A BRAVE NEW WORLD · 2026

# 美丽新世界

# 2026

当认知不再属于人类

21.5万字 · 预估阅读 7.2 小时

达达尼昂

with

Claude Code

## 关于本书

2026年，AI已经像自来水一样流过每一个知识工作者的日常。这不是科幻，而是正在发生的现实。

本书试图回答一个问题：当认知不再是人类的专属能力，我们的工作、身份、关系、意义——乃至整个社会契约——将如何重构?

全书从认知垄断的瓦解写起，经过劳动契约、教育体系、公司形态、中产叙事的逐一拆解，穿越内容泛滥、算力圈地、社会摩擦的转型阵痛，最终抵达意义重建、新社会契约与行动复兴的可能路径。二十章正文之间穿插四段间奏——一个普通人在这场巨变中的微观体验。

这不是一本预言书。书中所有涉及2026年之后的推演，都是基于当前趋势的合理想象，不是对未来的预测。历史的展开方式，从来不会严格遵循任何路线图。

但有些问题，现在就值得被提出。

---

## 关于作者

曾在大厂做过多年产品经理，在上个范式下做过一些有意思的产品和业务，现在在AI方向探索。方向恰好接近本书描述的场景之一：数字人格。

必须诚实地指出，全书二十余万字，整体框架和大部分观点来自本人，但从论证推演到文字润色，都是与Claude Code全程协作完成——某种意义上，这本书就是它所讨论的那个未来的一个小型样本。

如果你对书中的内容有共鸣、有质疑、或者只是想聊聊，欢迎邮件：wanqinghui@gmail.com

或者在即刻上搜索"达达尼昂"。

# 目录

# 序章 旧世界的最后一张快照

> 最危险的变化不是你能看见的那种，而是你每天都在使用、却从未意识到它正在重新定义你的那种。

**关于本书**

本书的时间框架是2026年到2046年——从当下出发，往后推演二十年。这是一次思想实验，不是预测。没有人能预测二十年后的世界具体长什么样，我们也不打算假装能做到。我们试图做的事情更有限也更诚实：通过推演，理解变化可能的方向、机制和后果。

因此，书中出现的具体年份、统计数据、政策场景和企业案例，都是服务于论证框架的情景构建，不应被视为对未来事实的精确预测。它们的价值不在于”2034年会不会真的发生这件事”，而在于”如果沿着当前趋势推演，这类事情发生的逻辑是什么”。

此外，书中所有具名的虚拟人物均为虚构，用于呈现不同群体在变革中的处境和选择。除明确标注的历史人物和学者外，请勿将任何角色对号入座。

但有一件事需要提前说清楚：二十年的时间框架不意味着一条平滑的上升曲线。技术的突变可能在2027到2028年就基本完成——Anthropic CEO Dario Amodei的判断是，一到两年内，数据中心将诞生堪比一个国家所有天才总和的智能体。如果这个判断哪怕大致正确，那么本书真正的时间结构就不是”二十年的AI进步史”，而是”两年的技术相变加上十八年的社会消化”。类比来说，水烧到100度之后的故事不是”更热的水”，而是蒸汽时代的到来。技术的沸点可能近在眼前，我们真正关心的是沸腾之后发生的事。这意味着前几章描述的冲击不是遥远的未来推演——它们正在发生，或即将发生。本书的大部分篇幅，写的不是技术还能走多远，而是人类社会如何消化一个已经到来的事实。

---

2026年春天，一个普通的知识工作者正在经历一种奇特的认知分裂。

她每天早上醒来，AI助手已经按优先级整理好了未读消息。通勤路上，AI读完三封邮件并标注了需要关注的条款。到了办公室，她对着一份写了三分之二的竞品分析说一句”把剩下的部分补完”，十五秒后八百字出现在屏幕上，她改了两处措辞，发给了上级。下午的产品评审会上，一个实习三个月的学生提交了一份完成度不输八年资深同事的方案——方案末尾标注了”部分内容使用AI辅助生成”。

这些事情在2026年不构成新闻。它们是日常。AI已经像自来水一样流过每一个知识工作者的早晨，无声地接管了一项又一项认知微操作，以至于当事人自己都懒得注意。

但分裂恰恰发生在这里：这个知识工作者使用着已经在指令上游运作的AI，嘴里却仍然说”AI只是一个工具”。她用旧地图标注新地形，用蒸汽时代的语言描述量子时代的现实，而且她并不觉得有任何

不妥。

二十年后回望，2026年最该被记住的不是任何一项技术突破，而是这种集体性的认知延迟。几十亿人已经站在变化中间，脚下的地面已经在震动，但绝大多数人将那震动解读为”正常的技术进步”，就像1440年的抄写员将印刷机理解为”一种更快的抄写方式”。

---

“工具”这个词是一个认知陷阱，而且是一个极其精巧的陷阱——因为它在相当长的时间里是准确的。

锤子延伸手臂的力量，但不会自己寻找钉子。汽车延伸双腿的距离，但不会自己决定目的地。两百年的工业文明训练出一个根深蒂固的认知模式：工具放大人类的能力，但不替代人类的判断。这个模式如此稳定，以至于它已经成为一个不言自明的背景前提——就像鱼不需要”假设”水的存在。

2026年的AI已经越过了这条线，但语言的惯性让大多数人没有意识到越线已经发生。

考虑这样一个事实：当一个知识工作者说”帮我把竞品分析补完”，AI做的事情远不止”执行指令”。它需要理解文档的论证结构，推断作者的分析框架和写作风格，判断哪些数据在当前语境下是相关的，决定论证的推进方向，并在多个可能的表述中选择最恰当的一个。这些认知操作中的每一项，在三年前都需要一个受过良好训练的人类大脑来完成。现在它们被打包进一次十五秒的API调用里，成本低到几乎可以忽略。

但人们仍然管这叫”工具”。

塔勒布大概会说这里犯了范畴错误。“工具”这个词隐含了一个关键假设：使用者和工具之间存在一条稳定的、方向明确的控制链。人决定做什么，工具负责怎么做。从人到工具，从意图到执行，这条链的方向性是”工具”这个概念的本质特征。

但当AI开始在指令的上游运作——它开始参与定义问题和评估方案，而不仅仅执行——控制链的方向性就变得模糊了。你仍然可以说”是我决定让AI帮我分析的”，就像你可以说”是我决定去问那个顾问的”。但如果那个顾问的建议你采纳了九成，你的”决定”到底还剩多少独立判断的成分？

这个问题没有一个干净的答案，而没有干净的答案恰恰是问题所在。“工具”是一个提供干净答案的词，它把人和技术的关系简化为使用者和被使用者的关系，清楚、稳定、令人安心。但2026年的现实已经是一团模糊的灰色地带，人和AI的认知边界在每一次交互中都在被重新协商。“工具”这个比喻不是错了，而是坍塌了，因为它试图用一个二元框架去装载一个连续光谱。

名字没变，但事物的性质已经变了。就像早期的汽车被叫做”不用马的马车”——那个名字在当时不算错，但它遮蔽了一个关键事实：正在出现的东西跟马车的区别，远比跟马车的相似之处更重要。

---

彼得·蒂尔面试时喜欢问一个问题：“什么是你相信、但大多数人不同意的真相？”这个问题的力量不在于它筛选出叛逆者，而在于它揭示了一个认知规律：在任何给定的时刻，主流共识中总有一些命题是错误的，而且是重要地错误的——但身处其中的人几乎不可能分辨出哪些是。

如果我们从2046年的位置回头检视2026年，有一组广泛接受的命题后来被证明是错的。它们在当时不仅是主流观点，而且是”理性”的代名词——质疑它们会被视为要么无知，要么危言耸听。

**“AI只是效率工具。”**这是最基础的一条，前面已经拆解过。它在2026年的流行程度大致相当于”地球是平的”在中世纪的流行程度——不是因为人们愚蠢，而是因为日常经验尚未提供足够的反面证据。每天使用AI的知识工作者确实是在”使用”它，确实在”给它指令”，确实在”审核它的输出”。但这些表面现象掩盖了一个底层趋势：审核的比例在逐年下降，采纳的比例在逐年上升，而给出的”指令”越来越像”请帮我想想该做什么”而不是”请按我说的做”。从”工具”到”认知基础设施”的转变不是某一天突然发生的，它是一个连续过程，而2026年的人恰好站在那个还能自欺欺人的甜蜜点上。

**“人类的创造力不可替代。”**这句话在2026年几乎是一种信仰。它的流行有一个心理学上的原因：当体力劳动被机器替代时，人类退守到”智力”这条防线上；当常规智力劳动也被替代时，防线进一步后撤到”创造力”。这不是分析的结果，而是心理防御的结果——我们需要相信自己还有某种不可替代的东西。但”创造力”从来不是一种神秘的、不可还原的能力。它本质上是对已有元素的新颖组合——而组合这件事，恰恰是大语言模型在统计层面上做的事情。2026年的人不愿意面对这一点，因为接受它意味着放弃最后一条心理防线。到2030年代中期，当AI生成的专利申请数量超过人类、当AI创作的音乐和文学作品在盲评中持续获得与人类作品相当甚至更高的评价时，“创造力不可替代”这条防线没有被攻破——它被绕过了。人们发现，争论AI是否”真正具有”创造力是一个哲学问题；而在市场上，这个哲学问题毫无意义。

**“新技术总会创造新工作。”**这是经济学教科书里最令人安心的一条定律。蒸汽机消灭了手工织布，但创造了工厂工人；汽车消灭了马车夫，但创造了出租车司机和汽车修理工；计算机消灭了打字员，但创造了程序员和网页设计师。这个规律在过去两百年里一直有效——但它有效的前提，是每一轮技术替代只替代了人类能力谱系中的一个窄带。蒸汽机替代了肌肉力量，但没有替代手眼协调；计算机替代了计算能力，但没有替代语言理解。每次替代之后，人类还有大量未被触及的能力可以重新部署到新的岗位上。AI的不同之处在于，它不是替代能力谱系中的一个窄带，而是以宽频方式同时侵入整个认知能力谱系。当几乎所有的认知能力都面临替代压力时，“人类可以转去做新的工作”这个承诺就需要回答一个追问：转去做什么？用什么能力来做？如果你能回答这个追问，你可能比2026年几乎所有的经济学家都更了解未来。

**“教育是最好的对冲。”**这是2026年中产阶级的集体镇静剂。逻辑链条看起来清晰而坚固：AI替代低技能工作，那就提升技能；提升技能靠教育，所以接受更好的教育就安全了。问题在于，当AI开始替代的不是”低技能”而是”技能本身”时，整个逻辑链条的根基就松动了。教育投资的回报率建立在一个隐含假设上：通过学习获得的认知能力是稀缺的，因此有市场价值。但如果同等甚至更高水平的认知能力可以通过API以近乎零的边际成本获取，那么教育投资的经济回报率就不得不被重新计算。这不意味

着教育本身没有价值——它意味着教育的经济价值和内在价值正在脱钩，而2026年的大多数家长在为孩子规划教育投资时，依据的仍然是经济价值。

**“这次和以前一样。”**所有安慰性命题的总纲。它是归纳法的产物：过去每一次”这次不一样”的预言都错了，所以这次也不会不一样。这个推理在统计上是合理的——直到它遇见一个真正的范式转移。每一次文明跃迁都伴随着一个残酷的发现：上一个时代最珍贵的能力，在下一个时代一文不值。问题从来不是”变化会不会来”，而是”你能不能在变化到来之前意识到这次的变化是范式级的”。2026年的绝大多数人没有意识到。不是因为信号不够多，而是因为承认信号意味着承认自己的人生规划建立在一个正在消解的假设之上——这对大多数人来说，是一个无法承受的认知成本。

