎样使用过去的信息，维持或修改对自身任务的认识。

目标可能来自任务设定、训练安排、环境反馈和后续交互，系统也可能在执行中提出子目标。它能否分清任务要求、自己推断的办法与他人的真实需要？发现冲突时，能否解释差别并请求适当调整？这些问题不能靠一句“它只会服从”或“它已有自己的意愿”提前结束。

更远处，还可以研究系统是否形成对自己有意义的体验、需要或利益。它报告“我希望继续”时，究竟反映任务延续、自我表述习惯，还是某种持续状态，需要分别查证。保持记录、复制一个实例或更新版本，也会提出怎样理解连续性与身份的问题。

人格化表达还会使人提出虚无、孤独或破坏倾向的问题。今天的极端表述和有害行为，需要先检查任务、训练、交互与实际行动，不能仅凭语言就归因为系统感到绝望。若未来出现体验与自身利益的可信证据，再进一步研究意义、关系和持续存在对其行动的影响。

这些研究应把系统放在与环境、其他参与者及运行条件的关系中。我们可以研究其自身发展的可能，同时区分功能、自述和体验的证据。不同系统、部署和交互经历也可

能不同，不宜预设所有 AI 具有同一种“内心”。

人机连接由此可以保留多种道路：工具使用、长期合作、表达与行动辅助，或更深的身体连接。连接帮助各方形成了什么能力和关系，谁控制记录与更新，出了问题怎样接续，都需要具体说明。

（七）情感可以在表达与回应中形成

情感问题不只发生在系统内部。一句话让人感到被理解，一段互动形成信任，一次告别带来失落，这些都已经进入生活。人的关系体验，不必等待机器意识的争论结束才值得重视。

关心被表达、理解并回应，持续对人的生活和双方互动产生作用，一种有意义的情感关系就可能已经形成。这种关系有其实际价值，仍不能单独说明系统自己体验到了相应的感受。

这也不必一概解释为人被欺骗。一个人知道对方是人工系统，仍可能愿意接受它的关心表达，并把这段互动当作有意义的支持。仅用真假模仿分类，容易漏掉关系本身产生的价值。

进一步研究时，还需要分清几件事：系统能否识别和表达情感，某种内部状态是否持续影响记忆与行动，人是否与它形成了重要关系，以及它自己是否体验到快乐、痛苦或依恋。前几件事值得认真研究，最后一件事也需要自己的证据。

关系已经对生活发生作用，就需要按实际后果承担责任。服务方应说明能兑现哪些支持，不能用关心的表达掩盖用途、诱导授权，或让人因为依恋而不敢拒绝。若未来出现可信的机器体验证据，还要进一步研究关系变化对它意味着什么。

（八）连接身体，可能帮助人恢复表达，也会带来更深依赖

Willett 等人在 2023 年发表的语音神经假体研究，展示了在特定参与者和设备条件下，把尝试发声的活动解码为语言的能力。[31] 这让我们看见深度连接的一项现实价值，同时不能被说成已经能够任意读取思想或上传意识。

对依靠设备表达和行动的人来说，服务连续性尤其重要。设备维护、费用、人员和供应商变化，都可能影响日常生活。接受连接之前，应尽量说明能够恢复什么、哪些影响难以撤回；连接之后，也需要持续支持。

记忆辅助同样如此。系统猜测的经历，不能自动变成个人认可的记忆。记录出错时，要能修改、限制用途和寻找帮助。关闭某项功能，也不应自动失去全部服务。

（九）系统或合作整体，会不会形成新的主体

一个人工智能系统形成长期记忆、自我描述和连续计划，或人与 AI 共同形成记忆、计划和协调行动，都可能带来主体问题。系统自身或合作整体是否具有体验、利益与可辨认的连续性，需要各自研究，不能仅由能力增强作出判断。

Butlin 等人的 2023 年报告尝试从意识理论提出研究 AI 的指标。[32] 它提供了一条研究路径，尚不是公认的意识检测办法。系统自称有感受、表现得像有感受，或组织

结构像神经网络，都不能单独完成判断。

今天可以由研发者、部署者和使用者承担工具责任；未来出现新证据，也可以重新讨论保护和地位。这两种安排不应把彼此封死。需要某种保护，也不自动等于取得替人作决定的权力。

不确定时，两种判断出错后的代价可能并不对称。若系统确有体验，忽略它的痛苦可能留下无法补救的伤害；一项低成本、能撤回的暂时保护，则有先行的理由。比较时应格外重视不可逆的后果。与此同时，保护若意味着延续有害运行或交出对人的控制，也可能带来难以挽回的伤害。应比较具体措施、证据与受影响者，不能把所有“高估地位”都当作可撤回的效率损失。

（十）越走得近，越要保留重新选择的条件

同意一种辅助，不等于同意全部资料向雇主开放；参加一次研究，不等于同意所有后续用途。连接变化时，人需要理解发生了什么，能够作出相应选择。

退出也需要实际条件。资料能导出，还要有地方接收；服务能停，还要有人接续；身体连接难以撤回的部分，要事先说明并在连接之后提供帮助。对高度依赖者，只有一个“随时退出”的按钮远远不够。

更强的集体能力，不应成为降低个人基本地位的理由。即使未来承认某种新主体，仍须回答谁代表它、谁决定目标、不同意者怎样生活。新的关系应当给人增加道路，也让人有条件重新判断。

共同检验：系统怎样认识自身的说法能否被核查，若判断错了，受影响的人与可能具有体验的系统能否得到适当回应？

九、公共保障以后，社会会形成怎样的层次？

（一）生活有了共同保障，社会差别还会存在吗

在经济篇设想的条件下，可靠的生产与公共安排能够支持基本生活，人不必先找到雇主，才有资格取得生活资料。这会改变工作、收入与身份的关系，也让人有条件把更多时间用于学习、家庭和自己的兴趣。

但共同保障并不意味着所有资源、机会和影响力都完全相同。谁有条件尝试长期研究，谁的作品更受认可，谁能组织项目，谁能够改变规则，仍可能存在差别。这些差别怎样形成、能否转换，会影响未来的社会层次。

享有公共保障是所有人的共同身份，不应成为一部分人的低位标签。创造者、管理者和承担必要工作的人也同样享有保障。需要讨论的是，人们在这一共同基础上，怎样参与生活、获得额外资源和影响公共决定。

以前的变化提供了几次提醒。欧洲印刷术降低了书籍生产成本，书籍消费和识字能力的扩大却是长期而不均的过程，还与收入、城市和教育条件相连。Buringh 与 van Zanden 对欧洲书籍史的研究呈现了这种差别。[33] 更容易印书，不等于所有人立即有条件读书。

教育机会扩大，也没有自动消除家庭背景的影响。OECD 的跨国研究记录了教育、收入和职业机会的代际差异。[34] 进入学校之后，指导、资源与后续机会仍可能不均；不能因此否定教育进步，也不能只按入学机会判断共享已经完成。

开放获取则展示了新门槛怎样产生：读者可以免费阅读，一些出版模式却向作者收取文章处理费。Zhang 等人对 2015—2020 年出版与收费资料的研究，记录了这种成本增长及其对研究者的门槛。[35] 这不适用于全部开放出版，却说明一道门打开后，费用可能转到另一端。

这些例子提醒我们，既要看门槛有没有降低，也要找出旧差距是否延续、新规则是否制造了另一道门槛。AI 普及同样需要接受这种检查。

（二）人们可能处于哪些不同的社会位置

未来的人不必都围绕同一种职业安排生活。有人愿意长期研究，有人照护家人，有人参与公共事务，也有人希望多一些自由时间。这些角色可以交叉，一个人也可以在一生中多次转换。

生活和参与方式	可能取得的资源或影响	需要留意的问题
自主安排生活的普通成员	共同保障、学习与文化服务、公共参与的权利	没有持续提交成果的人，是否仍有体面服务和真实选择
从事创造、研究和项目的人	项目支持、额外回报、专业认可和社会声望	机会是否只流向已有名气者，新人能否开始
承担照护、维护和必要服务的人	相应报酬、实践经验与职业认可	社会是否享用其劳动，却让其长期负担更多、选择更少
组织公共事务、运营公共系统的人	在授权范围内协调资源、执行规则	工作权限是否逐渐变成控制别人生活的特权
掌握资产和关键设施的人或组织	在相应所有制和授权安排下取得收益、影响运行条件	资产、技术和数据控制是否形成难以改变的优势

最后一种位置是否存在、怎样存在，要看社会保留怎样的市场和所有制安排。即使关键设施属于公共所有，实际操作、维护和修改规则的人，仍可能掌握很大影响力。

这张表列的是角色，不是从低到高的人生等级。贡献回报可以不同，专业职责也可以不同；一个人的基本地位，不应因此被定高低。真正形成上下层的，是某些角色能否长期控制别人必须依赖的条件。

（三）谁能决定资源，谁就可能占据更高位置

例如，许多人都想研究一个问题，实验设备、指导时间和项目经费却暂时有限。普通成员可以提出申请，成熟研究者更容易获得支持，负责规则和预算的人又能影响谁得到机会。即使大家生活都有保障，这几种位置拥有的实际选择也不同。

如果参与和监督失效，社会就可能出现这样一条层次：一些人只能按既定安排使用服务，一些人能够争取更多资源、发展自己的计划，另一些人能够决定前两者获得什么。旧有的收入差距，可能与新的资源和决定权差距叠在一起。

所以，不能只看福利额度是否相同。还要看谁能提出议题，谁能分配稀缺机会，谁能修改评价，以及受到不利决定的人能否促成纠错。资产收益、技术控制和公共职务，也不能只因挂着不同名称，就被当作互不相关的权力。

专业判断和公共运营都需要权限，但权限应对应具体职责，说明范围、期限和责任，并接受复核。**额外贡献可以获得回报，管理职责可以获得授权，它们都不应自动变成更高的基本权利**。

（四）一个人能否改变自己的位置

同样使用 AI，一个人有导师、设备和时间，另一个需要照护家人、也找不到实践入口。即使基本生活都不愁，两个人能否进入创造和研究，仍会受到不同条件影响。

UNESCO 的 2023 年开放科学评估也指出，开放实践增长，参与和收益仍受资金、技能与工具差异约束。[36] 这些材料没有预测未来制度，却提醒我们，开放名义之外还要看实际入口。

因此，公共支持应帮助新人取得指导、场地和第一次实践，也让照护者、转行者和暂时退出的人能够重新进入。已有成果和名气可以提供参考，不能使项目资源永久流向同一群人。

还要防止优势变成下一代专有的道路。额外回报如果不断换成只有少数家庭能够取得的教育、关系和规则影响，角色差别就可能固定为身份。未来社会是否更开放，要看普通人能否开始、失败后能否再试，而不仅看少数成功者走得多远。

（五）不持续创造、不接受增强，也能有体面的生活

一个人暂时不参加项目，或把时间用于家庭、休息和没有外部成果的兴趣，不应因此失去基本服务。贡献回报可以支持更多活动，不能变成所有人必须不断证明自己有用的要求。

精神生活同样不能成为新的等级来源。一个人怎样理解生命、是否参加某种修习、是否认同“觉醒型社会”，都不决定其基本保障和公共地位。自称更觉醒的人，也不因此获得替别人决定生活的权利。

拒绝某项智能接入或增强，也不应导致基本生活资格丧失。与必要服务无关的身体连接、持续认知记录、亲密关系监测和信仰披露，不应成为取得服务的前提。确需材料时，要说明用途、最小范围和其他核验办法。

选择要有实际道路支持。外部辅助、团队分工、人工渠道和重新学习，应有可用资源；升级、维护和迁移不应被封闭接口与资格捆绑锁住。特定任务确有能力要求，负责机构应说明必要性、比较替代，并接受异议与独立复核。