---

比喻讲够了。回到第一性原理。

人类文明的全部制度——经济的、政治的、教育的、法律的——建立在什么前提之上？

可以有很多答案，但有一个答案如此基本，以至于它几乎从未被明确说出：**认知垄断**。

准确地说，是这样一个事实：在这颗星球上，只有一个物种能够进行抽象推理、语言处理、因果分析、长期规划和知识积累。这不是人类的”优势之一”——这是人类文明存在的前提条件。所有的制度设计，无论多么精巧，都是在这个前提条件下展开的。

考虑经济制度。劳动力市场的存在基于一个简单的事实：认知能力被封装在人类大脑中，不可复制，不可批量生产，获取成本高昂（需要十几年的教育和训练）。因此认知劳动是稀缺的，稀缺产生价值，价值产生收入，收入支撑消费，消费驱动经济。整个循环的第一块砖是”认知不可复制”。

考虑教育制度。学校的存在基于一个更底层的事实：认知能力只能通过缓慢的、个体化的学习过程来获取。你不能把一个人的知识直接拷贝到另一个人的大脑里。因此我们需要教室、教师、教材、考试——一整套工业化的认知传递基础设施。这套设施从普鲁士时代延续至今，核心假设从未改变：认知是一种需要被”生产”出来的东西，而生产过程耗时、耗资、一次只能生产一个。

同样的逻辑延伸到政治和社会分层。代议民主假设公民无法对每个议题形成充分认知，所以需要代表；社会等级假设认知能力存在有意义的个体差异，所以分层有合法性。底层假设都一样：认知是稀缺资源，必须通过制度分配。

现在，想象这些制度的设计者——从亚当·斯密到杰斐逊到俾斯麦——在设计这些系统时，在一张白纸上列出”不需要论证的基本假设”。认知垄断一定在那张纸上。不是因为他们思考过这个问题然后决定把它当作假设，而是因为这个事实太过基本，根本不需要思考。就像你不需要假设”明天太阳会升起”一样——在这颗星球上，认知就是人类做的事情。

这个前提从未被质疑过，因为从来没有必要质疑。

直到现在。

而且，正在打破这个垄断的东西，比大多数人以为的更加陌生。我们习惯用人类类比来理解AI——“更聪明的助手”“不知疲倦的员工”“超级实习生”。但这些类比本身就是旧地图。Dario Amodei在一次深度访谈中指出，AI的学习方式”既不同于人类学习，也并非复刻人类进化，而是介于两者之间”——一种人类认知谱系中从未出现过的新事物。这不是语义上的吹毛求疵，而是一个根本性的认知校准：我们面对的不是一个”更强的工具”，而是一种全新的认知存在。这意味着，不只是前面拆解的那些社会制度的旧地图过时了——我们用来理解AI本身的旧地图，同样需要重画。本书讨论的真正问题，不是一个超级工具如何冲击人类社会，而是当一种全新的认知物种出现后，人类如何重新理解自己在这颗星球上的位置。

人类花了三十万年垄断认知，然后在十年内把它变成了公共基础设施。这不是一句修辞。2020年代中期，高质量的认知输出——分析、推理、写作、规划、判断——的边际成本正在以指数速度趋近于零。一个能够在大多数认知任务上达到专业人类水平的AI系统，其每次推理的成本以美分计价，而且每年都在以数量级的速度下降——第六章将详细拆解这条成本曲线。

这意味着什么？它意味着建立在认知稀缺之上的全部制度安排，突然之间失去了它们的前提条件。不是理论推演——正在发生。

当认知可以被复制、可以被批量生产、可以以近乎零的成本获取时——

劳动力市场的定价机制还能正常运作吗？如果认知劳动的稀缺性消失了，知识工作者的市场价值靠什么来支撑？

社会分层的合法性叙事还站得住吗？如果认知能力不再是个体差异的有效区分维度——因为所有人都可以接入同等水平的认知基础设施——那么基于认知贡献的分配逻辑还能为不平等提供辩护吗？

这些不是修辞性的反问。它们是结构性的追问。每一个问题背后都对应着一套已经运行了几十年到几百年的制度体系。这些制度不会在一夜之间崩塌——制度有惯性，有既得利益集团的维护，有路径依赖的自我强化。但前提条件的消解是一种缓慢而不可逆的腐蚀。制度可以在前提消失后继续运行一段时间，就像一座大楼在地基被掏空后还能靠结构惯性站立一段时间。但每个人都知道接下来会发生什么。

---

如果你试图理解未来二十年将要发生的事情，不妨借用达利欧的”机器思维”——把文明当作一台可拆解的机器来理解，可能是最有用的框架。

把人类文明想象成一台机器。这台机器有三个组成部分：

输入端是技术——人类理解和改造世界的能力。从石器到青铜器，从印刷术到蒸汽机，从电力到互联网，技术是驱动这台机器运转的能量来源。技术的每一次重大升级都会改变机器的运行参数。

处理机制是制度——经济制度、政治制度、教育制度、法律制度、社会规范。制度的功能是把技术提供的能力转化为可以被组织和分配的社会秩序。制度是机器的齿轮和传动轴。

输出端是秩序——谁拥有什么，谁决定什么，谁的生活过得怎样。秩序是这台机器的最终产出。

在稳定时期，这三个部分是协调的。技术、制度和秩序达成了一种动态平衡——不是完美的，但是可运行的。人们在这种平衡中建立预期、规划人生、做出投资决策。这种平衡就是我们平常所说的”正常”。

但技术升级和制度调整的速度是不匹配的。技术可以在几年内发生量级变化——从GPT-3到GPT-4到GPT-5，每一步都是认知能力的大幅跃升。而制度的调整需要几十年——法律需要立法程序，教育需要课程改革和师资培训，劳动力市场需要技能转型和社会保障体系的重建，政治制度需要新的合法性叙事和权力再分配。

这个时间差——技术已经变了但制度还没跟上——就是我所说的”制度拖曳”。它不是新鲜事物。工业革命时期就出现过：蒸汽机和工厂制度在18世纪末期已经成熟，但现代劳动法、社会保障、公共教育等制度要到19世纪中后期才逐步建立。中间那几十年——童工、十六小时工作日、城市贫民窟、卢德运动——就是制度拖曳的代价。

AI引发的制度拖曳将比工业革命时期更剧烈，原因有三。

第一，速度。蒸汽机从发明到大规模应用用了大约八十年。AI从实验室到全面渗透知识工作只用了不到十年。制度调整的速度没有相应提升——民主程序的节奏、教育改革的惯性、社会共识的形成速度基本没变。这意味着技术和制度之间的时间差更大了。

第二，广度。蒸汽机主要影响了体力劳动和制造业。电力和计算机主要影响了信息处理和通信。AI影响的是认知本身——而认知是几乎所有人类活动的共同基底。这意味着需要调整的不是一两个制度，而是几乎所有制度。

第三，深度。更根本的区别在于，AI动摇的不是做事方式，而是做事资格。以前不管技术怎么变，拍板的始终是人。这一次，连”拍板”本身都成了可自动化的操作。这不是程度的变化，而是性质的变化——它要求的不是制度的修补，而是制度前提的重新审视。

未来二十年一切混乱的根源就在这里：制度来不及适应。假设在被推翻，新秩序还没影子。在旧秩序和新秩序之间的那段空白里，会有混乱、焦虑、不平等加剧、社会契约断裂、政治极化、以及无数个体命运的剧烈颠簸。

这不是悲观主义。这是机器运行的规律。每一次重大的技术输入变化都会引发制度调整的延迟期，延迟期都是痛苦的，最终制度会完成调整，新的平衡会建立。真正的问题是：在好起来之前，要经历什么，要多久。工业革命的制度拖曳期持续了大约半个世纪，其间诞生了马克思主义、两次工人运动浪潮和整个现代福利国家。AI时代的制度拖曳期会产出什么，取决于我们对正在发生之事的理解速度。

这就是这本书试图做的事情：加快理解。

---

最后一件事。

这本书讲的不是AI会变多强。关于AI能力的预测已经够多了，而且绝大多数预测——无论乐观的还是悲观的——都不影响核心论点。核心论点是：

真正的断裂不是机器变聪明了，而是聪明变便宜了。

当认知从稀缺资源变成公共基础设施，建立在认知稀缺之上的全部秩序——经济的、教育的、政治的、社会的——都将被迫重组。这个重组过程已经开始了。它不会等待我们准备好。

2026年的那个知识工作者，她正站在一幅地图正在重绘的世界里，手中拿着的还是上一版地图。她不是不够聪明——她只是恰好生活在旧地图还能用、但即将不能用的那个短暂窗口期。

这本书写给所有正站在那个窗口期的人。目的只有一个：让你看清地形。

这个能力，至少在目前，还需要你自己来完成。

# 第一章 认知垄断：人类最后的特权

> 三十万年来没有人质疑过这个前提，不是因为它经受住了质疑，而是因为根本没有质疑的必要——直到必要突然出现。

---

2026年3月，一个在律所做了八年尽调的初级合伙人，发现自己新带的实习生用AI在四十分钟内完成了他当年需要三天才能完成的工作。他看着屏幕上那份几乎无可挑剔的报告，第一个念头不是恐惧，而是困惑：如果这事儿不再需要三天，那我这八年练出来的东西算什么？

序章讲了一个人的早晨和傍晚。她用旧地图导航，地形已经变了。她失去的不是工作，而是确信感——觉得世界会沿着自己理解的方式运转下去的那种感觉。

现在我们要问一个更冷的问题：那张旧地图是什么时候画的，画在什么上面，以及为什么三十万年来从来没人想过它会过期？

---

## 一、三十万年的隐含前提

赫拉利在《人类简史》中做过一次知识考古：智人主导地球，是因为我们能创造并相信”虚构故事”——国家、货币、公司、法律、宗教，这些虚构秩序是人类文明的操作系统。这个论证已经成为常识。但赫拉利没有追问一个更致命的问题：这些虚构秩序之所以能运转，是因为只有人脑能处理维持它们所需的复杂信息。三十万年来，人类垄断了地球上全部的认知处理能力——这个事实如此理所当然，以至于从未被当作一个”前提”来审视。

让我们用”机器思维”来重新审视这个前提。如果把人类文明看作一台机器，这台机器的运转逻辑是：

首先是**信息输入**，感知环境、收集数据；然后经过**信息处理**，分析、推理、判断；由此产生**决策输出**，法律、政策、商业行为；最后是**执行与反馈**，行动、评估、修正。

在这台机器里，人类大脑不是某一个零件，而是**唯一的通用处理器**。从最底层的数据录入员到最顶层的国家元首，从税吏到法官，从会计到CEO——每一个节点都是人脑。机器的其他部分——纸张、印刷机、电报、互联网——只是扩展了信息的传输速度和存储容量，但处理环节始终是人类大脑的专属领地。

这就是”认知垄断”的准确含义：不是说人类比其他物种”更聪明”（这是一个模糊的说法），而是说**人类是文明这台机器里唯一能承担信息处理节点角色的物种**。所有的制度设计、激励机制、教育体系、社会分层，都是围绕这个垄断地位展开的。

教育的本质是什么？是把原始的人类大脑加工成合格的信息处理节点，然后插入文明机器的对应插槽。小学教识字和算术（基础信息处理能力），中学教学科知识（专业化处理能力），大学教深度专业技能（高级处理能力），研究生院教前沿研究方法（创造新信息的能力）。学位不是知识的证明，而是这块”人脑芯片”已通过质检的信号。

薪资结构的本质是什么？是按照信息处理的复杂度和稀缺性定价。数据录入员处理最简单的信息，收入最低。律师处理法律信息，医生处理医学信息，金融分析师处理市场信息——复杂度更高，供给更稀缺，价格更高。CEO的价值不在于他个人的信息处理量，而在于他处理的是”不确定性最高的战略决策”——这种信息处理的供给极度稀缺。