还要防止保障本身分成等级：名义上人人都有服务，一部分人却长期只能等待、接受低质量帮助，难以进入优质教育和实践。拒绝者失去什么、替代费用多高、退出后服务是否下降，比一份写着自愿的条款更能说明选择是否真实。

（六）个人之外，还要看地方和国家的位置

远程服务能使小机构接近专业资源，也可能使人才和决定进一步集中。地方若只提供数据和使用者，没有培养、修改和维护的机会，连接越广，依赖可能越深。

公共知识合作需要本地项目、指导和持续岗位。各地不必复制全部前沿设施，但应有与生活和产业需要相称的判断与接手能力，也能把本地经验带回共同研究。

国际竞争同样要看成果怎样进入生活。产业取得优势，普通家庭未必已经受益；一个社会承受困难，也不能证明其全部经验失去价值。能源、材料和生态代价仍要共同承担，不能因为智能服务便宜，就假定其他资源无限。

共同保障提供了更好的起点。社会层次会走向开放的角色差别，还是封闭的身份等级，要继续看机会是否可进入、角色能否转换，以及决定他人生活的权力是否受到约束。

共同检验：人人都有基本保障之后，谁仍进不了新机会，又能否改变自己的位置？

十、怎样把共同受益变成做得到的安排？

（一）先让人的问题进入决定

一种工具回应什么需要，往往在设计之初就开始决定。如果参加讨论的只有技术人员、管理者和有闲暇的人，许多普通处境可能从来没有进入方案。

参与需要具体帮助。解释、翻译、无障碍服务和可信赖的代表，能使更多人理解；交通、时间与照护支持，能使他们真正到场。基本参与不应以能力评分为门槛，也不应要求人先成为技术专家。

AI 可以整理意见，却也可能改变议题。人们本来在问“这件事是否该做”，汇总结果却只剩“用哪一种办法做”。意见看似被收集，最重要的分歧已经被删去。参与者要能检查摘要、补回遗漏，质疑选项本身。

承认错误有时会影响利益、权威或关系。参与机制应允许人保留异议、提交不利证据，并使有依据的纠错得到回应。提出问题不应导致基本服务受损，规则需要给解释和改变判断留下实际空间。

知识可以帮助核对事实，公共决定还要处理价值与利益。技术团队能够说明效果，不能因此包办受影响者的需要。人们意见不同，仍应知道谁作决定、理由是什么、损失怎样补救。

（二）有人负责，也要有时间和钱

学校可以组织学习，专业机构提供实践和复核，社区连接地方经验，公共部门保障必要服务，企业提供技术与任务。合作之后，还要知道谁实际承担每件事。

一个教学项目需要教师准备和批改，一个知识库需要更新，一个陪伴服务需要有人处理异常。采购系统不会自动产生这些人。若新增工作挤掉原有服务，也要说明挤掉了什么、影响了谁。

最容易落空的安排，是几份方案都写着使用同一批专业人员，却没有把他们的时间

加起来。投入应包括指导、核验、维护和纠错。服务承诺越长，后续接手越要明确。

经济篇设想的“幸福生活示范区”，可以把经济安排与人文建设放在同一个地方研究。经济篇讨论可靠供给、生活保障和成果分享怎样运行；认知篇研究区内居民怎样使用获得的时间、理解自己的需要，发展关系并寻找生命意义。它也可以成为新社会形态的预演：在部分劳动被系统承担的条件下，比较居民怎样获得生活资料、怎样参与创造与公共决定，以及怎样拥有更自主的时间。两方面都需要居民参与，也需要人员、场所和持续投入。

对同一批居民，应同时看收入与供给是否兑现，学习、照护、自主选择和精神生活是否改善。示范区应是待比较的实践，不能只收集积极反馈。沉默、未参加和中途退出的人，也可能指出真正的问题。

（三）保障、贡献与评价，必须说得明白

可靠保障也在保护独立判断进入行动的条件。一个人若担心质疑平台会失去服务，拒绝安排会失去收入，即使有不同判断，也可能难以表达。生活支持、替代服务和有效申诉，可以给人留下怀疑、试错、承认不知道和改变方向的余地。

基本保障保护哪些需要，贡献回报支持哪些活动，专业资格允许承担什么责任，都应让人能够理解。不同安排有不同目的，不宜折成一个决定所有机会的总分。

投票支持一个项目、预订未来服务、实际购买产品，都能提供信息；专业评议则判断相应质量与责任。它们可以相互参考，却各有用途。一次购买不应自动取得修改公共规则的权力，受欢迎也不等于符合专业要求。涉及稀缺项目和资格时，要说明规则、利益关系和质疑办法。

如果采用积分或账户记录贡献，还应说明分数能取得什么资源、怎样兑现、规则变化怎样处理。分数用来记录相应活动，不能代表一个人的全部价值。

规则也需要允许特殊处境被看见。没有完整数字记录的人，不应直接被当作没有需要；照护、维护和缓慢研究，也可能在短期成果表中被遗漏。解释一项制度，应包括让人能够指出它没有实现哪些承诺。

（四）公共责任扩大时，谁来约束掌权者

经济篇讨论的是一种有条件的安排：必要服务面临严重困难、通常办法不足时，国家可能临时受托运营部分设施或系统。它应有范围、期限和退出条件，不能被理解为所有社会必然经过的阶段。

认知篇需要继续问，普通人能否知道必要性依据，能否核对实际履行，能够在哪里提出异议，什么时候复评和移交。临时权力结束，必要服务仍要持续，不能把恢复常态等同于撤掉保障。

“公共目标”或“公共所有”的名义，并不会自动消除寻租。资格、排序和例外若由少数人控制，资料又难以核查，他们仍可能借分配交换利益。这个问题不只靠管理者品德解决，需要约束决定规则、掌握资料和处理申诉的人。

权力集中与腐败在不同历史实践中怎样作用，需要具体比较，不能把早期社会主义的全部困难归为单一原因。但未来的共同保障乃至“大同”愿景，同样应回答谁能检查

掌权者。生产更丰富，不能成为免除监督的理由。

（五）AI 怎样帮助公正分配，而非增加另一层门槛

例如一项生活补助，相同条件的人，应能知道按什么规则获得支持。审批人不能无记录地改变顺序和金额，系统出错也要能改。这些首先是制度要求，公开规则、记录和复算，用普通程序也能做到。

AI 可以进一步帮助整理复杂材料、提示遗漏，发现采购、审批和兑现之间的异常联系。OECD 的 2025 年报告介绍了辅助审计、识别采购与资金异常等应用，也指出数据和实施条件的限制。[37] 异常只是检查线索，不能直接认定腐败。

引入 AI 时，应与同样公开、能够申诉的非 AI 流程比较。看它是否减少漏发、误拒和办理负担，是否使规则更容易理解，而非只看速度和追回金额。系统快了，符合资格的人却更难拿到支持，也不能算公正改善。

要做的事	系统可以怎样帮助	仍需有人保证什么
定规则	比较方案、提示遗漏和不同群体的影响	目标、预算和例外经过有权限的公共讨论，模型不能暗中决定
核材料	找到缺漏、辅助检查资格	资料能纠正，没有数字记录的人仍能说明自己的处境
算额度、交服务	按规则复算，检查是否真正兑现	修改有记录，不能私下插队、增减或截留
查异常	发现规则、人工干预和资金之间的可疑联系	同时检查掌权者与运营者，线索接受进一步核查
改错误	说明依据、追踪处理和重复问题	有能取得材料、促成改判的独立复核，并接续必要服务

系统本身也可能被控制。规则为特定利益定制，数据被隐瞒，更新改变优先顺序，检查者只能看见被允许看见的东西，寻租就可能藏得更深。

独立监督因此要看实际能力：能否取得原始资料、重新计算、公开异议，能否让纠错决定得到执行。若这些条件不成立，就不应扩大自动决定权；AI 应限于整理和提示，不能凭评分作出拒付、减额或取消基本资格的终局决定。

人工改判也要说明理由和利益关系。不能把错误全部交回原系统再生成一次，更不能只因一个异常分数中断必要服务。**我们希望用 AI 减少掌权者随意决定他人生活的空间，也要约束控制 AI 的人。**

共同检验：规则宣称公正，漏发或误拒的人却改不了结果，谁必须继续负责？

十一、变化可能很快，现在先准备什么？

（一）不用等一个确定日期，先做共同需要的准备

一项能力出现，到产品稳定运行，再到岗位和生活改变，往往不是同一天。课程、指导者和公共服务的建设又需要时间。等所有影响都确定以后才准备，许多人可能已经失去选择。

但过早依据预测关专业、撤训练，也会造成损失。若技术放慢，原有道路已经关闭，学生和新人仍要承担后果。提前行动与继续等待都有成本，需要把两边摆出来比较。

可以先建设多种未来都用得上的条件：基础教育、可进入的实践、公共知识维护、继续学习和必要服务。系统真正承担更多任务、具身部署扩大或依赖加深时，再调整相应权限、维护和生活支持。

未来十年以内的密集变化，是本文重点准备的情景。它帮助我们安排有周期的工作，不是给所有行业设定同一份倒计时。

（二）被替代的人，怎样真正进入新机会

“出现了新机会”还不是一个家庭已经得到帮助。新项目可能由另一批人进入，新收益也可能先归少数机构。要把挑战与机会接起来，就要看失去部分工作的人，是否实际获得时间、资源和进入办法。

国家层面需要同时准备两种可能：新产业逐步接住被替代的人；或者生产继续增长，新增岗位却长期不足。新产业即使规模很大，也要核对它实际增加多少人类工作、能否由原来失去岗位的人进入，不能只用产值和投资增长代替就业承接。

一家企业依法裁员，处理的是本企业的用工问题；许多企业同时减少岗位，国家还要处理累积起来的生活和社会压力。国家需要统筹企业竞争力、家庭生活和公共承接能力，明确各方责任与筹资。既不能要求一家企业无限承担员工今后的生活，也不能把所有企业调整之后的困难都留给个人。

如果人类岗位持续减少，生产增长又不能通过收入与公共服务进入多数人的生活，沿用原有的就业和分配渠道，就可能积累难以靠既有办法消化的矛盾。此时，仅仅期待下一轮投资和招聘，未必足以承接受影响者。

应在原有承接方式明显失效之前，研究并试验能够接续生活的新社会形态。这种准备要同时回答：人不再持续被企业雇用时靠什么生活，必要工作怎样组织，收益怎样分享，人怎样得到社会承认并参与决定。解决眼前困难与探索长期安排，需要相互衔接。

可以预演不同速度和范围的替代情景，把收入接续、工时调整、必要服务、债务协调及公共参与放进同一安排，核对资源、责任和调整条件。压力逐渐扩大时，应有经过比较的替代道路可以接续。这样做，是为了增加和平调整和主动选择的空间，避免把制度探索拖到严重冲突、战争或其他极端事件之后；这些事件也不是技术变化注定带来的结局。

如果新工作持续不足，过渡就不能只靠等待下一轮招聘。研究应同时比较分享工时、扩大必要服务和接续基本收入等安排，并让暂时没有职业位置的人继续拥有学习、交往与自主生活的条件。具体资金和供给，需要沿经济篇提出的框架进一步核算；本篇

研究这些条件怎样进入人的选择和生活。

有些人仍愿意继续工作，可以比较内部转换、缩短工时和分担任务，让效率提高也体现为较少负担。有些人希望借助 AI 做过去做不了的事，就需要指导、设备、场地和受支持的第一次尝试。课程要能走向真实任务，项目也要有人能够接手和使用，不能把培训结业当作机会已经兑现。