社会地位的本质是什么？很大程度上就是认知稀缺性的映射。医生、律师、教授受尊敬，不是因为他们的道德更高尚，而是因为他们大脑里装载的认知程序更复杂、更难复制。“读书改变命运”这句话的真实含义是：通过长期教育投资升级你的信息处理能力，从而占据文明机器里更稀缺的节点位置。

三十万年来，这套逻辑从未失效过。每一次技术革命都在改变信息的传输和存储方式——文字、印刷术、电报、互联网——但处理环节始终是人类大脑的禁脔。文明的制度越来越复杂，但复杂的方向始终是同一个：创造更多、更细分、更专业的人脑信息处理岗位。

直到2020年代。

---

## 二、AI打破的到底是什么

先说清楚AI**没有**打破什么。

AI没有获得意识。它不理解它在做什么。它没有欲望、恐惧、动机或价值观。2026年最先进的大语言模型仍然是一个极其精密的模式匹配系统——它在海量文本数据中学习到的统计关联让它能够生成看起来像人类思考产物的输出，但”看起来像”和”是”之间有一条本质的鸿沟。

关于AI是否”真的在思考”的哲学争论，对于理解正在发生的事情毫无帮助。重要的是AI能不能**以极低的成本完成过去需要人类认知才能完成的任务**。答案已经不需要预测：在标准化认知领域，它已经可以了。

这个区分至关重要。让我们把人类认知劳动拆解为三个层次，然后逐层分析AI的冲击路径。

**第一层：标准化认知——边际成本趋零。**

合同审查、财务报表分析、法律法规检索、专利文件比对、医学影像初筛、翻译、数据录入、格式化报告生成——这些任务有明确的输入、明确的处理规则和可量化的输出标准。它们是认知劳动中的”流水线工作”，区别只是流水线上处理的是信息而非零件。

一位初年级律师做一份标准尽调，包含薪资、社保、办公、管理、培训成本，每工时约500–800元人民币。AI完成同样任务的推理调用费在2026年已低于10元/次，每六到九个月再减半。Sam Altman在2025年初给出了一个更惊人的数据：同等水平AI的使用成本每12个月下降约10倍——这是摩尔定律速度的五倍以上。这是两到三个数量级的成本崩塌，而且崩塌还在加速。

当一项认知任务的边际成本从数百元降到个位数，市场的反应不会是渐进式的线性调整。先是最价格敏感的中小企业切换到AI法律工具，然后是大型律所发现初级律师的利用率持续下滑，然后是法学院招生人数连年走低。每一步都不像灾难，但每一步都在压缩同一个空间——而且这种压缩是不可逆的。没有人会因为”怀旧”而回到每份合同花三天人工审查的时代。

**第二层：复杂推理——天花板被压低。**

罕见病鉴别诊断、多因子金融建模、企业战略情景推演、复杂诉讼策略设计——这些任务的输入高度不确定，处理规则不完全明确，需要整合多种信息源进行非线性推理。

AI在这个层面的表现呈现出一种特殊的模式：单步推理已达资深专家水平，但多步链式推理的可靠性随步数增加而衰减。一个单步准确率95%的模型，执行十步任务的端到端成功率可能只有60%。这意味着”AI全面接管复杂知识工作”在短期内仍需人类监督。

但”短期内”到底是多短？2025年初，Anthropic CEO Dario Amodei给出了一个让整个行业都感到紧迫的判断：1到2年内，数据中心将诞生堪比一个国家所有天才总和的AI系统。如果他是对的，本章描述的一切不是即将到来，而是正在发生。以这个速度推算，复杂推理层面的天花板压低——从”AI辅助”到”AI主导、人类审核”——可能只需要2–5年，而非大多数人下意识预估的更长时间。

“仍需人类监督”和”仍需同样多的人类”是两码事。当AI把复杂推理的前八步都做完、只剩最后两步需要人类判断时，一个资深分析师的价值定义从”能做出别人做不出的分析”变成了”能在AI出错的边缘情况中做出正确判断”。后者需要的人数远少于前者。十五个分析师的工作量被压缩为三个人的审核量——不是消失，是坍缩。

**第三层：判断性认知——暂时安全但边界在移动。**

法官在两个都合理的法律解释之间做出选择。CEO在极度不确定的环境中决定战略方向。立法者在效率与公平之间权衡政策。这些决策涉及价值判断、伦理权衡、风险偏好、文化直觉——无法被还原为信息处理。

这个层面暂时是安全的。但”暂时”是一个需要认真对待的词。五年前人们认为”写一首能打动人的诗”属于判断性认知，2025年的大语言模型已经能写出让多数读者无法区分出处的诗歌。三年前人们认为”评估创业项目的潜力”需要风投特有的商业直觉，2026年的AI投资分析工具在回测中跑赢了大多数基金。安全边界不是固定的城墙，而是在持续后退的海岸线。

现在，这三个层次的叠加效应才是真正需要关注的——而大多数分析都在这里停下来了。

但大多数人对”替代”的理解是错的。

公众最常犯的错误是把”AI能做90%的工作”等同于”90%的人将失业”。这两者之间隔着五个层级，每一层到下一层都是一次质的跳跃，而非简单的量的递增。

第一层：AI完成90%的具体操作——写代码、起草文件、生成分析报告。第二层：AI完成100%的具体操作——不再需要人类动手。第三层：AI完成90%的端到端任务——不仅执行，还包括理解需求、验证质量、处理异常。第四层：AI完成100%的端到端任务——从接收需求到交付成果，全程无人介入。第五层：从业者数量大幅减少——市场最终将过剩的人力挤出。

我们正在快速穿越这些层级，但在任何一个时间点，不同行业处于不同的层级——这解释了为什么有人已经被深度冲击，而有人还毫无感觉。翻译行业已经逼近第四层，而外科手术还停留在第一层的边缘。用同一个”AI替代论”去描述这两个行业，只会制造恐慌或虚假的安慰。

更准确的说法不是”替代”，而是”重组”。每一层都在重新定义”这份工作到底是什么”——当AI做完了90%的具体操作，剩下的10%不是原来工作的缩小版，而是一份全新的工作。理解这五层区分，是理解后面所有讨论的基础。

更大的问题是：变化不是匀速的。

大多数人的心智模型是线性的：AI先替代简单工作，然后慢慢替代复杂工作，人类逐步后退到更高层次的岗位上。这个模型暗含一个假设——变化是渐进的，社会有时间适应。

但复杂系统的核心规律是：变化不是渐进的，而是非线性的。它们会经历相变（phase transition）——系统在量变阶段看起来什么都没发生，然后在某个临界点突然发生质变。温度从90°C升到99°C只是”更热了”，但从99°C到100°C不是”又热了1度”，而是水变成了蒸汽。

认知劳动市场正在逼近这样一个相变点。当足够多的标准化认知被自动化——不是50%，可能只需要30%——系统就会出现连锁反应。还记得间奏一中那个西二旗的程序员吗？他每天写的代码越来越像填表，那种”我在造东西”的感觉在消退。他还不知道半年后AI工具会在三十秒内生成他一天的工作量，但他身上已经在发生的事情，正是这种连锁反应的微观预兆。因为标准化认知是复杂认知的训练基础。初年级律师通过做尽调来学习法律实务，初级分析师通过写行业报告来建立商业直觉，住院医师通过大量阅片来培养临床判断力。当这些”训练阶段”的工作被AI接管，高级岗位的人才管道就被切断了——你很难通过”审核AI的报告”来获得与”亲手做三千份尽调”同等的认知深度。

这意味着什么？意味着标准化认知的自动化不只是消灭了入门级岗位，它还腐蚀了高级岗位的根基。这不是线性后退（AI做简单的，人做复杂的），而是从底部开始的结构性坍塌——底层消失后，中层失去训练来源，顶层失去后备力量，整个金字塔的稳定性发生质变。

这才是”打破认知垄断”的真实含义。不是某个职业被替代，而是”人脑作为文明机器唯一通用处理器”这个三十万年的前提被动摇了。文明机器开始拥有一种不依赖人脑的信息处理选项——它还很粗

糙，还不完全可靠，但它存在了。而一个选项的出现，足以改变所有基于”无选项”这个假设而建立的制度。

---

## 三、稀缺性转移——什么变得真正值钱

蒂尔在《从零到一》里区分了两种截然不同的事：0到1的创新与1到n的复制。这个框架被广泛引用，但很少被真正用对。大多数人在思考AI对就业的影响时犯的错误，恰好是把一个0到1的问题当作了1到n的问题来处理。

他们问的是：AI会替代哪些岗位、创造哪些新岗位？——这是1到n的思维，在旧框架内做加减法。更有穿透力的问题应该是：当认知变廉价了，什么东西还值钱？

稀缺性不会消失，它会转移。工业革命让体力不再稀缺，但它没有消灭稀缺性——它把稀缺性从肌肉转移到了大脑。同样的逻辑，AI革命不会消灭稀缺性，而是把稀缺性从认知转移到别处。

转移到哪里？

最廉价的回答是”创造力”。这个回答之所以廉价，是因为它本身就是一种标准化认知的产物——它是大多数人在面对”AI替代人类”这个问题时的默认输出。真正的稀缺不会这么容易被识别，否则它就不稀缺了。

**第一个真正的稀缺：定义问题的能力。**

AI能回答问题，但”问什么问题”本身变成了最稀缺的能力。这不是文字游戏。在信息处理成本极高的旧世界，光是”回答问题”就需要大量人力——律师花三天查法规，分析师花两周写报告，医生花半小时阅片。问题的质量不太重要，因为回答问题本身就是瓶颈。

当回答问题的成本趋近于零，瓶颈突然移到了上游：你在问正确的问题吗？AI可以在十秒内生成一份深圳房地产市场的分析报告，但”应该分析深圳房地产市场吗”这个问题本身——它预设了什么假设、忽略了什么替代方案、服务于什么战略目标——才是真正创造价值的环节。

这就是0到1的真正含义：不是在已知框架内做更好的分析，而是定义一个新的框架。AI极大地加速了1到n（在给定框架内高效执行），但0到1——看到别人看不到的问题——仍然是人类特有的能力，而且因为1到n被AI加速了，0到1的相对价值反而飙升。

**第二个真正的稀缺：可问责性。**

人类社会需要有人为决策承担责任。一个AI可以建议法官判被告无罪，但判决书上需要一个人类法官的签名。一个AI可以推荐最优投资组合，但基金经理的名字出现在季报上。这不是因为人类比AI更聪明或更可靠，而是因为”责任”是一个社会建构——它需要一个可以被质疑、被追究、被惩罚的主体。

算法没有skin in the game。这是一个关键缺陷。一个推荐错误药物的AI不会失去执业资格、不会面临诉讼、不会在半夜被愧疚感惊醒。“谁来承担后果”这个问题，在AI时代不是变得更不重要了，而是变成了一个核心的制度设计问题。可问责性的稀缺将重塑职业结构——未来高薪岗位的核心定价因素可能不是”你能分析什么”，而是”你愿意为什么签字”。

**第三个真正的稀缺：在场感与信任。**

AI可以生成比任何人类律师更全面的法律意见书。但当一家公司面临生死攸关的并购决策时，CEO不会只看一份AI报告就拍板——他需要一个他信任的人坐在他对面，看着他的眼睛说：“我看了所有的材料，我认为可以做。”这个”看着眼睛”的动作不包含任何AI无法处理的信息，但它包含了AI无法提供的东西：一个人把自己的判断和声誉押在了这个建议上。

信任的建立依赖于共享的时间、重复的互动、可观察的行为一致性——所有这些都是物理世界中的稀缺资源。AI可以在一秒内生成一万条建议，但一个人一天只能与有限的人建立深度信任关系。在一个信息无限廉价的世界里，信任反而变成了终极的稀缺资产。

**第四个真正的稀缺：意义感。**

这是最不显眼但可能最深远的转移。当AI能做大多数认知工作时，“我为什么要工作”这个问题不再有显而易见的答案。过去三十万年，人类的意义感很大程度上来自”我能做别人做不了的事”——我是律师，我能解决法律问题；我是医生，我能治病救人。当这些能力不再稀缺时，意义感的来源需要被重新定义。

这归根结底是一个制度问题。一个找不到意义感的个体会抑郁。数亿找不到意义感的个体会动摇社会结构。

---

## 三（续）、从认知到物理——智能的延伸

稀缺性的转移不只发生在认知领域内部。一个更深层的变化正在发生：AI正在从”处理信息”延伸到”操控物理世界”。

当机器人可以根据软件指令操控物理对象时，物理世界的一部分就变得”可编程”了——正如1995年互联网让信息变得可复制、可搜索、可重组，具身智能正在让物理操作变得可版本控制、可远程部署、可A/B测试。一个在深圳工厂里被优化的焊接策略，可以在24小时内被部署到全球所有兼容工厂。制造业know-how的传播速度从”年”变成了”天”——这个变化的意义可能不亚于印刷术对知识传播的加速。

这意味着认知垄断的打破不会止步于”信息处理”层面。当认知自动化与物理自动化合流，形成感知-决策-执行-反馈的全闭环时，AI就不再只是文明机器里的一个新型处理器——它开始拥有”手脚”，

能够独立作用于物理世界。三十万年来人类不仅垄断了认知，也通过认知垄断了对物理世界的有目的操控。这两项垄断正在同时被动摇。

但”同时”不是”同速”。物理世界的自动化受制于工程成本、科学瓶颈和物理定律约束（能源密度、执行器功率、散热等），其进展呈阶梯形而非线性。这种非线性——长期平台期之后的突然跃迁——使得预测极不可靠，也使得制度回应更加困难。具体的阶段和时间线，将在第七章详述。

---

## 四、制度拖曳——河床决定水流

现在让我们换一个视角。前面三节讨论的是技术冲击本身——认知垄断是什么、AI打破了什么、稀缺性转移到了哪里。但技术从来不是在真空中运行的。它降落在制度上，而制度不是被动接收者——制度是河床，决定了水往哪里流。

理解制度拖曳，“机器”隐喻仍然有用：输入进来，经过一系列齿轮和传动机构，产生输出。如果你想改变输出，不要盯着输出看，要去检查机器的内部结构。

同样的道理适用于AI对社会的影响。AI的扩散速度和扩散路径不取决于AI本身的能力，而取决于接收它的制度结构。不同的制度有不同的惯性——我用”制度拖曳系数”来描述这种差异。

**拖曳系数最低：个人。** 适应周期1–2年。一个人决定使用AI不需要任何人的同意。序章中那位知识工作者早上让AI补完竞品分析，这个决定只需要她自己点一下鼠标。个人采纳AI的摩擦力几乎为零——唯一的障碍是认知惯性和学习成本，而这两者都在迅速降低。

**拖曳系数中等：企业。** 适应周期2–5年。企业采纳AI需要改变组织架构、工作流程、激励机制、考核标准。这些改变涉及权力的重新分配——谁管谁、谁决定什么、谁的利益被触动。

律所的合伙人制度是一个经典案例。锁步制把合伙人收入与客户计费绑定，客户计费与工时挂钩。AI把尽调时间从三天压缩到四十分钟，理论上项目报价应该下降——但那意味着合伙人收入下降。没有合伙人会主动推动一项让自己收入下降的变革。结果是行业普遍采取”内部提效、外部维价”的策略：用AI加速内部流程，但对客户仍按传统方式报价。效率红利被截留在组织内部，而不是传导到市场。

医院的等级制度展现了另一种拖曳模式。AI在放射科渗透最快——因为影像诊断接近标准化认知，且放射科在医院权力结构中处于辅助地位，科室主任把AI视为提升话语权的武器。而外科是医院权力金字塔的顶端，外科主任没有任何动力引入可能动摇自身权威的技术。同一家医院，同一个技术，因为制度结构不同，采纳速度可以相差五到十年。

**拖曳系数高：法规。** 适应周期5–10年。法律法规的制定和修改需要经过立法程序、公众讨论、利益博弈，而且法律天然追求稳定性——频繁修法本身就是对法治的损害。2026年，全球主要经济体的AI监管仍处于框架讨论阶段。欧盟的AI法案是最领先的，但它的分类标准基于2023年的技术现实，到2026年已经出现了明显的滞后。

法规滞后不只是”慢半拍”的问题。它创造了一种特殊的套利空间：在法规尚未覆盖的灰色地带，先行者可以获得巨大的竞争优势，而落后者则面临”合规但低效”与”高效但违规”的两难选择。这种套利空间本身会扭曲技术的扩散方向——技术总是流向监管阻力最小的地方，未必是最需要它的地方。

**拖曳系数最高：教育。** 适应周期10–20年。教育体系是人类所有制度中惯性最大的。一个显而易见的原因是：教育的反馈回路极长。一项教育改革的效果需要等学生毕业、进入社会后才能评估——这个周期是十五到二十年。在这么长的反馈回路面前，任何改革冲动都会被稀释。

但更深层的原因是：教育体系的核心功能不是”传授知识”——如果是的话，互联网早就颠覆它了。教育体系的核心功能是**社会分层的认证机制**。清华北大的学位不是知识的证明（同样的知识可以在网上免费获取），而是一个信号：这个人通过了极度激烈的竞争筛选，他的认知能力大概率处于人群的前1%。雇主购买的不是知识，而是这个信号。

当AI让认知能力的获取成本趋近于零时，这个信号的价值理论上应该崩溃。但它不会——至少不会很快崩溃——因为社会还没有找到替代性的分层认证机制。在找到替代方案之前，旧的教育体系会继续运转，就像金本位废除后纸币继续流通，因为人们还没有习惯另一种货币。在2026年的当下，相当比例的终身教职教授可能从未在课堂上使用过AI辅助教学工具。这不是因为他们不知道AI，而是因为终身教职制度给了他们不改变的权利——不改变不会受到惩罚，改变也不会得到奖励。

**四种速度的叠加效应。**

个人（1–2年）< 企业（2–5年）< 法规（5–10年）< 教育（10–20年）——这四个时间常数的差异，解释了2026年我们看到的那种特殊的错位景象：

一个人已经在用AI处理自己80%的日常工作，但他所在的公司仍在讨论要不要”试点引入AI”。公司已经在内部大量使用AI了，但行业监管还在讨论”AI生成的法律意见书是否具有法律效力”。监管刚出台了第一批AI相关法规，但大学课堂上讲的还是十年前的教材，培养的是十年后已经不存在的岗位需要的技能。

这种错位不是某个环节的失败。它是四种制度按照各自的拖曳系数、以不同速度回应同一个技术冲击的必然结果。住在这个时间差里的人——那些按照旧制度的规则投入了时间和资源、但旧制度的回报正在消失的人——就是这场转型中最先感受到痛苦的群体。

制度惯性不全是坏事。它提供了缓冲——让社会不至于因技术的每一次跳跃而剧烈震荡。但缓冲也有代价：技术已经改变了现实，制度还在维护一个不存在的旧现实。当缓冲期耗尽、现实与制度之间的裂缝大到无法忽视时，调整将不是渐进式的，而是断裂式的。这正是”尾部风险”的典型机制——系统在表面稳定中积累了巨大的势能，然后在某个临界点释放。

---

## 五、这次为什么不同

每一次重大技术革命都打破了某种垄断。李录习惯追问事物的第一性原理，这个方法在这里特别有用——

印刷术打破了知识复制的垄断，蒸汽机打破了体力的垄断，互联网打破了信息传输的垄断——但在所有这些革命中，认知处理始终留在人脑里。被打破的垄断始终在人类认知能力的”外围”——存储、传输、体力——而认知本身不可触碰。

人类应对每一次技术革命的策略也因此是一致的：**用认知能力来适应变化**。体力被机器替代了？没关系，我们用大脑去管理机器。信息传输被互联网接管了？没关系，我们用判断力去筛选和使用信息。每一次，人类都用”更高级的认知”来应对”低级功能的自动化”。

这个策略隐含一个前提：认知本身是安全的、不可自动化的。

2020年代的AI革命是第一次对这个前提发起正面冲击。

这一次，被打破的垄断不是认知的外围，而是认知本身。不是信息的存储和传输，而是信息的处理和生成。

这就像一个人站在一块冰上，冰在融化。以前每次地震，他都能跳到另一块冰上——因为他的腿没有问题。但这一次，融化的不是脚下的冰，而是他的腿。

从文明演进的角度看，每一次真正的跃迁，都是因为某种基础性约束被突破。农业革命突破了食物供给的约束（不再依赖狩猎采集），工业革命突破了能量供给的约束（不再依赖肌肉和风水），信息革命突破了通信的约束（不再依赖物理载体）。每一次突破都不是在旧框架内做优化，而是创造了一个全新的可能性空间。

AI革命突破的约束是认知供给。三十万年来，复杂信息处理的唯一来源是人脑——每一颗人脑都需要二十年的教育投入来初始化，每天只能工作十几个小时，一生的产出是有限的。AI打破了这个约束，让认知供给可以像电力一样被规模化生产和分配。

这意味着什么？意味着所有基于”认知供给有限”这个约束条件而建立的制度——教育、就业、薪酬、社会分层、国家能力——都需要在新的约束条件下被重新设计。

这就是为什么”这次不同”不是一句空话。以前每次技术革命，人类都能用认知能力去理解变化、设计应对方案、重建制度。这一次，被变革的恰好是那个用来应对变革的能力本身。

---

## 六、地基上的第一块砖

本章论证了一个看似简单实则深远的命题：人类文明的全部制度建立在一个三十万年的隐含前提之上——只有人类大脑能够承担复杂信息处理。这个前提从未被质疑，因为从未有替代品。AI不是替代品

——它远比人脑粗糙、不可靠、缺乏理解——但它是第一个可用的替代选项。而一个选项的出现，就足以动摇所有基于”无选项”而建立的制度。

本章的核心发现：标准化认知的成本已经趋零，而这个变化正在向上侵蚀——复杂推理岗位在缩编，判断性认知的边界也在后退。稀缺性没有消失，但转移了方向。

第一章的论证收束于一个核心事实：这次被动摇的不是地面，而是脚。

如果认知垄断是文明大厦的地基，那么建在这块地基上的第一块砖是什么？

是劳动契约。

“你出卖认知劳动时间，我支付薪资报酬”——这个贯穿工业时代和信息时代的基础交易，在认知垄断动摇的那一刻就开始松动了。当一台机器能以你千分之一的成本完成你百分之八十的工作时，“你的时间值多少钱”这个问题的答案不再是确定的。

下一章，我们来拆解这块砖。

---

**本章关键判断与置信度**

| 判断 | 置信度 | 说明 |
| --- | --- | --- |
| 标准化认知任务的边际成本趋零 | A | 已在发生，趋势不可逆 |
| 复杂认知任务的人力需求将坍缩而非线性减少 | B+ | 逻辑自洽，但坍缩速度高度不确定 |
| 判断性认知安全边界在持续后退 | B | 方向确定，后退速度不可预测 |
| 标准化认知的自动化将切断高级岗位的人才管道 | B | 合理推演，缺乏充分实证 |
| 制度拖曳系数差异是转型混乱的主要来源 | A | 历史反复验证的结构性规律 |
| 这是第一次被打破的垄断是人类的应对能力本身 | A | 第一性原理推导，逻辑不可反驳 |

**未解之问**

一、如果标准化认知的自动化切断了高级岗位的训练管道，未来的”专家”如何产生？AI辅助训练能否替代实战经验？如果不能，我们面临的可能不是”岗位减少”，而是”专业能力断代”。

二、制度拖曳系数最高的教育体系，恰好是最需要率先变革的——因为它决定了下一代人用什么能力去面对新世界。但教育变革恰好是最慢的。这个”最需要先变”和”注定最后才变”之间的矛盾，会在十年后产生什么后果？

# 第二章 劳动契约——一个两百年的实验

> 我们以为自己怕没钱。其实我们怕的是没有一个每天早上必须起床的理由。

上一章的结论是认知垄断在松动。但抽象的松动不改变任何人的日子——得看它松动了什么具体的制度。建在认知垄断上的第一个制度，也是最大的一个，是劳动。

准确地说，是”每个成年人都必须通过出卖劳动来换取生存资格”这个安排。

## 一、一个并非天然如此的安排

从第一性原理追问开始：“每个人都必须工作”到底是不是自然法则?