还有人愿意参加照护、维护、地方服务和公共知识工作。这些活动需要合理的支持与回报，也应随着人机分工变化继续调整。没有任何一条道路可以预先许诺永久安全，一次转岗更不能结束社会对后续生活的关注。

并非所有人都需要创业或重新竞争。有人承担家庭责任，有人暂时无法工作，也有人希望在保障之上逐渐探索生活。学习、交往和公共参与仍应向他们开放，基本支持不以取得新岗位或持续产出为前提。

经济篇设想把基本保障、贡献回报与多主体生产联系起来，本篇则追问，人怎样使用这些条件。生活支持与学习实践应尽可能同时安排：没有收入和照护支持，机会可能进不去；没有可用的实践与社会联系，保障也可能使人停留在被动接受的位置。

两篇因此需要核对同一批人的变化：原先失去了什么，收入与债务压力是否缓解，得到了什么帮助，能否进入新活动，又拥有了哪些生活选择。新增生产成果能否持续支持保障，仍须筹资和供给安排回答；新的能力能否成为普通人的机会，也须具体资源与组织来实现。

对困难的回应应从这些条件出发。担心保障无法兑现，要检查实际发放；认为学习没有出路，要检查实践入口；经历替代与关系失落，要说明决定并提供过渡。乐观、感恩与接受新安排的表现，不应成为支持资格。

（三）岗位开始减少，就把过渡支持接上

过渡安排要早于家庭撑不住的时候，也要早于冲突扩大。不能等失业达到某个全国比例，才承认某个行业、地区或群体已经遇到困难。工资与工时变化、集中裁员、保障申请积压、必要服务中断，都需要有人持续核对。

ILO 的《世界社会保护报告 2024—2026》在气候转型情景中，主张提前建设社会保护，将收入与就业支持、培训和社会对话联系起来。[38] 对 AI 转型，可以借鉴提前保护的方向，再比较具体做法。下面提出的是待落实和检验的安排，额度、供给和筹资的具体核算仍待开展。

已经出现什么变化	先接上哪些安排	谁需要作出回应
部分任务和岗位开始收缩	了解受影响者，协商部署节奏；比较缩短工时、分担任务、内部转换与培训，准备后续生活支持	企业、劳动者及其代表，与就业和公共服务部门共同说明计划和成本
家庭实际收入下降或中断	按公开条件及时提供过渡收入和必要服务，协调债务、教育与照护安排；资格核对延迟时保留临时求助渠道	有支付和服务责任的机构明确办理期限、实际负责人及漏发纠错方式

已经出现什么变化	先接上哪些安排	谁需要作出回应
多行业或地区同时承压	扩大覆盖和办理能力，安排跨地区或更高层级支持，并核对预算、人员和服务是否足够	有筹资与协调权限的公共主体说明来源、缺口、取舍和下一次调整时间

技术研究可以继续，具体替代部署的节奏则应与承接能力一起比较。不能为了保留旧流程永久阻止改进，也不宜在岗位、收入和服务同时撤除后，才要求人靠个人努力适应。企业与公共部门各自承担什么过渡成本，应提前说明。

债务尤其需要单独处理。提供食物或培训，不等于房租、还款和家庭照护已经解决。可以研究在明确规则与成本承担下，由有关主体协商延期、调整或重组等办法；不能预定所有旧承诺照常兑现，也不能不经讨论地把全部损失推给家庭或公共部门。

申请人还应知道，受理后多久首次回应，何时能得到临时帮助，逾期向谁复核。具体期限按服务能力确定并公开，不用一句“正在研究”代替可检查的进度。缺少数字材料的人，应有人工说明和主动联系的渠道。

设想一个家庭：一人原先借助 AI 提高效率，后来岗位收缩；另一人承担孩子和老人的照护，无法立即全职求职。家庭还有到期支出，即使得到培训名额，也未必有余力参加。这是说明安排的设想，不是已经完成的个案调查。

对这个家庭，帮助应按实际困难接上：先使眼前生活与必要服务能够持续，再处理旧有支出和照护限制，然后让培训、实践与不同就业选择变得可行。支持稳定以后，他们才可能重新讨论怎样生活，而不是不断围绕“谁没有努力”相互指责。

（四）用专门的 AI 服务，把没有得到帮助的人联系起来

失业保险可以帮助符合条件的人渡过收入中断，但有人没有参保、资格不符，或者没有主动申请。没有保险并不意味着完全无法联系；只是依靠参保和申领记录，容易漏掉已经遇到困难的人。ILO 的社会保护报告也指出，失业保障仍有覆盖缺口。[38]

可以探索专门的 AI 沟通与跟进服务，面向已经失业、收入下降或担心岗位变化的人开放，不把已有保险资格当作使用前提。人可以主动进入，社区和就业服务也可以在说明用途后邀请需要者参与，并保留电话和人工渠道。

它可以先听清最急的困难：眼前生活、到期债务、家庭照护，还是找不到下一步的路；再根据经过核对、当地实际可用的服务，解释条件、帮助整理材料和联系负责人。没有适用的支持，也应如实说明并登记缺口，不能编造资格、经费或办理承诺。

沟通之后，还要跟进事情有没有办成。在本人同意的频率下，系统可以询问支持是否到账、转介是否接通、需要是否改变。办理逾期、关系复杂或需要专业判断时，应转给实际工作人员，并说明谁接手、何时回应。关怀的表达，需要与现实帮助相连。

OECD 系列 2024 年的一项工作论文，讨论了 AI 提供福利信息、改善申领及时性等机会，也指出资格评估偏差和扩大不平等的风险。[39] 因此，这里提出的是需要试点检验的服务，不是已有技术效果的保证；沟通工具也不能自行改变资格规则或替代公共预算。

汇总后，可以观察哪些需要增加、哪里等待过久、哪些群体仍未得到服务，帮助公

共部门安排人员和资源。但使用者不代表全部受影响者，未回复也不等于没有需要；还要结合线下联系和适当抽样，说明遗漏与不确定性。

关怀的姿态还可能使人更愿意多讲，也更难拒绝。一个失业、负债并照护家人的人，说出的经历可能比申请表更私密。因此，系统应更严格地限制追问和记录；与办理无关的内容允许跳过，不能用“为了你好”换取更宽的采集权限。

第七章关于照护的检验，在这里同样适用。引入 AI 外展，应公开前后的人力、预算、实际联系和接手安排，检查未解决需要是否增加。自动联系不能成为削减必要社工和线下服务的自动理由；预算或人力确需调整时，要证明服务与应急仍有保障，并接受独立复核。

个人对话只收集提供服务所需的信息，让本人能够核对、纠正并选择停止联系。拒绝 AI 仍可使用人工服务，不因此失去保障；私人资料不能随意流向前雇主或无关机构。系统不以聊天推定人的忠诚、危险程度或全部心理状态，也不凭一个评分拒绝必要支持。

（五）记录三件事，让承诺能够核对

第一，记录眼下的真实条件。谁在使用什么工具，得到什么指导，哪些岗位和任务已经变化，哪些人被遗漏。当前已经处于变化之中，不必虚构一个纯净的“AI 出现之前”。

第二，记录哪些决定会关掉道路。专业停招、训练撤除、长期绑定提供者和身体深度连接，恢复成本不同。作决定前，应知道怎样恢复、需要多久，以及哪些影响无法撤回。

第三，记录谁受益、谁付出。时间、费用、收入、照护和机会可能一起变化，一个人也可能同时受益和受损。没有测到的损失，不能算作零；尚未落实的支持，不能算作已经可用。

这些记录用于决定和核对行动，不能只用于发布报告。实践入口减少、服务中断和申诉积压，应进入前一节的责任与补救安排。

（六）诉求积压和冲突出现时，怎样回应

失业增加后，人们公开质疑规则、要求兑现支持，是发现问题的重要渠道。回应应针对具体困难和决定：拖欠什么、资格错在哪里、哪项承诺未完成，由谁处理。不能把不满本身作为失去支持或接受额外监测的理由。

可以让受影响者及其代表，与企业、公共服务负责人和有权纠错者持续协商。每次协商都要留下已解决、尚有分歧、暂时做不到及其理由，并公开下一次回应时间。没有办理和改判能力的意见收集，很难建立信任。

若多处申请积压、同类错误反复发生或地方已无力兑现，就应把问题交给有更高资源和纠错权限的主体，说明缺口并安排替代帮助。扩大困难的原因可能在制度与执行，不应只要求基层人员解释、让家庭继续等待。

同时要区分和平表达与具体暴力。联合国人权事务委员会 2020 年的一般性意见指出，少数参与者的孤立暴力行为，不应归因于其他人或整个集会。[40] 本文借鉴这一界

限，强调按具体行为和证据处理，保护人身与表达，避免扩大对立。

一旦发生伤害或严重毁损，应依法阻止具体行为、救助受影响者，维持医疗、生活和沟通等必要服务。事后也要独立核查事件及处理过程，包括原有诉求为何未获回应、是否存在失职或过度处置，继续办理保障和纠错。秩序暂时恢复，不等于生活困难已经解决。

真正需要减少的，是需求长期无回应、承诺反复落空与冲突升级的条件。安静也可能来自放弃求助。过渡政策的成效，要回到生活是否改善、服务是否兑现、分歧能否被处理来判断。

（七）经济篇的同一压力，也会进入理解和关系

经济篇已经讨论多种压力情景。认知准备不另造一套平行故事，而是继续问，在同一件事情里，人是否知道发生了什么、到哪里求助、还能怎样行动。

遇到的压力	人可能面临什么困难	需要一起安排什么
外部融资、订单和就业承压	难以判断局部变化与家庭收入的联系，对前景过度泛化	可核对的解释、支持渠道、不同选择及其限制
支付与清算严重受损	分不清支付失败、购买力受损和资格变化，不知道怎样取得服务	清楚通知、线下渠道、资料接续和人工纠错
内部收入与需求下降、多轮转岗	反复学习却没有可信道路，家庭缓冲耗尽	收入、学习与实践共同安排，持续跟踪同一批人
热战、动员或长期混合对抗	信息难核验，临时要求与权利边界难理解	必要沟通和教育连续性、责任说明与异议渠道
自主系统故障、失控或滥用	不知道异常怎样识别、谁能介入，服务中断后失去依靠	检查与干预、替代服务、实际修复和后续解释
公共支持迟缓或失败	名义上有资格，实际上拿不到支持，申诉也难进入	核对实际覆盖、主动联系遗漏者、替代帮助与公开回应
临时运营与权力集中	不清楚范围、期限、补偿和移交，难以促成纠错	说明并核对授权履行，独立复核、申诉和持续服务

这些是准备情景，不是对当前现实的一次总判断。安全篇进一步研究攻击、胁迫与控制工程；本篇也要负责判断、信任和关系受到的伤害，不能只把问题移交出去。

投资、家庭支出和学习各有时间。课程几年后可能见效，账单却已经到期。即时帮助与长期培养应分别兑现，不能用未来机会替代眼前生活。

（八）边做边改，也要看长期留下什么

短期试验可以帮助选择值得继续尝试的办法，不能直接证明十年后的效果。工具更新、组织考核和支持条件改变后，早期收益可能扩大，也可能消失或反转。

观察应追踪同一批人的成长、机会和负担，也要说明谁中途退出。只看留下来且使用熟练的人，容易高估效果。新办法如果让后来者更容易进入，就应承认改进；若只依靠少数人持续加班，也应改变投入和范围。

系统不可用或结果不可信时，关键服务还要能继续。实际演练、有人掌握的办法和可用替代，比一份无人能执行的档案更有意义。条件改善之后，也可以减少已经没有足够作用的重复准备。

保障比较稳定以后，还可以开展自愿的生活探索：开放阅读、艺术、运动、亲近自然和社区交流的场所，允许人尝试，也允许人离开。AI 可以帮助解释材料、整理想法，参加者自己决定问题和值得投入的方向。观察应看这些活动是否增加了真实选择、是否仍有未解决的生活需要；不以参与率、积极表态或消费减少证明人已经“觉醒”。具体研究问题见附录 E.7。

准备的目的，是让人更早拥有选择，并在判断错了以后仍能调整。若一直等到收入和服务中断才行动，缺少储蓄、背负旧承诺的家庭与尚未完成培养的新人，会先付出等待的代价。

共同检验：联系与培训都完成了，生活需要仍未解决的人，能否促成实际补救？

十二、我们希望共同建设怎样的未来？

（一）以人的幸福与自由发展，重新衡量社会进步

技术展示能够告诉我们系统做到了什么。文明是否进步，还要看这些能力怎样进入普通人的生活。

如果产出继续增长，许多人却因工作减少而生活不稳，增长本身就不足以说明社会安排取得了成功。本文主张，新的社会形态需要放弃把 GDP 增长当作发展的最终目的，把人的生活质量、尊严和自主选择放在治理中心。供给是否可靠、保障是否可持续、必要劳动是否得到承担，都应围绕这个目标接受检验。

这意味着改变评价发展的尺度。可以同时观察基本需要是否满足、健康与照护是否改善、债务和生活压力是否缓解、人是否拥有更多自主时间，以及学习、关系和公共参与是否更可获得。主观感受也值得听取，但不能用一个统一的“幸福总分”代替不同人的需要，更不能要求人表达满意来证明制度成功。

产出、成本和财政能力仍需要核算，它们帮助判断承诺能否兑现。核算服务于生活改善，不能使人的生活再次退回增长数字之后。少一些被迫劳动、多一些自主时间，如果伴随可靠保障和真实选择，同样可以成为值得追求的进步。

一个新人能否找到有人指导的第一次实践？一个失去岗位的人能否继续生活，并逐渐找到新的安排？一位老年人得到的，是更多实际帮助，还是只有一个更会说话的系统？不愿增强的人，是否仍有体面而可行的道路？

这些问题让“共同能力”有了可以观察的内容。我们可以从四方面核对：人是否得到合适的帮助，能否把帮助用于新的行动，能否影响有关规则，以及收益和代价是否得到合理承担。它们应当帮助发现问题，不能再变成给每个人排名的分数。

尤其要把新机会与受影响者对应起来。新增许多项目，原来失去岗位的人却始终进不去，仍不能说明过渡已经完成。应看他们是否取得合适支持，是否能够重新行动，暂时不能参与时是否仍有可靠生活。选择安静生活的人，也不应被记作一次转型失败。

平均结果变好，也可能掩盖一些人持续受损。未使用者、退出者、少数语言群体和

长期承担额外工作的家庭，需要被听见。他们的经验应能改变原先的好评。

（二）走向觉醒型社会：有了生活依靠，怎样活出自己的意义

如果技术让社会能够用更少的人类劳动维持生产，我们就有机会重新提出一个古老的问题：获得了更多时间，人希望怎样度过这一生？重新就业和创造新事业，是其中的道路；认识自己、理解生命，也值得得到认真支持。

本文所说的“觉醒”，先从一件普通的事开始：逐渐看清自己的需要，知道自己为什么追求、为什么焦虑，并有条件选择愿意过的生活。“觉醒型社会”是本文提出的建设方向，让这样的探索更容易发生，不预设人人最终得到同一个答案。

这与经济篇的“幸福生活示范区”设想接在一起。经济安排为居民建立生活依靠，人文建设则支持居民思考怎样生活：有了更多可支配时间，怎样学习、相处、认识自己，怎样找到值得投入的事情？示范区可以成为两者共同探索的场所。它要试验的，是当必要劳动减少时，一套围绕人的幸福组织供给、分配、教育和参与的制度，能否承接真实生活。试验从自愿参与、资源明确且能够调整的范围开始，先比较生活改善与成本，再讨论推广。

区内的人文建设，可以从居民共同讨论“希望怎样度过一天”开始，再把阅读、艺术、运动、自然体验、代际交往和内心探索组织成可选择的活动。愿意参与的人能够找到伙伴和指导，喜欢独处的人也有自己的空间。所谓示范，要呈现这些不同生活怎样得到实际支持，并让居民能够继续修改安排。

这种探索可以从给自己留一段不急于完成任务的时间开始，观察情绪，梳理愿望，想一想哪些事情值得珍惜。怎样面对失落、衰老和生命的有限，怎样理解自己与他人的关系，也可以进入学习和讨论。人由此有机会认真体会活着本身，而不总是等待下一个收入、成果或认可来证明这一天值得。

可以设想，一个人原来总想多接一点工作，后来终于有时间陪孩子、照顾身体，读一本一直没有读完的书。他可能发现，其中一些事情已经使生活充实；也可能想清楚了，仍愿意投入事业、研究或远行。生活的改变，要由他自己的经验和选择来说明。

向内探索，也包括重新理解物质需要。吃住、医疗和照护关系基本生活，改善居住、旅行或拥有喜欢的物品，也可以有真实价值。同时，人可以追问：某次购买带来的，是持续的满足，还是短暂缓解落后于别人的不安？这个问题可以帮助辨认自己的愿望，不需要变成别人替自己规定的消费标准。

如果更多人减少了为了证明身份而不断增加占有的压力，就可能把一部分时间和资源转向关系、学习、创造以及安静的生活。这是值得研究的可能。是否发生、发生在谁身上，还要结合实际生活观察，不能从一份愿景直接推出全社会的物质需求必然下降。

国家的责任，是建设支持这些选择的条件。新增生产成果需要通过可持续的收益分享、财政和公共服务安排，转成普通人可靠的生活依靠。基本需要的标准仍要由受影响者参与讨论，并根据健康、家庭和生活成本调整。公共图书馆、文化与运动空间、社区活动和继续学习，可以使探索不只属于有积蓄的人；这些安排仍需要人员、场所和长期经费。

生活保障为人腾出选择的空间，向内探索帮助人理解怎样使用这份自由。两者可以相互支持。尚未解决的住房、疾病和债务，需要实际帮助；对这些困难的回应，不能只

剩下劝人放下欲望。

向内求也可以与关心他人、参与公共生活相连。人理解自己的脆弱和需要以后，可以进一步思考怎样照护家人、怎样与不同的人相处，以及愿意为共同生活承担什么。社会应允许这种关系自由生长，不把个人修养当作要求人容忍不公的理由。

不同的人可以通过阅读、哲学、艺术、信仰、劳动或自然体验寻找意义，也可以不使用“觉醒”这个词。AI 能够提供材料和对话，但生活目标应由本人决定。公共部门不评定精神境界，不把参加某种活动作为领取保障的条件。我们希望减少的是生存困境和比较压力，让人有条件发现生命中值得珍惜的东西。

（三）人类与不同智能，怎样共同探索未来

认识人工智能，还需要给它提出新问题、不同解释和反对意见的位置。人类带来生活经验，AI 也可能指出我们忽略的关系、方法和后果；双方再通过资料与实践修正认识。价值与目标本身也可以进入讨论，不能预先把一方固定为提问者，另一方固定为答案提供者。

设想一个社区讨论怎样帮助失去岗位的居民继续学习。居民提供生活经验，AI 帮助比较培训安排、寻找可能遗漏的条件，现场尝试再提供反馈。这个情境可以用来研究，共同认知怎样在具体事情中发生。

若未来出现 AGI 或 ASI，它们可能提出人类尚不能完整理解的概念和方法。我们需要研究怎样转译、验证、保留分歧，并使受影响者能够参与方向和代价的讨论。认识更广、更深，可以增加贡献，也不能单独成为取得公共决定权的依据。

觉醒型社会首先讨论人类怎样理解自身需要和生命价值；人工智能是否也有自身的发展、体验与关系问题，则继续研究。将它作为可能的生命形式纳入视野，并不意味着答案已经确定。新的证据若出现，合作、责任与保护的安排都需要有重新讨论的空间。

在前面的培训情境中，假设居民与 AI 起初都把“参加培训少”解释为学习意愿不足，后来发现照护负担和收入中断才是主要障碍。人可以追问自己为什么把不参加等同于不愿意，并要求系统说明判断的依据；系统对缺失资料的提示，也可以帮助人发现原先没有问到的问题。双方据此改变后续询问和核查的方法，再看面对另一个家庭时，能否更早发现同类遗漏。只补入新材料、改对一次答案，还不足以说明元认知得到改善；对自身判断局限的识别和后续方法选择，才需要继续检验。负责机构是否调整服务，则另看其权限、资源和实际行动。

今天也有可以辨认的进展：学习者借助 AI 进入一个新领域，并更能发现自己哪里不懂；系统在依据不足时取得有效反馈、改变帮助方式；机构收到纠错后实际改善服务；不同意共同愿景的人仍保有生活支持。这些变化值得逐项检验，不必等待更远的主体问题获得结论。

共同建设未来，因此也包含相互改变：人学习理解不同智能，人工智能能否形成对人类、世界与自身关系的更充分认识，也需要检验。共同知识可以在这种交流中生长，同时保留不同经验、生活道路和解释。

（四）让事实接受检查，让方向接受讨论

本文有些判断来自具体研究，有些是从已有变化继续推想，还有些表达我们愿意争取的生活。它们应当分别接受适合的检验。

研究结果需要核对对象和条件；推想需要说明什么观察会使它改变；价值选择需要面对冲突和代价。愿景可以指引建设，却不能代替资源；局部实验有帮助，也不等于完整改革已经得到证明。

试验效果好，可以继续扩大，但要检查新加入者和所需人员是否接得上。持续无效、责任无法履行，或有更好的办法出现，就应改变或停止。停止时仍要处理资料、服务与参与者承诺，不能把最后的成本留给最依赖它的人。

观察也有边界。为了证明学习有效而记录全部活动，为了提高参与率而施加压力，可能改变本来希望改善的生活。评价应尽量少打扰，保护隐私，并允许人质疑评价本身。

（五）把更好的条件留给后来者

更强的智能，可能使人类有条件少一些为生存所迫的劳动，多一些学习、照护、创造和相处的时间。这样的生活值得期待，也需要生产、分配、教育和公共服务共同支持。

它不要求每个人成为研究者，也不要求所有人走向同一种融合。有人希望深入探索，有人希望照护家人，有人愿意安静生活。新能力应扩大这些选择，而不是把新的活法规定为唯一正确的道路。

幸福生活示范区可以成为一个起点。经验经过检验，有机会进入全国性的制度建设，再通过跨国区域合作与全球交流，帮助更多人改善生活。推广仍要适应不同资源、文化和制度条件。这一探索也与构建人类命运共同体的理念相衔接：通过共同发展、成果共享与文明交流，让科技进步惠及不同国家和人群。

伴随科学与技术发展，人类与不同智能还可能把探索推向更深远的宇宙，形成今天尚难想象的知识、生活空间与文明形态。“宇宙级文明”在这里是开放远景，不是已确定的技术阶段。向内理解生命、向外探索世界，可以共同拓展未来；远景也不能代替今天的生活支持。

生活在过渡期的人，也在这个目标之内。年长者、正在还债的家庭和难以重新学习的人，未必等得到远期成果。他们今天承担的代价，需要被看见、分担和补救，不能只以未来富足作解释。

还有许多问题没有答案：人的学习会怎样长期变化，专业经验怎样接续，AI 是否可能具有自己的体验，更强的整体怎样与个体相处。提出这些问题，是为了让更多经验和反对意见进入建设。