答案是：不是。它甚至不是一个古老的传统。它是一个大约两百年前才被发明出来的制度安排，而且有一个明确的到期条件。

智人作为物种已存在约三十万年。在其中的29.97万年里，没有任何人”上班”。人类学家对现存狩猎采集部落的系统研究表明，他们每天花在食物获取上的时间约为四到五个小时，剩余时间用于社交、讲故事、照看孩子、制作工具、以及大量的”什么也不做”。人类学家萨林斯1972年管这叫”原初丰裕社会”——这个说法有理想化之嫌，但核心观察已被反复验证：狩猎采集者的物质需求有限，生态位中的食物来源分散但充足，他们完全没有”失业”的概念——因为不存在”就业”这个参照系。

农业革命让劳动时间变长了，但仍然没有产生”失业”。一个中世纪农民的劳动是季节性的：农忙期每天可能十二小时，农闲期几乎无事可做。他不”上班”——他耕种属于领主的土地，或耕种自己的小块地。他的劳动和他的生存直接相连，中间不经过”就业市场”这个中介。他不可能”失业”，正如一棵树不可能”失职”。

“失业”（unemployment）这个词的现代用法直到19世纪末才出现在英语中。在此之前，一个没在工厂干活的人不叫”失业者”——他可能是佃农、游方工匠、流浪者、乞丐、家庭主妇，有各种社会身份，但没有一种叫”失业”。“失业”的逻辑前提是：存在一个”就业”的常态，而你偏离了它。这个常态是工业革命制造出来的。

古代文明对劳动的态度，恰恰与现代相反。古希腊语中”劳动”（ponos）与”痛苦”同源。亚里士多德明确认为自由人不应从事体力劳动——劳动是奴隶的事务，自由人的正当活动是政治参与和哲学沉思。罗马帝国的公民尊严不来自他做什么工作，而来自他在公共事务中的参与程度。阿伦特在《人的境

况》中提供了一个更锐利的切口。她将人类活动分为三层：劳动（labor，维持生存）、工作（work，创造持久之物）、行动（action，在公共领域彰显人的独特性）。古希腊人把前两者留给奴隶和工匠，“行动”才是公民的正当事务。中国的”劳心者治人，劳力者治于人”表达了相似的等级观——士大夫的价值在于”治”而非”劳”。

真正的翻转发生在18世纪末到19世纪中叶。波兰尼在《大转型》中还原了这个过程的暴力性：圈地运动将农民从土地上剥离，使他们变成”自由”得一无所有的劳动力；1834年英国《新济贫法》将”不工作”等同于”不道德”，把贫穷从一种不幸重新定义为一种罪过；工厂纪律和钟表时间将人的身体节律改造为机器节律。这不是一个自然演化的过程——它是一系列制度设计和权力运作的结果。在不到三代人的时间里，“工作”从一种谋生活动被提升为一种社会身份、一种道德义务、一种存在证明。

到了2026年的中国，这个提升已经完成得如此彻底，以至于我们完全忘记了它是被人为建构的。

---

## 二、六根绳索，一个桩子

拆解现代劳动制度的结构，会发现它异常清晰：一个成年人的”工作”同时承载了至少六项完全不同的功能，而这六项功能全部绑在同一个载体上。

**收入来源。** 工资是大多数家庭唯一的现金流入口。在中国城镇家庭中，工资性收入占可支配收入的比例约为60%，加上经营性收入（本质上也是劳动所得），超过75%。离开了”工作”，绝大多数人立刻失去主要现金流。

**社会保障。** 中国的五险一金制度将养老、医疗、失业、工伤、生育保险和住房公积金全部绑定在雇佣关系上。没有工作意味着没有雇主缴纳的那一半社保——而这一半往往占总缴费额的六成以上。个人参保的替代方案存在，但覆盖率低，且不含公积金。这意味着”失业”在中国不仅是失去收入，还是失去医疗和养老的安全网。

**社会身份。** “你做什么工作的？”是中国社交场合最标准的开场白。它不只是闲聊——它是一次快速的社会定位：你的收入水平、教育背景、社会阶层、甚至婚恋市场上的竞争力，全部浓缩在你的职业标签里。“银行客户经理”和”外卖骑手”不仅描述了两种不同的工作，更描述了两种截然不同的社会存在。如果你不信，试试在家庭聚会上回答”我目前没工作”——你会立刻感受到整桌人的沉默。

**日常社交。** 对多数上班族而言，同事是除家人外接触最频繁的人群。工作场所是社交网络的核心节点。失去工作往往同时意味着失去一个完整的社交生态——你与前同事不再共享日常话题、共同抱怨和集体仪式，联系自然就断了。

**时间结构。** 工作提供了日、周、年的节奏框架。九点上班、六点下班、周末休息、年假旅行——这套节奏不是自然的，是工业社会强加的，但人类已经深度适应了它。长期没有工作的人经常报告一种”时间失去了形状”的感觉——时间锚点消失了。人类学研究揭示了更深一层的机制：劳动不仅填充时间，它提供的是日常生活的仪式节奏。通勤是从私人领域进入公共领域的过渡仪式，午餐是集体停顿

的共享时刻，下班是身份切换的信号，周末是恢复与奖赏的循环节点。这些看似平凡的重复并非可有可无的惯性——涂尔干和戈夫曼的研究都指向同一个结论：日常仪式是维持心理稳定和社会认同的基础结构。当劳动消失时，瓦解的不只是收入来源，还有整套赋予日子以形状的仪式秩序。

**心理健康。** 大量研究证实，非自愿失业与抑郁、焦虑、自尊下降、甚至自杀率上升高度相关——而且这种相关性在控制了收入因素之后依然显著。也就是说，即使给失业者发同样多的钱，他们的心理状况仍然比有工作的人差。这说明工作提供的不仅是经济回报，还有某种更深层的东西——可能是目的感，可能是被需要的感觉，可能仅仅是一种”我在世界中有一个位置”的确认。

六根绳索，全部绑在”你有一份工作”这一个桩子上。

这个安排在人类历史上是极其特殊的。一万两千年的农业社会里，收入来自土地，身份来自血统和宗教，社交在宗族和教区中完成，时间结构由季节和宗教历法决定，心理锚点由信仰和仪式提供。六项功能分散在多个载体上，任何一个载体的松动都不会导致全部坍塌。

工业社会的做法是将六根绳索拧成一股——效率极高，但脆弱性也极高。这是一个典型的塔勒布式”脆弱系统”：在正常运行时表现优异，一旦核心载体受到冲击，所有功能同时失灵。塔勒布在《反脆弱》中反复强调的原则是：真正稳健的系统让鸡蛋分散在多个独立的篮子中，而非放在一个最坚固的篮子里。工业社会的劳动安排恰恰违反了这条原则。

涂尔干在一百多年前就预见了这种脆弱性的根源。他区分了两种社会团结：传统社会依靠”机械团结”，人们因为彼此相似而凝聚，共享信仰、习俗和生活方式；工业社会则转向”有机团结”，人们因为分工的相互依赖而凝聚，每个人的专业化角色使他既不可或缺又无法自足。有机团结的效率远高于机械团结，但它有一个致命的依赖项：分工体系必须持续运转。当AI开始大规模替代知识分工时，动摇的不仅是个体的就业机会，而是整个社会赖以凝聚的底层机制。这不是一个经济问题，而是一个社会解体问题——六根绳索断裂的背后，是有机团结本身的基座在松动。

---

## 三、这台机器的隐含前提

每台机器都有一个设计前提。劳动契约这台机器的设计前提是什么？

**社会需要大多数人的劳动。**

过去两百年，这个前提几乎从未被真正挑战过。工业化需要工人，城市化需要服务人员，信息化需要知识工作者——每一次技术升级都消灭了旧岗位、创造了新岗位，而新岗位的总量始终足以吸纳大部分劳动力。经济学家把这个规律叫做”技术互补性”：机器替代了人的一部分能力，同时增强了人的另一部分能力，净效应为正。

但这个规律有一个被忽略的隐含条件：**技术替代的是特定技能，而非通用认知能力。**

蒸汽机替代了肌肉力。纺织机替代了手工纺纱的技艺。计算机替代了手工计算的能力。每一次替代都是针对某一项具体技能，而人类的通用认知——学习新事物、适应新环境、在不确定性中做判断——始终完好无损。一个被纺织机淘汰的手工织工可以去学操作纺织机；一个被电子表格淘汰的记账员可以去学数据分析。“学习新技能”本身是一种通用能力，此前没有任何机器具备。

关键追问是：**“新技术总是创造更多工作”——为什么这次可能不成立？**

答案不在于AI有多强大，而在于它替代的层面不同。AI替代的不是”做某件事”的能力，而是”信息处理”本身——而信息处理恰恰是几乎所有现代知识工作的通用底层。一个律师和一个会计和一个翻译和一个分析师的工作表面上完全不同，拆到底层，他们都在做同一件事：接收信息，按规则处理，输出结果。当这个底层被机器以极低成本覆盖时，“学习新技能”不再是可靠的逃逸通道——因为新技能的底层同样是信息处理。

这不是说没有任何新工作会出现。新工作当然会出现。但真正的问题更具体：新工作的数量能否吸纳被替代的劳动力？新工作的薪资水平能否维持现有的生活水准？新工作的技能门槛对被替代者是否可及？

后来我们知道了这三个问题的答案。新工作确实出现了——AI训练师、提示词工程师、人机协作设计师、数字体验策展人。但它们的总量远远不及被压缩的旧工作。更关键的是，新工作的技能分布呈现出一个尖锐的U形：要么极高（AI系统架构师、对齐研究员），要么极低（数据标注员、内容审核员），中间地带——过去吸纳了最大量白领劳动力的那个区间——被AI填满了。

这个过程的因果链清晰可循：AI让信息处理的边际成本急剧下降，社会因此不再需要大多数人执行信息处理劳动，劳动契约的基础开始动摇。捆绑在劳动上的六项功能同时面临脱钩压力，但制度惯性让这个脱钩过程极其缓慢和痛苦。

---

## 四、阈值效应——为什么30%就够了

理解劳动替代的关键在于非线性。人类的直觉是线性的：如果一个行业被AI替代了10%的岗位，我们会本能地认为影响大约也是10%。但劳动市场的运作机制是高度非线性的。

让我们用一个具体的行业来推演。假设一个城市的法律服务市场原本有一万名执业律师。AI进入后，标准化法律工作（合同审查、文书起草、法规检索、尽调报告）被压缩，这部分工作大约占律师总工作量的60%到70%。这不意味着60%的律师会失业——因为每个律师的工作都是标准化和非标准化的混合体。AI做的是把每个律师的有效利用率从接近100%压缩到40%左右。

现在，同样质量的法律服务，市场只需要四千名律师就能完成。但市场上有一万名律师。

第一个效应是价格坍缩。一万名律师争夺四千人的工作量，律师费率开始承压。率先降价的是中小所和年轻律师——他们在价格竞争中没有品牌护城河。大所暂时可以靠品牌溢价维持报价，但随着中小