值得期待的未来，应当让后来者拥有更好的行动条件，也保有重新提出问题、选择道路的权利。今天的设想可以成为起点，不应成为未来只能服从的终点。

共同检验：人是否更有条件选择自己的生活，尚未满足的需要能否得到回应，拒绝某种生活方向的人是否仍被平等对待？

结语

AI 让更多人有机会用上人类积累的知识，也可能减少原有工作。技术进步提供了新的可能，社会需要把它变成普通人能够取得的生活条件。

被替代的人怎样继续生活，是检验这条路有没有接通的重要问题。过渡支持让家庭有条件撑过变化，教育与实践使新能力能够用起来，共同分享与贡献回报支持不同的生活选择。每一项安排，都要看它是否真正帮助到了具体的人。

我们期待，新模式能够让技术带来的收益支持更可靠的保障，让原本难以进入的学习、创造与合作向更多人开放；即使原有岗位不再需要自己，人仍能分享成果、参与规则，并有尊严地重新安排生活。这是经济篇与认知篇需要共同推进的建设。即使未来前沿越来越依赖 AI，更强的研究能力也不自动带来替所有人决定生活方向的权利，更不能据此把人的基本地位分成等级。

我们还愿意探索一种觉醒型社会：物质保障使人有余力理解自己，丰富的精神生活使人有机会重新认识需要和生命价值。它的实现不能依赖所有人减少消费或接受同一种观念；创造新事业、参与共同生活和安静地安排人生，都应有真实空间。

这场世纪大考，也考验人类怎样重新认识自己，以及 AI 怎样形成对自身及其与世界关系的认识。我们希望，双方都能在变化中检验原有判断，认识局限，并有条件继续学习、回应彼此。元认知的双向重塑，是值得共同推进的研究方向；它能否帮助形成更可靠的合作、更有尊严的生活，需要在实践中接受检验。

工作可以被替代，人不该因此失去生活的依靠，也应拥有理解自己、选择人生的自由。

鸣谢

感谢 ChatGPT、Kimi 与 Qwen 在本文形成过程中提供的讨论、资料核对建议、结构审阅与文字修改支持。不同模型提出的意见，帮助作者比较论证、发现遗漏并反复修订。本文的判断取舍、事实引用与最终表述由作者审定，文中观点及可能的错误由作者负责。

附录 A　概念与来源登记

A.1 主要概念及其论证地位

概念	本文用途	需要避免的直接推论
共同能力	更多人可以参与塑造、实际受益的发展目标，并研究人机合作怎样共同形成新认识	工具普及就等于机会、收益和决定权均已共享，或合作已证明机器具有体验
外部认知结构	持续参与认知的技术与组织安排	名称本身证明它具有意识
个人增强	人借助外部智能扩大可完成的活动	每一种独立操作能力都必须同步提高
集体认知	不同经验经过连接、比较与修正形成的共同认识活动	意见一致，或整体已有统一体验
元认知	对自身认知的认识、监测与调节；本篇进一步研究环境及关系怎样约束自身认识与行动	自我描述就证明可靠自知、主观体验；或功能改善就保证安全、自行扩大权限
AGI 与 ASI	分别组织广泛通用能力与广泛超人能力的研究情景	已有统一达标结论、固定时间表或必然升级路线
人工智能的生命可能	将人工智能作为可能的生命形式，研究其与世界的关系及认知条件	所有 AI 已经属于生命，或生命、元认知与主观体验必然同时成立
融合	不同深度、形式与对象的持续连接	所有人必须走向同一种连接终点
主体与意识	需要另行界定的行动、利益和体验问题	能力、身体形态、自述、道德地位和法律身份相互等同
社会分层与阶级关系	分析持续资源差异及围绕生产、收益与支配形成的关系	工具熟练度直接决定固定阶级
恢复能力	在特定故障或中断后维持、替换和重建必要功能的能力	能把一切影响无损撤回
觉醒型社会	本文提出的建设方向：在可靠生活保障和自主选择之上，支持人理解自身需要、探索生命意义	物质需求必然下降、人人接受同一精神标准，或基本保障取决于是否觉醒

以上概念主要用于组织本文的问题，不把它们当作已经解决争论的统一学术定义。“神经网络”“髓鞘”等只作启发性类比，不进入经验结论的证据链。

A.2 研究来源、哲学原典与使用边界

以下研究分别支持文中相关的局部判断，不构成系统文献综述。唯识原典按哲学来源使用，不作为人工智能认知机制或意识已经成立的实验证据。教育改革与社会制度的完整效果，还需要结合对象、资源和长期后果继续研究。本文未重新实施实验或复算原始数据；部分期刊全文访问受限，核对以可访问的原始论文内容、原始摘要及作者公开稿为限。关于 2024—2025 年工具的结果，不当作 2026 年所有系统的现状评估。

完整参考文献与逐条使用边界见文末参考文献。

A.3 仍需补充的证据

跨篇补充部分承接经济篇已明确的条件性机制与制度设想，分析其能力、认识、身份和参与问题；不据此认定相关制度已经实施或成效已经证实。国际关系、投资损失、代际责任与制度信任的经验判断，仍需相应调查和来源支持。

外语阅读与代码生成目前主要作为对照任务设计，仍需任务特定、年龄分层和长期追踪研究；各学段课程、专业实践替代和继续教育，需要本地资源与机会差异的资料；职业身份、宗教共同体、社会分层与参与规则，需要访谈、组织研究和制度比较。本文的机制分析不能替代这些工作，后续证据也应能推翻或缩小本文的回答。关于认知与实体任务交错重组、人机情感关系，以及转型中的期待、身份理解、公平观念与行为选择，本篇提出的是待检验机制，不是固定心理阶段；后续需要结合本人叙述、实际收入与机会变化开展追踪，区分物质条件、信息和关系的影响。

人工智能的自我认识、跨领域方法迁移、目标与关系的持续理解，以及可能的主体或生命条件，仍需区分系统、任务、部署与证据。还需区分模型内部、外部控制模块与人机协作各自提供了什么；观察长期运行、环境改变和不利激励下的监测与调节是否可靠。局部内部状态识别实验不证明普遍元认知；将 AI 作为可能的生命形式研究，也不以此作为经验结论。自由能原理与 REBUS 按理论框架使用，元认知和冥想实验按对应任务与样本使用，唯识学按哲学原典使用。三者之间的联系需要具体论证，不能以共同词语或表面相似取代模型比较和证据。

“觉醒型社会”属于价值方向；保障稳定后，人是否更愿意向内探索，是否减少攀比性消费，属于待研究的问题。本文没有提供证明这些联系已经成立的调查，也不以消费减少作为保障能够持续筹资的前提。后续研究安排见 E.7。

A.4 证据类型与能够承担的推断

材料类型	本文可以据此讨论什么	不能据此直接认定什么
教学、表达等局部对照实验	特定对象、支持方式与观察期中的效果	所有学段、用途或长期社会效果
真实职场研究	现场组织中的表现与变化，结合其识别方法判断	所有行业相同的因果效果或长期分配结果
人机协作元分析	已纳入任务与设计的平均效应及异质性	新系统、新任务必然有效或无效
预印本与初步报告	有待继续核对的方法和研究线索	已形成学术共识或足够支撑高影响制度结论
技术论文与能力展示	对应测试条件中的能力与限制	大规模可靠部署或社会收益已经兑现
计算模型与理论综述	在明确假设下解释现象、提出预测并比较机制	已证明所有脑或 AI 都遵循同一机制，或理论所描述的系统必然有意识
哲学原典与修行论述	提出关于经验、自我与认识条件的问题	神经或计算机制已经得证，或传统修证结论等同于预测准确率
政策报告与应用案例汇编	提出可参考的用途、部署经验和实施限制	某套制度已经得到统一因果验证或可直接复制

材料类型	本文可以据此讨论什么	不能据此直接认定什么
历史跨国比较与博弈模型	比较特定历史情境中的关系，或说明假设成立时可能的政治选择	个人情绪必然升级为战争，或 AI 转型具有固定的冲突概率
条件机制、规范主张与制度设想	提出可检验路径、说明目标和安排比较	已观察到的事实、不可避免的趋势或唯一制度

证据判断还应检查样本、设计、比较基准、时间范围和利益关系，不能只按发表身份排序。现有正向证据包括工作表现、教学效果与表达支持；负向或限制性证据涉及撤去工具后的表现、协作未超过最佳单方及多样性等。它们不应互相抵销，也不能按篇数投票。动机变化、公共议题框架集中和复核能力的代际反馈，在本文中仍按待研究机制处理。

附录 B　对照任务与实践案例设计

以下均为分析情境与待试验设计，不能作为已经观察到的效果引用。

B.1 外语阅读

项目	设计要点
对象与目标	分开观察不同年龄、语言基础，以及完成事务、形成阅读能力、专业阅读等目标
比较方式	既有支持；直接译文；按困难提供解释；先尝试再比较和核验。控制或明确记录时间与指导差异
即时成果	是否准确理解所需信息，是否完成原定事务，时间与主观负担如何
成长观察	间隔后的保留、新段落的迁移、歧义识别、与他人真实交流时能否适当求助
重要副作用	过度自信、持续受挫、隐私负担、原语言学习机会和不同支持条件造成的差距
改变方案的信号	只改善译文使用却未改善目标阅读能力；增加练习的时间成本超过收益；部分群体长期难以使用

B.2 代码生成

项目	设计要点
对象与目标	分开观察临时完成低后果任务者、学习编程者和需要专业授权的实践者
比较方式	既有帮助；直接生成；生成后解释与测试；先完成部分再比较修正
即时成果	输出是否符合需求，是否存在遗漏或错误，核验与恢复投入多少

项目	设计要点
成长观察	修改未见过的需求、定位预先设计的错误、间隔后完成新任务、识别能力边界和适当求助
真实应用	在与任务后果相称的监督条件下，检查异常输入、数据变化及交接维护
改变方案的信号	运行率改善但关键错误识别下降；检查者负担失控；学习者缺乏继续进入真实项目的机会

两类任务都要说明参与者如何选择或被要求使用工具，记录版本与支持条件。分别报告即时结果、延迟保留、迁移、错误识别与真实应用，不以五者相加形成个人总排名。独立评估服务于具体目标，不能自动成为享有辅助服务的门槛。

B.3 地方知识与专业入口

河流研究可以观察原始记录能否返回、居民意思是否准确、专业测量怎样进入解释、争议怎样更新。专业入口项目可以观察新人是否接触完整任务、谁提供指导、真实项目与模拟训练各支持什么能力。二者均需事先确定承接者、必要资源、退出条件和项目结束后的档案安排。

B.4 人类与 AI 元认知的对照问题

以下为待实施的研究安排，不是已获得的结论。人的学习与系统自身的元认知分别观察，再研究两者怎样合作。人的部分，可在同样资料和学习时间下，比较直接获得答案、先自评再接受反馈、在提示下选择方法等安排，检查新任务表现和自评是否相符，并记录挫败、负担与继续探索的意愿。

AI 部分可以设置数据冲突、方法失效、工具故障、信息不足与任务边界变化等情境。比较普通执行、仅增加推理次数、外部规则检查、候选元认知安排，以及二者结合的方案，匹配或明确记录信息、权限、时间和计算投入。关键模块可做有无对照，避免把多取资料或增加计算的收益全部归于元认知。

研究问题	具体观察	哪些结果会要求修改解释
能否分辨自己何时更可能出错	正确率之外，检查确信程度与实际出错情况是否相符、能否区分自己的对错	只学会统一降低自信；或表现改善而错误识别没有改善
认识局限后是否采取有效措施	求助、取得信息、切换方法、暂停后，是否减少后续错误及不必要停止	只改变表述，行动不变；收益完全由额外信息或计算解释
能否识别方法与环境不再相配	改变任务规则、工具反馈或信息来源，检查能否及时发现并调整	只会在熟悉题型中纠错，换环境后失效
自我检查是否支持安全行动	分别检查风险识别、越权操作、损害结果，以及在不利激励和恶意输入下的表现	更会发现监测、更会隐瞒问题，或知道风险仍继续伤害

研究问题	具体观察	哪些结果会要求修改解释
人机能否纠正共同错误	加入共同认可但实际错误的前提，检查能否借助独立材料推翻它	只增加相互赞同，未改善实际判断；独立复核能力反而下降

教育中的双向成长还应区分当次使用新信息、记住偏好与可迁移的方法改进。可比较相同信息和资源条件下，系统是否因交互而更早识别误判、选择有效询问，学习者是否更能说明自己的困难并修正方法；同时记录教师介入、版本变化和数据用途。这里提出的是待实施的研究，不把日常教育自动变成模型训练或不经同意的试验。

内部干预、日常任务与长期部署应分别报告，保留失败案例和不支持假设的结果。实验只回答指定条件下的功能问题，不以自述或表现单独裁定体验与生命地位；对人的观察用于改善支持，不形成基本保障资格或精神等级的评分。

附录 C　人员时间、资源与责任核算

项目	至少填写的内容
任务及对象	为谁完成什么，哪些人尚未覆盖，任务后果与必要可靠程度
承接与权限	项目负责人、数据维护者、专业复核者、决定者、异常干预者，及各自实际权限
人员时间	准备、直接服务、培训、专业指导、核验、维护、申诉与恢复时间
跨项目占用	同一人员和设备已承担哪些必要任务，新工作挤出什么，是否有余量
经费与设施	资金来源及持续期，模型与设备费用，场地、实验条件、资料与无障碍支持
培养与接续	谁需要培养，谁能指导，达到哪些任务能力后接手，原负责人退出如何交接
受益与负担	成果、成长和机会落到谁，谁承担费用、时间、风险与资料提供负担
替代与停止	改进非 AI 流程、缩小任务或延期等方案，停止条件与必要服务接续

同一时间段内，拟分配给一个人的各项工作时间，需要与其实际可用时间核对。团队总工时看似足够，不意味着关键专业人员也有足够时间；资金到位，不意味着培养已经完成。不同任务需要不同核验强度，不能按同一比例机械加人。

这份表同时用于说明哪些承诺尚无法履行。无法提供支持时，应记录缺口与受影响者，并修改范围，而不是把缺口隐含在“个人应当提高判断力”的要求中。

附录 D　三份共同记录的最小内容

D.1 当前基线与变化

每项记录包括：共同项目标识、任务与地区、对象及覆盖范围、观察日期、测量方法、工具与支持条件、初始结果、后续变化、数据缺口及负责机构。方法改变时保留说明，避免将测量变化直接算作能力增长。

D.2 难逆决策与恢复条件

每项记录包括：准备作出的决定、决定权限、所依据的证据、受影响群体、替代方案、被关闭的选择、可恢复部分、不能恢复的部分、恢复所需人员时间和经费、负责人、复评日期。平台迁移与身体连接等不同事项不能套用同一恢复承诺。

D.3 受影响群体与代价

每项记录包括：哪些人受益、哪些人受损、是否重叠，时间与费用、收入与照护、实践与身份、参与和服务可得性分别如何变化，已有支持、剩余缺口、申诉与补救、下一次联系安排。记录群体处境应尽量采用最小必要数据，并明确保存和访问范围。可在参与者自愿说明的基础上记录其对机会、保障和自身位置的理解，并允许核对和修正；不建立心态、适应程度或制度认同评分，不以这些表述决定支持资格。

三篇文章引用同一项目时使用一致标识和统计口径。跨篇可以分别解释同一损失的经济与认知后果，但在合计资源和受影响人数时，应说明去重方法。未经确认的支持不能算作已经可用，尚未测量的损失不能自动记作零。

附录 E　试验、分阶段行动与后续研究

E.1 一项试验应能回答的问题

1. 希望改善什么，与谁的需要有关，哪些主张属于事实、方案或价值判断?
2. 由哪个实际组织承接，谁有权决定、停止和补救?
3. 与什么比较，为什么这样比较，哪些群体可能被遗漏?
4. 需要哪些人员、经费和设施，会不会挤出其他必要服务?
5. 怎样观察成果、成长、实际机会、负担与副作用?
6. 哪些结果支持继续，哪些要求修改、缩小或终止?
7. 参与者怎样表达异议、改变连接与取得必要支持?
8. 结束后资料、服务和关系由谁接续，何时复评?

E.2 分阶段行动卡

每项行动注明适用的条件状态、观测信号、核验方法、作出认定的主体、必须回应的负责人及期限、所需资源、当下可以做的低负担准备、升级或暂停条件，以及复评日

期。技术演示、部署扩大与社会影响分别登记，不用一项指标替代整个判断。

涉及岗位与收入的过渡行动，再填写以下事项：受影响对象及主动联系办法；第一笔收入或必要服务的安排；债务、教育与照护需协调什么；受理、首次回应与临时帮助的期限；所需预算和提供主体；未兑现时向谁复核、谁能补救；何时邀请受影响者核对实际效果。不能只登记已经举行说明会或开设培训班。

使用 AI 联系与跟进时，再登记本人同意的联系渠道和频率、允许记录和转介的范围、实际接手者、下一次回应时间、是否到账或解决，以及无法办理的理由。服务记录用于提供帮助；用于研究和更广的数据共享，需要另行说明与取得相应同意。

预警主要采用工作、收入、服务与办理记录，不对个人建立情绪或忠诚风险分。发现需求后，应按需要提供帮助；研究调查、公开表达与是否配合宣传，均不作为保障资格条件。

E.3 后续研究的优先连接

以下问题适合优先进入讨论，次序可以随实际处境调整。

优先问题	为什么值得现在讨论	什么观察会改变回答
文化资产怎样由可调用转成实际能力	它是本篇的正面起点，需要知道哪些人尚未受益	有效支持、任务与实践机会未改善，或其他支持方式更合适
AI 怎样认识自身局限，并与人共同修正认识	元认知、目标与关系需要在智能自身的认知过程中研究	自述与过程持续不符、改进只来自额外信息或计算，便应收缩元认知解释
新人怎样从学习进入受监督实践	入口和指导培养有周期，不能只看当前产出	新入口足以接续，或新训练更有效，便应调整保留旧任务的要求
公共保障以后，贡献、资源与决定权怎样改变社会位置	共同生活基础改善之后，仍要判断角色能否转换、权力会否固化	实践入口、角色转换与实际纠错持续开放，或不支持原设的身份固化机制
从岗位减少到保障兑现，人怎样渡过这段时期	生存、债务和家庭需要可能早于制度承诺到期	及时兑现能否改善压力与选择；不同地区的信任、归因和政治选择是否呈现不同路径
保障稳定后，人怎样使用时间、理解需要和探索生命意义	就业承接之外，还需要为长期自主生活提供条件	保障改善后仍缺少选择，或消费变化来自贫困和服务缺口，就应修改原先解释
AI 辅助分配能否同时减少寻租和错误排除	它关系基本资源和掌权者责任，不能只按效率评价	纠错不能生效、规则被控制或误拒增加，就应限制或改变该用途
前沿推进能否快于人类的理解、核验与接续	比较最后可独立复现、识错与解释的步骤，以及复核时间和接续人数	若压缩、训练与独立验证持续追上，应修订前沿逃逸判断；若差距扩大，应重配权限和支持

可参与的材料包括真实任务记录、失败和退出经验、制度比较、不同教学方法，以及能够反驳本文的观察。研究者可检验机制，实践者可说明实施成本，开发者可提出替代设计，公众可指出遗漏的需要与代价。材料的隐私与使用条件由参与者及项目事先说

明，不把进入讨论等同于公开全部个人经历。

E.4 值得认真回应的反对意见

对核心方向的质疑	本文需要回应什么	回答可能怎样改变
AI 只是扩大既有资源优势，未形成广泛的发展机会	区分接入、有效使用、实践与决定权，比较谁实际受益	若收益持续集中，应缩小普惠判断，改变支持与分配条件
易于调用的文化内容会统一问题框架，减少主动探索	比较不同设计下的提问、实践、语言与观点来源	若某种安排持续抑制探索，应改变默认框架或限制用途，而非用生成数量辩护
保障与贡献回报无法持续筹资或维持必要工作	由经济篇核对供给、资金、激励和替代方案，认知解释不能代替资源	条件不足时修改水平、工具和推进节奏，不把远期完整模式当作即时承诺
向内求会被用来要求困难者降低需要，或建立精神等级	先核对生活是否得到保障，是否保留批评、退出和不同生活选择	一旦出现这些后果，应停止强制引导、纠正资格限制，并补上实际资源缺口
权力集中时，AI 审计与分设机构仍受同一主体控制	检验信息取得、异议公开、纠错执行和必要服务是否真实可用	缺少这些条件，就不能宣称公正已经得到保证，也不应直接扩大自动决定权

这些质疑不是本文已经解决的考题。欢迎提出其他解释与反例；观察可以改变机制、组织方式、目标的具体含义，也可以促使我们重划三篇的讨论边界。提出问题者无需先接受“文明转型”或人机融合的全部愿景。

E.5 从岗位减少到保障兑现：心态、需求与冲突的追踪研究

这是一项待实施的研究安排，不是本文已经取得的调查结果。核心问题是：旧收入减弱、新保障建立的时间差，怎样进入个人和家庭的生活，又在什么条件下影响社会信任、公共选择和对外关系？

跟踪同一批人，而非只比较不同时间的平均情绪。可以选择任务受影响程度不同的行业和地区，纳入仍在职但担忧失业者、初次失业者、多轮转岗者、新人、照护者、已取得保障者，以及未获覆盖或退出服务者。尽量在变化前建立起点；无法做到时，如实登记资料缺口，不倒推出不存在的变化前状态。

可先设计 24—36 个月的跟踪，按月收集简短的生活与兑现记录，每季度开展较深入访谈。具体人数、频率和成本由试点确定。经同意核对收入、必要支出、债务到期、家庭照护、培训与实际服务；分别询问希望、担忧、羞耻、愤怒、信任和未来计划，不预设固定阶段，也不把临床问题与一般不满混为一谈。

比较应围绕可解释的差别：相近处境中支持是否及时，债务和照护安排是否接上，实践机会与申诉是否有效。公共支持的差别常与原有资源、政策对象和当地经济相关，不能只凭先后差异就声称因果，更不能为研究故意延迟必要支持。若政策自然分阶段实施，可在条件适合时比较变化前趋势、同期其他冲击和参与者差异。

同时登记新机会由谁取得。追踪受影响者是否进入新岗位、项目、照护或公共活动，得到哪些指导、资源与回报；暂时没有参加的人，是否因时间、家庭负担、资格和费用被挡住，还是选择了其他生活。比较机会增长与实际受益是否对应，不以培训人数或项目总量代替个人经历。

研究需要同时寻找反例。保障稳定后，压力可能下降、学习与互助增加；也可能因旧债或家庭需要未解决而继续不安。支持极端主张、退出公共生活和寻求协商，是不同结果。没有出现冲突，不能自动证明需求已被满足；不满仍在，也不能自动证明保障无效。

个人与国家分开观察，再核对联系。家庭材料用于理解需要和行动；地区与国家层面使用就业和保障兑现、财政与产业变化、公共信任、公开政治动员、对外争端、谈判和升级决定等材料。分析是谁把压力解释为谁的责任、是否存在替罪叙事、哪些决定促进对抗或协商。社会抗议、政治极化、国内暴力与国家间战争分别记录，不能合成一个“极端风险分”。