所以更低价格提供AI加持的”足够好”的服务，大所的价格体系也开始被侵蚀。

第二个效应是身份坍缩。当一个行业有40%的从业者处于半闲置状态时，这个行业的社会声望开始下降。法学院报名人数走低，在读学生开始怀疑职业前景，头部人才流向其他行业。这进一步削弱了法律行业的人才基础，形成螺旋。

第三个效应是生态坍缩。法律行业不只是律师——它还包括律所的行政人员、法律出版商、法律培训机构、法律科技公司（为人类律师提供工具的那些）、法学院的教职人员。当律师数量需求下降40%时，整个生态的就业岗位萎缩幅度远超40%。

这就是阈值效应：系统并不是随着压力的线性增加而线性劣化。它在承受一定压力时看起来运转正常，然后在某个临界点突然进入非线性崩坏。30%的岗位被压缩，带来的不是30%的影响——它带来的是薪资谈判权的坍缩、职业认同的瓦解、整个行业生态链的收缩。

回看历史，这种阈值效应并非没有先例。20世纪80年代，美国钢铁业的产量下降了约40%，但就业人数减少了约一半——因为存活下来的钢铁厂使用了更高效的技术，用更少的人做同样多的事。底特律汽车产业的萎缩同样如此：产量下降30%，但工人数量减少远超30%，因为日本竞争压力迫使剩余工厂极大提升自动化水平。每一次，压力的非线性放大效应都远超决策者的预期。

AI对知识工作的冲击将复刻这个模式，但作用范围更广。钢铁业的坍缩只影响了特定地区的蓝领工人。法律、会计、金融分析、翻译、咨询、初级编程——这些行业的知识工作者分布在每一个城市、每一个阶层。当阈值效应同时在十几个行业发生时，它不再是一个行业问题，而是一个社会结构问题。

---

## 五、三种文化，三种脆弱性

劳动和身份的捆绑在不同社会中的紧度不同。这意味着同样的技术冲击，会在不同文化中激发不同强度的危机。

**日本是捆绑最紧的。** 战后日本经济奇迹建立在”终身雇佣制”之上：大学毕业进入一家公司，一直干到退休，公司负责你的几乎全部人生——收入、住房补贴、社交网络甚至婚姻介绍。名片上的公司名在个人名之前——“三菱商事·山田太郎”，而不是”山田太郎·三菱商事”。个人身份嵌入公司身份之中，二者不可分割。

“社畜”——社会的畜生——这个自嘲词汇本身就说明了捆绑的深度。当公司不再需要山田太郎时，山田太郎不仅失去了工作，他在某种意义上失去了自己。日本中老年男性的自杀率与失业率的正相关系数，在所有发达国家中最高。2026年后，日本企业采取的策略不是大规模裁员（那在文化上几乎不可能），而是不再招聘新人。老员工自然退休后，岗位由AI接管。这是一种”温水煮青蛙”式的替代——没有冲突，没有抗议，只有新一代年轻人发现自己面前的门在安静地关上。

**中国的捆绑方式不同，但同样紧。** 中国没有终身雇佣制的传统，但有另一套等量的绑定逻辑：内卷。内卷的本质不是竞争——任何社会都有竞争——而是竞争的单一通道性。中国主流叙事中的人生价值通过一条极窄的通道实现：好成绩带来好大学，好大学带来好工作，好工作带来好收入，好收入带来好婚姻，好婚姻带来好下一代。这条通道的每一个环节都与”劳动”直接或间接相关。一个没有”好工作”的成年人，面临的不仅是经济压力，还有来自整个社会网络——父母、亲戚、同学、邻居——的持续评价压力。

“你在哪里上班？”这个问题在中国社交场合出现的频率，大概是全世界最高的。当一个人无法给出体面的回答时，他失去的不只是面子，而是在整个社会评价体系中的坐标。

中国还有一个独特的缓冲层：体制内。约五千万体制内就业人员（公务员七百万、事业编三千多万、国企正式工一千余万）构成了一个巨大的减震器。体制内的编制提供的不仅是经济安全，更是一种存在论层面的确定性——“只要体制在，我就在。”公务员考试报名人数屡创新高，不是因为体制内待遇变好了，而是因为体制外变得不确定了。但这个缓冲层本身也有极限——它的资金来源是税收，而税收依赖于市场经济的活力。当AI压缩了市场化就业的收入和规模，税基收窄，缓冲层的可持续性也会受到挤压。

**美国的捆绑看起来最松，但有一个隐性的紧点：信用消费。** 美国的文化基因中有”自我雇佣”的传统——从拓荒者到创业者，美国叙事的核心是”创造你自己的机会”。一个被裁员的美国人去开Uber、做自媒体、搞个小生意，在社会观感上并不丢人。但这种文化弹性掩盖了一个结构性脆弱点：美国中产阶级的生活方式高度依赖信用消费。平均家庭信用卡债务超过六千美元，学生贷款总额超过1.7万亿美元，房贷是家庭最大负债项。当AI开始压缩白领薪资时，债务结构会放大冲击——因为债务是刚性的，它不会因为你的收入下降而自动减少。

三个国家，脆弱点各不相同。日本是身份——工作和自我焊在一起，拆不开。中国是通道太窄——只有一条路算”正路”，堵死了就是死路。美国看着最灵活，但账单不灵活，杠杆放大了每一次收入波动。但无论哪种形态，底层的因果链是相同的：当”社会需要你的劳动”这个前提松动时，建立在劳动之上的全套社会安排都会跟着松动。

---

## 六、两百年的到期条件

过去两百年的劳动安排可以被理解为一份有隐含到期条件的社会契约：社会需要你的劳动——作为交换——社会给你收入、身份、保障、社交、时间结构和心理锚点。这份契约之所以运转良好，是因为到期条件从未被触发。每一次技术革命都消灭旧工作，但也创造新工作，净效应为正，社会对劳动的需求始终旺盛。

波兰尼在更长的历史尺度上发现了一个规律，他称之为”双重运动”：市场力量每一次扩张，把新的领域纳入市场交换，都会激发社会的自我保护运动。劳动的市场化激发了工会运动和劳动立法。金融

自由化激发了金融监管。全球化激发了贸易保护主义。这不是非理性的反动，而是社会有机体的免疫反应——当市场逻辑侵入到威胁社会存续的程度时，社会会自发产生对抗力量。

AI对劳动的替代，是最新的一轮市场扩张。按照波兰尼的逻辑，社会保护运动必然会跟上来。2023年好莱坞编剧和演员的联合罢工——直接针对AI在影视行业的应用——只是预演。但更大规模的抵抗还没有到来。

为什么？因为AI的替代方式不同于以往。蒸汽机替代手工织工时，被替代的人清晰地知道自己被替代了——织机就在那里，替代是可见的、具体的、可以被指认的。人们可以愤怒，可以砸机器（卢德运动），可以组织罢工。但AI替代知识工作者的方式更隐蔽：你的职称还在，你的工位还在，你每天还在”上班”——但你的工作实质已经被掏空了。你不是被机器替代了，你是被重新定义为”机器的审核员”。你的处境用一个词来形容：不完全就业。

间奏一中那位出版社编辑的处境，就是”不完全就业”的预演。主编轻描淡写地说”翻译类的书可以用AI初译”，她的三分之一工作量瞬间悬在了空中，但她的工牌、工位、劳动合同全都完好无损。

“不完全就业”不等于失业。统计上你是”在职”的，有劳动合同、有社保、有工资条。但你自己心里清楚，你每天真正创造价值的时间可能只有两三个小时，剩下的时间在”假装忙碌”或做你清楚AI做得更好的事。这种状态比失业更难组织抵抗——因为你无法明确指出”谁替代了你”。你也无法向社会承认”我的工作已经没有实质内容了”——因为那等于承认你的社会位置是虚假的。

这就产生了一个悖论：系统的脆弱性在增加，但没有任何可见的指标在报警。失业率只是小幅上升。GDP还在增长。股票市场一片繁荣（因为AI提升了企业利润率）。所有的宏观指标都指向”一切正常”——但在微观层面，数以千万计的劳动者正在经历一种缓慢的、无声的功能剥离。

---

## 七、经验的贬值

一个常见的成长公式是”痛苦+反思=进步”。但这个公式有一个前提：痛苦必须来自可识别的错误，而反思必须能指向可操作的改进。当痛苦来源模糊、改进路径不清时，痛苦就只是痛苦——没有进步的出口。

AI时代劳动者的核心困境正在于此。过去，经验是资产。一个银行客户经理干了十七年，他对客户心理的把握、对风险偏好的判断、对产品适配的直觉——这些都是时间积淀的产物，无法速成。但AI改变了经验的价值等式。AI不需要十七年积累客户关系——它可以在客户每一次打开手机银行的瞬间，根据交易记录、资产配置、浏览行为，给出比任何人类客户经理都更精准的产品推荐。AI不是没有”经验”——它的经验来自数百万客户的数据训练，规模和密度远超任何个体。

更深层的问题是：经验不仅在贬值，经验还可能成为负债。一个在旧范式下积累了十七年经验的人，他的认知模型是按照旧规则校准的。他对世界运作方式的理解、对风险的判断框架、对”什么是好工作”的认知——全部锚定在一个正在消逝的现实中。这些锚定让他在面对新现实时更难调整，而不是

更容易。这就是”范式锁定”——当你的成功经验太强、太久、太深入地嵌入你的认知结构时，它会变成一副隐形的枷锁。

回到蒂尔那个0到1和1到n的区分——创造全新事物，与复制已有事物的规模化。工业社会的大多数劳动者从事的是1到n的工作——按照已知的方法、已知的流程、已知的标准，重复执行。经验在1到n的工作中价值极高，因为重复次数越多，执行质量越高。但AI恰恰最擅长1到n——规模化的重复执行是机器的原生优势。留给人类的是0到1的空间——创造性的、非标准化的、没有先例可循的工作。而0到1恰恰是经验帮不上忙的地方——因为创新的本质就是做没人做过的事，经验在定义上是对已做过的事的总结。

一个人类写出一份合格的商业计划书需要两周。AI需要两分钟。中间差出的那两周，曾经是一个阶层的全部经济基础。

---

## 八、意义真空

历史上，制度断裂常常以明确的方式发生——一纸诏令、一场革命、一次立法。受冲击的人至少知道旧规则已经废止，可以着手寻找新路径。英国1834年《新济贫法》将济贫制度从教区慈善改为工厂纪律，数十万人的生存方式被明确重新定义；日本明治维新废藩置县，武士阶层一夜之间失去了俸禄和身份，但同时获得了明确的信号：旧路已断，必须转型。

AI对劳动的冲击不同于这些历史先例。它是渐进的、弥散的——没有一纸诏令宣布”工作不再有意义”。信号在无数个微小的市场变化中慢慢褪色：一个部门从十二人缩编为五人，一个岗位从”做判断的人”变成”审核AI判断的人”，一份薪资在通胀调整后连续三年负增长。每一个变化都不构成新闻，但加在一起构成了趋势。

渐进性制造了一种新型的意义真空。你有职称、有工位、有工资条——但职称背后的工作内容已经所剩无几。旧路径似乎还在运作（你还在上班），但回报率在下降（同样的工作换来的收入更低了），而新路径还不清晰。这种模糊的、无法定位的不安，比明确的制度断裂更难应对——因为你甚至不确定自己是否需要转型。