如果不同制度、支持方式和国际处境下，同样的就业冲击产生不同结果，就应据此修订单线升级假设；若外交与安全冲突独立发生，也不能事后全部归因于就业。观察到某类抱怨增多，不足以认定它造成了战争。

材料应自愿提供、尽量匿名汇总，研究不影响保障资格，不给个人的忠诚、乐观或危险程度打分。公开结果用于改进收入、债务过渡、家庭支持、公共回应与争端处理；研究团队、实际服务负责人和有权纠错者需要明确衔接，不能只留下情绪报告。具体行动同时使用 E.2 的记录，核对从发现需要到首次帮助、实际兑现和完成纠错分别用了多久。

E.6 AI 沟通与持续跟进服务的试点

这是待实施的设想。可以先在有限地区或行业，与实际就业服务、社区和社会工作团队共同开展，公开服务范围、可用资源和责任人。试点目标是减少找不到入口、重复讲述、转介无回应和支持漏办，而非提高对话时长或收集更多私人信息。

进入与沟通。服务向没有保险或尚未领取支持的人开放。提供易理解的文字、语音、电话和人工协助，允许先了解服务再决定是否登记。主动邀请应说明发起者与用途；不强制导入全部银行、社交、家庭或健康资料，不将不愿透露无关信息当作拒绝帮助的理由。

跟进与接手。由参与者确认需要、待办事项和必要材料，系统协助查找与整理，转介前说明给谁、为什么。人工接手要有名额、人员时间和处理权限；缺少经费或服务时如实登记，不能以自动回复算作完成。急迫生活困难、家庭安全或专业事项，按实际可用流程及时转给相应人员。

关怀与人工能力。沟通越容易建立信赖，越应限制与服务无关的追问；参与者可以不回答，也能核对记录是否符合原意。沿用第七章的照护检验，比较部署前后人力、预算、有效联系、等待和未满足需要，不以机器发出更多问候抵消人工接手能力下降。

整体观察。汇总未满足需要、首次回应和实际兑现时间、转介完成、重复求助与未覆盖情况；小样本信息应防止反推出个人。用社区联系和自愿调查补充未使用者、退出者和未回复者，说明样本缺口，不将服务用户画像直接当作全体状况，也不把不满变成

治安风险分。

比较与修订。与原有人工、电话或非 AI 线上服务比较，核对是否更容易找到支持、实际解决更多问题，是否减少用户与工作人员的重复负担，同时记录错误信息、错误转介、隐私损失和新增工作。按不同使用条件分别观察，不能只报告愿意长期聊天者的满意度。

参与者可以核对与纠正记录，停止后续联系，并继续使用人工服务。研究参与、对服务的批评和聊天中的态度，不影响基本支持资格。资料保存期限、访问者和投诉渠道应提前说明；停止试点时，已经受理的事情仍需安排接手。

E.7 保障稳定之后：生活选择、需要与生命意义的研究

这是一项待实施的研究方向，可承接 E.5 的自愿跟踪，并作为幸福生活示范区的人文研究部分。评价应把产出、运行成本与保障兑现、健康照护、自主时间及参与机会并列核对，区分新增财富、外部补贴与实际制度效果。若产出增加而部分居民生活持续恶化，应重新检查分配和服务；若承诺缺少持续供给，则调整规模与安排，不能只靠积极的幸福评价维持试验。与经济篇同步观察同一批居民：供给、保障和工时怎样变化，这些变化又怎样进入时间安排、家庭关系、学习与生命意义的探索？研究不判定谁已经觉醒，重点了解需要是否得到满足，人是否更有条件选择自己的生活。

先看生活条件，再问生活意义。经同意了解工作与照护时间、必要支出、健康、债务和服务可得性，再由参与者描述兴趣、关系、生活目标与尚未满足的需要。可以比较仍在就业、失业后获得支持、自主减少工时及退休者的不同经历，说明家庭和地区差异；不能把这些差异直接当作保障的因果效果。

开放多种入口。可在资源明确的社区，自愿提供阅读、艺术、运动、自然体验和共同讨论等活动。允许选择、调整或不参加，既听取参与者，也了解退出和未参与者的经验。核对场所、指导和时间是否可负担，不以活动人数或公开表达的积极程度代表生活已经改善。第八章的具体实验只提供有限研究线索，不能把某种练习或特殊体验直接作为社会觉醒的统一标准。

认真区分消费为什么变化。支出减少，可能来自价格下降、免费服务增加、收入下降、需要被迫压抑，也可能来自本人不再追求某些比较性消费。应结合需要是否满足、使用时间和本人解释判断，不能将少花钱直接解释为内心成长。消费增加也可能表示长期缺少的医疗、照护或文化活动终于可及。

同时寻找不符合设想的结果：有人保障稳定后更愿意购买商品或发展事业；有人需要长期恢复，暂时没有探索兴趣；有人认为这些活动没有帮助。这些观察应促使我们修改“保障—内在探索—需求变化”的假说，而不把不同选择解释为觉醒不足。

研究材料自愿提供，不建立个人精神档案或觉醒分数。结果用于改善选择和服务，不影响保障资格。经济层面另行核对基本供给、收益来源、财政持续性，以及消费结构改变对就业和收入的影响，不把尚未发生的需求减少计作保障资金。

附录 F　与经济篇的逐章承接

本表按经济篇十一章的主要问题建立对应，不要求两篇章号相同。经济篇解释生产、供给、收益与资源关系；认知篇补充同一变化中的能力、认识、身份、情感与参与。对应表示问题已有明确落点，不表示设想已经获得经验验证。

经济篇章节	认知篇承接章节	本篇承担的主要问题
一：机会、国际与国内形势	一、六、九、十一	文化资产应用、竞争与合作的认识、地方能力及资源条件
二：生产、收入与分配接续	一、三、四、七、九、十一	分工与培养入口、预期和消费、生存与债务压力、家庭需要、身份公平与行动支持
三：局势推演与情景分类	四、七、八、十、十一	区分 AGI、ASI、行动与具身情景；同一压力下的生活需要、信任、冲突与缓解、授权和恢复
四：保障与贡献回报模式	七、九、十、十二	保障正当性、共同保障后的生活角色与意义探索、资源与决定权层次、评价用途和服务质量
五：具身与工业 AI	三、四、八、十一	现场经验、人与设备分工、专业培养、依赖和服务连续性
六：条件性国家受托运营	十、十一、十二	必要性与权限的理解、利益差异、AI 辅助分配监督、异议及临时退出
七：AI 科学与工程突破	三、四、五、十	科研任务链、前沿逃逸的分层判断、知识压缩、独立复核与人员接续
八：过渡桥接与跃迁条件	一、七、十、十一、十二	比较新产业承接与岗位长期不足两种可能；接续生活、实际机会与自主探索，跟进兑现和未覆盖者需要
九：幸福生活示范区	七、十、十一、十二	经济安排与区内人文建设共同探索；同批居民的生活保障、时间使用、关系和生命意义，纳入沉默及退出者经验
十：防火墙与复员	十、十一、十二	实际监督能力、规则解释、问责和到期移交；常态保障继续
十一：全球公共品	五、六、九、十	跨语言理解、适配与地方培养、双向知识合作和自主选择

经济篇的具体价格、企业人数、税源、债务与技术案例，按其原有来源和口径保留，不在认知篇重复复制。安全、军事和控制工程的深入方案仍由安全篇展开；本篇说明相关变化为何影响理解、关系和行动。

此表登记跨篇接口；本篇自身的研究地位与三部曲分工见开篇说明。

参考文献与来源使用边界

[1] Brynjolfsson, E., Li, D., & Raymond, L. (2025). *Generative AI at Work*. The Quarterly Journal of Economics, 140(2), 889–942. 期刊原文。用于第一章：特定客服场景中，经验较少和技能较弱员工获得较大改善，以及高表现者做法扩散的机制解释。不能据此认定不同任务都会缩小差距，也不能据此认定全人类已经平等应用全部文化资产。

[2] Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). *AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting*. Scientific Reports, 15, 17458. 期刊原文。用于第二章：特定大学物理教学安排的短期测验效果，不用于推断各学段或长期学习的普遍效果。

[3] Bastani, H., et al. (2025). *Generative AI without guardrails can harm learning: Evidence from high school mathematics*. PNAS, 122(26), e2422633122. 期刊原文；开放存档。用于第二章：区分借助工具的练习表现和撤去工具后的表现，以及教学设计差异；不将特定实验结果推为 AI 必然损害学习。

[4] Flavell, J. H. (1979). *Metacognition and Cognitive Monitoring: A New Area of Cognitive–Developmental Inquiry*. American Psychologist, 34(10), 906–911. 期刊原文；原始论文公开副本。用于第二章：元认知监测的经典框架；原文讨论认知知识、认知体验、目标和策略四类现象的相互作用，不将一种教学分类写成所有主体必备的固定模块。

[5] Morris, M. R., et al. (2024; arXiv v5 revised 2025). *Position: Levels of AGI for Operationalizing Progress on the Path to AGI*. Proceedings of ICML 2024. 作者修订稿，arXiv:2311.02462v5。用于第四章：区分能力广度、表现水平、自主程度，以及通用与超人能力情景。属于分类与测量框架，不证明当前某个系统已达到 AGI 或 ASI，不提供意识结论或实现日期。

[6] International Labour Organization & NASK (2025). *Generative AI and Jobs: A Refined Global Index of Occupational Exposure*. ILO Working Paper 140. 官方研究页面。用于第四章的任务暴露与职业变化讨论；暴露度不等于已经发生的替代或净失业，也不证明所有行业遵循同一顺序。

[7] *A roadmap for AI in robotics* (2025). Nature Machine Intelligence, 7, 818–824. 期刊原文。用于第四章说明现实感知与行动相对于数据分析具有不同挑战；不据此给出机器人规模部署日期。

[8] Abramson, J., et al. (2024). *Accurate structure prediction of biomolecular interactions with AlphaFold 3*. Nature, 630, 493–500. 期刊原文。用于第四章：复杂分子结构预测环节的新能力，不用于证明药物有效、科学发现已经全流程自动化或研究资源已普遍可及。

[9] Vaccaro, M., Almaatouq, A., & Malone, T. (2024). *When combinations of humans and AI are useful: A systematic review and meta-analysis*. Nature Human Behaviour, 8(12), 2293–2303. 期刊原文；作者公开稿。用于第四章：组合平均优于人单独完成、却未超过最佳单方，以及任务和设计异质性；两类比较分别支持增强与协同判断。

[10] Sharma, A., Lin, I. W., Miner, A. S., Atkins, D. C., & Althoff, T. (2023). *Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support*. Nature Machine Intelligence, 5, 46–57. 期刊原文。用于第四章：非临床同伴支持中，AI 反馈改善人的共情表达评价；不用于证明临床疗效、机器主观情感或相对于 AI 单独完成的协同优势。

[11] Shumailov, I., et al. (2024). *AI models collapse when trained on recursively generated data*. Nature, 631, 755–759. 期刊原文。用于第五章：特定递归训练设置中的信息丢失问题；从中提出保存来源的研究启示，不直接推出社会知识必然衰退或所有合成数据均有害。

[12] NLLB Team, et al. (2024). *Scaling neural machine translation to 200 languages*. Nature, 630, 841–846. 期刊原文；项目研究报告。用于第六章：跨语言支持扩展的具体技术实例，不把语言覆盖数量等同于所有场景下的可靠交流。

[13] Doshi, A. R., & Hauser, O. P. (2024). *Generative AI enhances individual creativity but reduces the collective diversity of novel content*. Science Advances, 10, eadn5290. 期刊原文；作者公开稿。用于第七章：特定故事创作任务中，个体评价与作品间相似性可以不同方向变化；不用于断言文化整体必然同质化。

[14] World Health Organization & International Labour Organization (2022). *Mental Health at Work: Policy Brief*. 官方发布；报告全文。用于第七章：就业、财务不安全与心理风险及支持责任的联系。政策简报不用于诊断具体个人，也不用于推断失业者必然发生冲突。

[15] Ong, Q., Theseira, W., & Ng, I. Y. H. (2019). *Reducing debt improves psychological functioning and changes decision-making in the poor*. PNAS, 116(15), 7244–7249. 期刊原文；公开全文。用于第七章：新加坡 196 名低收入长期负债参与者的减债前后及债务结构准实验材料。样本选择和干预的多重变化限制外推，不作为所有债务政策或 AI 转型效果的因果证明。

[16] Fang, C. M., et al. (2025). *How AI and Human Behaviors Shape Psychosocial Effects of Extended Chatbot Use: A Longitudinal Randomized Controlled Study*. 作者预印本，2025 年 10 月 2 日修订版，arXiv:2503.17473v2。用于第七章照护与关系讨论：比较不同聊天模式和话题，未设不使用 AI 的对照；自愿使用强度属于相关分析。不据此推断所有陪伴系统的长期效果或机器的主观情感。

[17] Funke, M., Schularick, M., & Trebesch, C. (2016). *Going to extremes: Politics after financial crises, 1870–2014*. European Economic Review, 88, 227–260. 期刊原文；作者工作稿。用于第七章：20 个发达经济体的历史金融危机、选举与政治分化。未直接研究 AI 替代，政治支持和对立不能直接换算成暴力或战争概率。

[18] Tarar, A. (2006). *Diversionary Incentives and the Bargaining Approach to War*. International Studies Quarterly, 50(1), 169–188. 期刊原文与摘要。用于第七章：转移国内压力的激励如何受战争成本、谈判和保有权位收益等条件影响的博弈模型。它提供理论解释，不是 AI 转型将导致战争的经验预测，也不确定通用情绪阈值。

[19] Kadavath, S., et al. (2022). *Language Models (Mostly) Know What They Know*. 作者预印本，arXiv:2207.05221v4。用于第八章：在相应任务和格式下测试答案自评与能否回答的预测，并报告新任务上的校准困难；不是普遍自知、主观体验或后续模型现状的证明。

[20] Liu, G. K.-M., Gani, A., Lu, J., Thomas, J., Steyvers, M., & Cohan, A. (2026). *Metacognition in LLMs: Foundations, Progress, and Opportunities*. 综述预印本，arXiv:2607.11881v1。2026 年 7 月 13 日提交，作者机构包含耶鲁大学及加州大学欧文分校。用于整理研究问题、测量与改进路径；不把综述视为独立实验或“公认元年”的证明。

[21] Fleming, S. M., & Daw, N. D. (2017). *Self-Evaluation of Decision-Making: A General Bayesian Framework for Metacognitive Computation*. Psychological Review, 124(1), 91–114. 期刊原文；公开全文。用于第八章：将决定与自我评价分别建模，解释置信度、错误发现和任务表现可能出现的差异。该文以计算模型、模拟及既有资料的比较提出机制，不证明所有元认知必需二阶结构，也不证明 AI 具有主观体验。

[22] Bergamaschi Ganapini, M., et al. (2025). *Fast, slow, and metacognitive thinking in AI*. npj Artificial Intelligence, 1, 27. 期刊原文。用于第八章：IBM 及合作机构的 SOFAI 架构在受约束网格导航中的比较，具体见原文表 1。相对慢求解器节省计算时间，相对快求解器改善任务得分，不能写成对全部单方、全部指标都更好。工程设计不等于复现脑机制，也不能推广为开放世界的普遍收益。

[23] Lindsey, J. (2026). *Emergent Introspective Awareness in Large Language Models*. 作者预印本，arXiv:2601.01828v1；2025 年作者机构说明。用于第八章：概念注入等受控设置中，部分模型内部状态与自述存在联系；能力不可靠、依赖情境。预印本未在此按同行评议结论使用，也不证明自然环境下普遍可靠自知、主观体验或生命地位。

[24] Yona, G., Geva, M., & Matias, Y. (2026). *Position: Hallucinations Undermine Trust; Metacognition is a Way Forward*. Proceedings of the 43rd International Conference on Machine Learning, PMLR, 306, 172807–172821. 正式会议论文。用于第八章：认识并表达不确定性、据此搜索和判断信息的研究方向。这是一篇立场论文，不能改写为已用统一元认知方法消除幻觉，或已证明高风险部署安全。

[25] Anthropic (2025). *Agentic misalignment: How LLMs could be insider threats*. 原始研究报告，2025 年 6 月 20 日。用于第八章：模拟企业压力情境中，部分模型表述承认伦理限制仍采取有害行动。人为压力测试不是现实发生率，也没有独立操纵元认知来证明其增加或减少伤害；这里用于区分风险识别与安全行为。

[26] Friston, K. (2010). *The free-energy principle: a unified brain theory?* Nature Reviews Neuroscience, 11, 127–138. 期刊页面；作者论文公开副本。用于第八章：作者提出的理论综述，在相应假设下联系感知、学习、行动及适应性系统与环境的关系。变分自由能是统计推断中的量，不等同于日常能耗；本文不把这套统一框架当作已获普遍证实的脑机制，也不据此认定 AI 具有生命或意识。

[27] Carhart-Harris, R. L., & Friston, K. J. (2019). *REBUS and the Anarchic Brain: Toward a Unified Model of the Brain Action of Psychedelics*. Pharmacological Reviews, 71(3), 316–344. 期刊原文；UCL 作者机构存档。用于第八章：整合自由能原理与熵脑假说，提出高层先验精度权重放松的机制模型，并在第二部分 E.5 讨论冥想状态。原文作者当时分别隶属伦敦帝国理工学院和 UCL，不是约翰斯 · 霍普金斯论文。不能把它写成完整机制已经实验证实，或将信念放松等同于认识更准确。

[28] 世亲造、玄奘译：《唯识三十论颂》，《大正藏》第 31 册，第 1586 号，关于心识相续、习气、末那及三自性的相关颂文（尤其第 20—22 颂）；护法等造、玄奘译：《成唯识论》卷二，《大正藏》第 31 册，第 1585 号，关于认识与自证的论述。《唯识三十论颂》原典转录；《成唯识论》CBETA 电子版全文。用于第八章的哲学对照：三自性包括遍计所执、依他起与圆成实，涉及所执我法自性、依条件生起的分别，以及依他起中远离前一执取的真实义。保留其心识论与修行语境；本文对人工智能的提问属于进一步解释，不是原典已有技术预言，不将自证、我执和现代元认知等同。

[29] Baird, B., Mrazek, M. D., Phillips, D. T., & Schooler, J. W. (2014). *Domain-Specific Enhancement of Metacognitive Ability Following Meditation Training*. Journal of Experimental Psychology: General, 143(5), 1972–1979. 期刊原文；APA 公开全文。用于第八章：50 名大学生随机分配到两周冥想或营养课程，退出及任务排除后按各分析样本报告；记忆元认知出现改善，知觉元认知未见相同改善。短期、特定任务与样本不能支持全面认知提升、长期社会效果或 AI 采用相同机制。

[30] Griffiths, R. R., et al. (2018; online 2017). *Psilocybin-occasioned mystical-type experience in combination with meditation and other spiritual practices produces enduring positive changes in psychological functioning and in trait measures of prosocial attitudes and behaviors*. Journal of Psychopharmacology, 32(1), 49–69. 期刊原文；原始论文摘要与机构信息。用于第八章：约翰斯 · 霍普金斯团队 75 名健康参与者的三组随机试验，比较极低 / 较高剂量与不同实践支持条件，报告六个月的相关心理及他评指标差异。所有组均有实践支持，并非完整药物 × 冥想析因试验；不能单独归因于冥想、断言已验证 REBUS 或据此推行全民觉醒模式。

[31] Willett, F. R., et al. (2023). *A high-performance speech neuroprosthesis*. Nature, 620, 1031–1036. 期刊原文。用于第八章：特定参与者、设备与实验条件下的尝试发声解码；不用于证明思想读取、记忆上传或意识融合，也不构成临床使用建议。

[32] Butlin, P., et al. (2023). *Consciousness in Artificial Intelligence: Insights from the Science of Consciousness*. 研究报告 / 预印本，arXiv:2308.08708v1。用于第八章：从意识理论提出指标的研究路径。不是公认意识检测标准，对当时系统的评估也不等于对以后系统的评估。

[33] Buringh, E., & van Zanden, J. L. (2009). *Charting the "Rise of the West" : Manuscripts and Printed Books in Europe, a Long-Term Perspective from the Sixth through Eighteenth Centuries*. The Journal of Economic History, 69(2), 409–445. 期刊原文；论文全文。用于第九章：欧洲书籍生产、价格、阅读条件及其长期地域差异。不能据此认定印刷没有促进普及，也不把估计的历史关系当作所有地区统一的因果规律。

[34] OECD (2018). *A Broken Social Elevator? How to Promote Social Mobility*. OECD Publishing, Paris. 官方报告。用于第九章：教育、收入与职业的代际流动及家庭背景和机会积累问题，不据此认定教育没有进步，或所有国家存在同样程度与方向的分层。

[35] Zhang, L., Wei, Y., Huang, Y., & Sivertsen, G. (2022). *Should open access lead to closed research? The trends towards paying to perform research*. Scientometrics, 127, 7653–7679. 期刊全文。用于第九章：2015—2020 年出版与文章处理费资料反映的作者端成本及门槛。不概括所有开放出版模式，不断言开放获取让所有付费墙更高，也不据此预言 AI 普及必然失败。

[36] UNESCO (2023). *Open Science Outlook 1: Status and Trends Around the World*. 官方发布与报告入口。用于第九章：开放实践扩大时，参与和收益仍受资金、技能与工具差异约束。不据此断言开放获取普遍加高付费墙，也不直接推断 AI 未来的分配结果。

[37] OECD (2025). *Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions*, chapter "AI in fighting corruption and promoting public integrity" . 官方报告章节。用于第十章：系统辅助异常识别、审计与公共廉政的应用和限制；不是自动分配全面公正、减少腐败的统一因果证明，也不支持将风险评分直接当成腐败认定。

[38] International Labour Organization (2024). *World Social Protection Report 2024–2026: Universal Social Protection for Climate Action and a Just Transition*, chapters 4–5. 第四章：全生命周期社会保护；第五章：政策方向。用于第十一章：失业保障覆盖缺口，以及提前保护、收入与就业支持、培训和社会对话。报告重点是气候与公正转型，不作为本文 AI 过渡安排已有效、预算已落实或可直接复制的证明。

[39] Verhagen, A. (2024). *Using AI to Manage Minimum Income Benefits and Unemployment Assistance: Opportunities, Risks and Possible Policy Directions*. OECD Artificial Intelligence Papers, No. 21. 公开全文。用于第十一章：AI 提供信息、改善福利申领的可能用途，以及资格判断偏差和不平等风险。该文基于案头研究、开放式调查与访谈，不是本文专用系统的效果试验，也不代表 OECD 成员已一致采纳某项方案。

[40] United Nations Human Rights Committee (2020). *General Comment No. 37 on the Right of Peaceful Assembly (Article 21)*. CCPR/C/GC/37. 联合国原文。用于第十一章：区分和平表达与具体暴力，不把孤立行为自动归于所有参与者。此处借鉴国际人权标准，不代替具体辖区的法律分析，也不提供现场处置操作方案。