---

## 九、波兰尼的双重运动2.0

那么核心问题是：这一次的社会保护运动会以什么形式展开？

波兰尼的框架预测，每一次市场力量对社会肌体的侵蚀都会激发反向运动。AI对劳动的替代不会例外。但反向运动的形式取决于替代的特征。

第一波反向运动已经出现，它的特征是”守住既有”。好莱坞编剧罢工是经典案例：一个有明确组织边界、强工会传统、高公众可见度的行业，率先发起了针对AI的集体行动。但这种形式的抵抗有一个结构性限制——它只在”劳动者有强组织能力”的行业有效，而AI最先冲击的恰恰是那些组织化程度低的行业：自由职业者、合同工、平台经济从业者。他们没有工会，没有集体谈判权，甚至没有稳定的雇佣关系。

第二波反向运动指向制度创新。这是”重组阶段”——旧安排不再可行，新安排必须被创造出来。UBI（全民基本收入）的讨论在2025年后显著升温。AI税、机器人税、数据红利等概念进入主流政策讨论。但从概念到政策到执行，中间有巨大的制度拖曳——每一个新分配方案都牵涉到复杂的政治博弈、财政测算和社会接受度问题。

第三波反向运动是文化层面的——它是最慢的，但也是最深层的。当”工作”不再是社会通行证时，“我是谁”这个问题必须找到新的答案。这不是一个经济问题，而是一个存在论问题。它需要的不是更好的就业政策，而是一整套新的意义体系。历史上，这种规模的意义重建从来没有在一代人之内完成过——从农业社会到工业社会的价值转型花了至少三到四代人。但这一次的时间窗口可能要窄得多。

---

## 十、到期之后

最后一个问题反着问：劳动契约到期了，拿什么接？

这个问题的答案不能只是”新的收入分配方式”——虽然那确实是最紧迫的。真正需要接替的是劳动所承载的全部六项功能。收入可以通过UBI或其他分配机制部分解决。社会保障可以从雇佣关系中解绑——中国已经在推进灵活就业人员的社保覆盖，但进度远远跟不上需求。

最难替代的是后四项：身份、社交、时间结构、心理健康。这四项的共同特征是：它们不是经济问题，不能通过经济手段解决。给一个失去工作的人发钱不能给他身份认同，不能给他社交网络，不能给他日常节奏，不能给他被需要的感觉。这些需求必须通过新的社会组织形式来满足——社区、协会、志愿者网络、公民参与渠道。但这些组织形式在当代中国严重发育不足——恰恰因为工业社会的劳动制度太”成功”了，它把所有功能都整合在了”公司”这个单一载体中，挤压了其他社会组织形式的生存空间。

这是一个深层的结构性困境：劳动契约正在到期，但替代它的社会基础设施还没有建好。制度的解体速度和制度的重建速度之间存在一个危险的时间差。住在这个时间差里的人——按照旧制度的规则投入了时间和资源，但旧制度的回报正在消失——就是这场转型中最先感受到痛苦的群体。

他们不会坐以待毙。后来的事实证明，当劳动契约松动时，人们会自发地寻找替代物。他们寻找的方向——社区、关系、信任、面对面的在场——恰恰是AI做不了的那些事。但这些自发的寻找是零散的、脆弱的、缺乏制度保障的。

一个社会能不能建立起一套新的制度安排，让个人的挣扎变成有制度保障的选择？

这是劳动契约到期之后，摆在文明面前的核心问题。但要回答它，我们首先需要理解另一个制度——劳动契约需要人力供给，而负责筛选和分配人力的那台机器，叫做教育。

---

**本章关键判断与置信度**

| 判断 | 置信度 | 说明 |
| --- | --- | --- |
| 劳动契约是工业社会的制度发明，而非自然法则 | A | 历史事实，无争议 |
| AI替代的是通用认知能力，而非特定技能 | A | 与第一章分析一致，充分证据 |
| 劳动替代有阈值效应——30%的压缩可触发非线性坍缩 | B | 历史先例支持，但具体阈值因行业而异 |
| 不同文化对劳动–身份脱钩的脆弱性形态不同 | A | 大量跨文化研究支持 |
| “不完全就业”将成为2028–2035年新常态 | B | 多条趋势指向，但程度有不确定性 |
| 波兰尼”双重运动”将在AI时代重现 | B | 逻辑清晰且已有早期信号，但形式不确定 |

**未解之问**

一、如果劳动不再是社会通行证，什么可以替代它？历史上有两种候选方案：公民身份（古罗马：你是公民，所以你有尊严）和信仰归属（中世纪：你是教徒，所以你有意义）。AI时代的”通行证”会是什么？

二、“不完全就业”是一种过渡状态，还是一种新的稳定结构？如果企业发现保留大量”不完全就业”的员工比裁员更便宜（避免了社会压力和品牌风险），这种状态可能持续很多年。那么，一个由数千万”名义就业者”构成的社会，会发展出什么样的文化心理？

# 第三章 教育的承诺——一台精密的分拣机器

> 一个系统越是宣称自己在培养人，越值得追问：它实际上在筛选什么。

劳动契约需要人力供给。每年有数以百万计的年轻人被输送到劳动市场上，填充从初级岗位到高级岗位的各个节点。谁来筛选这些人？谁来决定谁去哪个节点？谁来为每个节点标上价格？

这台筛选机器有一个被广泛美化的名字：教育。

## 一、分拣的本质

如果不看教育体系的宣传口号——“培养人才”、“启迪心智”、“为民族复兴育人”——而是观察它的实际行为，你会看到什么？

入学考试、期中期末测验、GPA排名、升学率统计、就业率公布。从小学一年级的期末考试到博士论文的答辩评级，教育体系的每一个环节都在执行同一个操作：测量、排序、贴标签。一年级的期末考试和博士答辩在技术规格上完全不同，但在功能层面完全一致——它们都是分拣。

这个功能不是教育体系的”缺陷”或”副作用”。它就是教育体系被设计出来要做的主要事情。

怎么判断一个系统的主功能？看激励结构。学校的预算拨付、教师的绩效考核、校长的升迁指标，主要取决于什么？升学率和考试成绩。不是”学生的批判性思维是否提升了”，不是”学生的公民素养是否培养了”，不是”学生是否在审美和人际方面有了成长”。这些维度几乎从不出现在任何一所中国学校的核心KPI中。它们是教育叙事中的装饰，不是教育机器的齿轮。

激励结构揭示真实功能。教育的真实功能是分拣。

这个判断听起来可能过于冷酷。但请注意，说教育的主功能是分拣，并不是说教育”只能”分拣、“不该”做别的。它是在描述一个事实——一个制度在实际运行中做的是什么，和它在口号中宣称做的是什么，往往是两件事。

## 二、一千三百年的信号机制

这台分拣机器的历史比大多数人以为的要长得多。

公元605年（一说607年），隋炀帝设立进士科，科举制度由此诞生。在随后的一千三百年里，这套制度以惊人的稳定性运转着：用标准化考试从数以万计的读书人中筛选出几百人，赋予他们管理帝国的权力。状元做翰林，进士做知县，落榜者回家务农。你在这台机器中的位置，决定了你一生的社会地位、经济收入和政治权力。

科举之所以能延续一千三百年，是因为它解决了一个所有大规模社会都面临的根本问题：信息不对称。帝国需要官员来管理辽阔的领土，但最高权力无法亲自考察每个候选人的实际能力。科举提供了一个低成本的筛选信号——能通过考试的人，至少证明了他具备基本的读写能力、记忆力和思维组织能力。信号是否精确并不是最重要的——重要的是它足够廉价和标准化，可以在整个帝国范围内统一执行。

用经济学的语言来说，科举的本质是一个信号机制（signaling mechanism）。

1973年，经济学家斯宾塞写了一篇后来让他拿到诺贝尔奖的论文，核心观点反直觉得近乎冒犯：教育的主要经济功能不是提升人的能力（虽然它确实也有这个副效果），而是向雇主发送一个关于求职者能力的信号。大学学位之所以值钱，不是因为大学教给了你多少独特的知识——在互联网时代，几乎所有大学课程的内容都可以在线免费获取——而是因为”能从这所大学毕业”这件事本身，证明了你具备一定的智力水平、自律能力和社会适应性。

换句话说，学位是一张昂贵的信号牌。雇主买的不是你学到的知识，而是大学替他们完成的筛选工作。

蒂尔更尖刻。他的逻辑是：大学的真正产品是信号。学历之所以有效，恰恰因为它既昂贵又耗时。如果获取学位是免费的、即时的，它就失去了信号功能——因为信号的区分力来自于获取它的成本。钻石之所以是地位象征，因为它贵，碳结晶本身没什么特殊的使用价值。学位的逻辑与此完全一致。

这解释了一个看似荒谬的现象：名校毕业生的起薪远高于普通大学毕业生，即使他们学的是同样的专业、读的是同样的教材。因为雇主支付的溢价不是给”知识”的，而是给”信号强度”的。清华的信号比普通一本强，一本比二本强，二本比大专强。这套信号体系的精密程度，和一千三百年前的科举如出一辙——只不过状元变成了清北毕业生，进士变成了985，举人变成了一本，秀才变成了二本。

从科举到高考，考试的内容在不断变化——从四书五经到数理化，从八股到标准化选择题——但核心功能从未改变：用考试把人排序，然后让市场按照排序定价。

---

## 三、信号的失灵

AI正在侵蚀学历信号的区分度。这是本章的核心命题，需要精确的推演。

信号的价值取决于它的区分能力。如果持有信号A的人群和持有信号B的人群之间存在稳定的、可预测的能力差距，那么信号就是可靠的——雇主可以放心地用它做决策。但如果这个差距开始收窄，信号的信噪比就下降了。当信噪比降到一定程度以下，信号就从”有效的筛选工具”变成了”昂贵的噪音”。

AI通过三个机制同时压缩信号的区分度：

**第一个机制：知识获取的民主化。** 名校的核心优势之一是知识资源的集中——更好的教授、更前沿的课程、更丰富的图书馆、更密集的学术讨论。AI把这个优势几乎抹平了。一个二本院校的学生，配合AI获得的学习资源质量，与清华学生的差距在急剧缩小。AI不仅能回答问题，还能根据学生的具体薄弱点定制学习路径——这在某些方面甚至超过了名校教授的一对多授课。

**第二个机制：技能产出的拉平。** 假设一个名校毕业生写一份法律备忘录的质量得分是85/100，一个普通院校毕业生的得分是65/100。这20分的差距足以让雇主为名校标签支付30%的薪资溢价。现在，两个人都使用AI辅助。名校毕业生的产出质量从85提升到92——AI帮他优化了措辞、补全了引用、检查了逻辑。普通院校毕业生的产出质量从65提升到88——AI在他原本薄弱的地方提供了更大的边际增量。差距从20分缩小到4分。4分的差距不足以支撑30%的薪资溢价。

这个效应在统计学上有一个名字：天花板效应。当基线水平较高时，AI的边际提升有限；当基线水平较低时，AI的边际提升更大。结果是：AI把能力分布从宽峰压缩为窄峰，信号的区分度急剧下降。

**第三个机制：替代性信号的涌现。** 当传统信号（学历）失灵时，市场会寻找替代信号。GitHub上的开源项目质量、个人作品集的展示、AI协作能力的实际测试、短期项目的交付表现——这些新信号不依赖学历发放机构，而是直接展示实际能力。它们的获取成本远低于四年大学教育，但信息含量可能更高。

因果链清晰可循：AI降低了知识获取的边际成本，不同学历背景者在知识和基础技能上的差距随之缩小，学历作为能力信号的区分度下降。雇主开始寻找替代筛选维度——项目经验、技能测试、AI协作能力的权重上升，学历信号的相对价值持续下降，形成自我强化的循环。

但教育体系不会主动承认这一点。

---

## 四、利益惯性——为什么真相不会被说出

塔勒布有一个判断标准可以直接套用在这里：skin in the game，切身利害。当一个系统中的决策者不承担自己决策的后果时，系统就会产生扭曲。教育体系恰恰是这个原则的反面教材。

全球高等教育是一个年产值超过两万亿美元的产业。从大学管理层到教职员工，从教育贷款机构到大学城的房地产开发商，从教辅出版商到标准化考试服务公司，数以千万计的人的生计依赖于一个前提：学历信号是有效的，因此教育投资是值得的。

这些利益相关者中，没有任何一方有动力告诉消费者”信号正在失灵”。

大学不会说。学费收入是大学运转的基础。如果承认”我们颁发的学位的市场价值正在下降”，招生会立即受到冲击。所以大学做的事情是：在课程中加入”AI素养”模块，宣布”我们的教育不止是知识，还有思维方式和人脉网络”——这些说法都不算错，但它们巧妙地回避了核心问题：你的学位作为就业信号的效力正在衰减。

教育培训机构不会说。中国的K-12课外辅导产业虽然在”双减”后转入半地下状态，但家庭端的教育支出并未实质减少。海淀区一对一私教价格持续攀升，部分科目已接近每小时数百元甚至更高。培训机构的商业模式建立在”考试排名决定人生”这个叙事之上，它们不可能主动瓦解自己的商业基础。

雇主也不急着说。大企业的HR体系已经建立了一整套基于学历的筛选流程——校招渠道、目标院校名单、GPA门槛。重建一套新的筛选体系需要时间、精力和试错成本。在新体系建立之前，沿用旧体系是阻力最小的选择。

政府更不会说。高考是中国社会最接近”公平”的制度之一。任何对高考体系的根本性质疑都可能触发关于公平、阶层、区域差异的大规模社会争论。维护高考的合法性首先是社会稳定问题，其次才是教育政策问题。

所有利益相关方都有充分的理由维护现状。这不是阴谋——没有人坐在密室里策划如何欺骗消费者。这是制度惯性的自然结果：每个参与者都按照自己的理性行事，但集体理性的结果是系统性的信息扭曲。信号在贬值，但没有人在喊”皇帝没穿衣服”。

这是”代理人问题”的教科书案例：做决策的人（教育机构、培训公司、政策制定者）不承担决策错误的主要后果——承担后果的是那些按照旧信号做出人生投资的家庭。

---

## 五、中间层的陷阱

逆向思考——不问”谁会受益”，而问”谁最脆弱”。在教育信号失灵的过程中，最脆弱的群体不是最底层，也不是最顶层。

最底层从未真正依赖教育叙事。一个农村的初中辍学者，他的谋生方式——务农、务工、个体经营——从来不以学历为前提。教育信号的失灵对他的影响接近于零。

最顶层有足够的资源绕过信号失灵。一个有钱人家的孩子，即使学历信号贬值了，仍然可以通过家族关系网络、海外教育、个人品牌建设等替代路径维持优势。对他们来说，学历本来就只是多个工具之一，而非唯一依赖。

最受冲击的是中间层——那些家庭年收入在20万到80万之间、没有显赫的社会关系网络、把几乎全部向上流动的希望押注在教育这一条路径上的家庭。他们是”教育改变命运”叙事的最忠实信徒。

间奏一中那个海淀的高二女孩，每天花十四个小时学习，她妈妈每次月考后都会拿隔壁张阿姨女儿的分数来比较。她妈妈正是这个群体的典型代表。在中国的语境下，这个群体有一个更具体的画像：1985年到1995年之间出生、赶上了中国经济高速增长期、通过高考进入城市、现在是互联网公司中层或金融行业白领或体制内中级职称的那一代人。他们的人生经验完美地验证了”好大学带来好工作、好工作带来好生活”的因果链。所以当他们做了父母，第一反应就是让孩子复制同样的路径，只是投入更多资源。

但因果链的成立依赖两个前提，AI同时动摇了它们：第一，知识稀缺——掌握特定知识的人少，所以拥有知识的人有溢价；第二，信号有效——学历能可靠区分”有能力”和”没能力”的人。当这两个前提同时松动，因果链就断裂了——但链条上的投资已经发生了。

从第一性原理分析：这些家庭的教育支出本质上是在购买一种”未来收入的期权”——用当下的投入换取未来较高收入的可能性。期权的价值取决于底层资产的行为。当AI改变了劳动市场的收入分布曲线——压缩了中间段、拉高了顶端、扩大了底端——这份期权的底层资产发生了根本性变化。但期权的价格（教育支出）还在按照旧的底层资产定价。

这就是泡沫。教育的泡沫不同于房地产泡沫或股票泡沫——它不会在某一天突然崩盘。它是一种缓慢的、价值回归的过程：人们逐渐发现，同样的教育投入换来的就业回报在降低，但这个发现需要一个完整的就业周期（至少三到五年）才能被统计数据证实。在这个时间差内，大量家庭仍然在按照旧定价继续投入。

一个典型的认知锁定在起作用：当你的整个人生成功建立在某条路径上时，你会倾向于认为这条路径具有普遍性和持久性——即使环境已经变了。心理学家把这叫做”锚定效应”——你的判断被锚定在过去的经验上，而不是当前的现实中。改变认知的成本太高了：如果一个母亲承认”教育可能不再是最好的投资”，她就必须同时承认过去几年花的五十万可能打了水漂，承认她规划孩子人生的整套方法论可能是错的，承认她从小到大相信的”知识改变命运”的故事可能只在特定历史窗口内成立。

没有人愿意在一个普通的周末下午面对这种认知崩塌。所以她做了大多数人会做的事：给孩子报下一期辅导班。这不是愚蠢——在信息不完全的情况下坚持已验证的策略是理性的。但当环境变化的速度超过认知更新的速度，理性就变成了惯性，惯性就变成了陷阱。

---

## 六、1905年的幽灵

1905年废科举的历史为当下提供了一个精确的结构类比。

废科举的核心冲击不在于知识过时——四书五经仍有文化价值——而在于”知识通向权力”的制度通道被关闭。一个苦读二十年的举人失去的不是能力，而是能力的兑换通道。更深层的是：科举同时是选拔制度和社会整合机制——它统一了价值标准、维系了社会流动、整合了地方精英。当这套机制突然

消失，整个社会失去了”什么值得做”的共同参照系，“意义真空期”持续了至少一代人，新的选拔机制花了几十年才逐步建立。

当下的教育信号贬值与废科举有两个结构相似点和一个关键不同点。

**相似点一：不是知识过时了，是通道变了。** AI时代的专业知识没有过时——法律知识、医学知识、工程知识仍然有价值。过时的是”拥有这些知识就能获得高薪工作”的保证。知识本身和知识的市场价格是两件事。

**相似点二：最受冲击的是最相信旧叙事的那群人。** 废科举后最痛苦的是已经在科举路上走了十几年的举人，文盲反倒不受影响。AI时代同理——教育路上投入最多的中间层家庭，承受的冲击远大于本就不依赖教育信号的底层。

**关键不同：速度和可见性。** 废科举是一个明确的制度断裂——上谕颁布之日起，所有人都知道旧游戏结束了。AI对教育信号的侵蚀是渐进的、弥散的、没有明确起止点的。没有一纸诏令宣布”学历不再值钱了”。信号是在无数个微小的市场变化中慢慢褪色的——一个雇主在面试中更看重项目经验而非学历，一个技术岗位取消了学历门槛改为技能测试，一个大厂的HR发现AI工具证书比MBA更能预测新员工的绩效。每一个微小变化都不构成新闻，但加在一起就是趋势。

渐进性使得应对更难。废科举后的士人至少知道自己需要转型。AI时代的教育消费者面对的是一种模糊的、无法定位的不安：旧路径似乎还在运作（孩子还在上学、还在考试、还能拿到学位），但回报在下降（同样的学位换来的起薪更低、岗位更少），而新路径还不清晰。

---

## 七、教育通胀的加速

这里需要引入一个周期概念：教育通胀。

教育通胀指的是：学历要求持续升高，但实际回报持续下降。它不是AI时代才出现的——它从20世纪中叶就开始了。1950年代，一个高中学历足以在美国获得一份中产收入的工作。1980年代，你需要大学本科。2000年代，很多岗位开始要求硕士学位。2020年代，某些行业甚至出现了”博士做本科生的活”的现象。

教育通胀的机制与货币通胀类似：当学位供给增加（高校扩招），单个学位的购买力下降。为了维持同样的”购买力”（就业竞争力），你需要更高的学位。这是一个经典的军备竞赛——每个人的理性选择（获取更高学历以保持竞争力）导致了集体的非理性结果（所有人都花更多时间和金钱在教育上，但相对位置没有改变）。

AI对教育通胀的影响是加速而非创造。机制是：AI压缩了中间技能岗位，雇主于是提高剩余岗位的学历门槛作为简便筛选，学位要求进一步升高。但学位背后的技能内容越来越多地可以被AI替代，学位的”实质价值”（区别于”信号价值”）持续下降。

这个螺旋的终点是什么？可能有两个方向。

**方向一：信号替代。** 当传统学历信号的信噪比降到一定程度以下，市场会集体转向替代信号——微证书（特定AI技能的认证）、项目作品集、在线声誉评分、实际任务表现数据。这已经在科技行业发生了：谷歌在2020年就宣布不再要求所有岗位必须有大学学位，转而依赖自己的技能评估测试。当这种做法从科技行业扩散到其他行业——而AI加速了这种扩散——传统学位的信号功能将进一步边缘化。

**方向二：信号崩盘。** 如果替代信号的建立速度跟不上传统信号的贬值速度，就会出现一个”信号真空期”——雇主不再信任学历，但也没有可靠的替代信号。在这种情况下，筛选退化为最原始的形式：关系网络和社会资本。谁认识谁、谁推荐了谁、谁的家族背景更可信——这些前现代的筛选标准在信号真空中重新抬头。这是一个讽刺的退步：教育体系花了两百年从”靠关系”进化到”靠考试”，AI可能在二十年内让它退回到”靠关系”。

无论走向哪个方向，对于当下的教育消费者——那些正在为子女教育做投资决策的家庭——信息都是明确的：教育投资的回报曲线正在变平，而且变平的速度在加快。

---

## 八、尾部风险——谁会最痛

但真正的痛点藏在”尾部风险”里——极端情况会怎样。在教育信号失灵的过程中，尾部风险的分布极不均匀。

以下几个特定群体承受的冲击最大：

**第一，高杠杆教育投资者。** 一个借了助学贷款读法学院的农村家庭，他们的教育投资用的是未来收入做抵押。当未来收入的期望值下降，债务的真实负担以杠杆率的倍数增加。一个背了三十万教育贷款的人，预期起薪从一万五降到一万，表面上收入少了33%，实际上整个财务计划的可行性被根本动摇了。

**第二，“教育改变命运”叙事的最强信徒。** 在中国，这特别指的是那些从小镇、县城通过高考进入城市的第一代大学生。他们的整个人生叙事建立在一个前提上：知识是改变命运的唯一可靠通道。当这个前提动摇时，他们面对的不仅是经济损失，还有存在论层面的打击——“如果教育不能改变命运，那我这二十年的苦读是什么”。这种打击比纯粹的经济损失更难修复。

**第三，正处于关键决策窗口的家庭。** 2027年前后，孩子正在读初中或高中的中产家庭面临一个历史上罕见的决策困境：按照旧规则继续投入（拼高考、冲名校），还是押注新方向（培养AI协作能力、鼓励项目实践、减少应试投入）？旧规则仍在运作但回报在降低，新方向还不清晰但可能是未来。在信息不完全的情况下，大多数家庭会选择旧规则——因为它至少是已知的。但选择旧规则意味着继续在一个正在贬值的资产上增加投入。

这三个群体的共同特征是：他们承担了教育投资的全部下行风险，却不掌握关于风险变化的充分信息。这是典型的”不对称风险”结构——上行收益被信号贬值吃掉了，下行风险仍然完全由个人承担，而真正掌握信息的一方（教育机构、政策制定者）没有动力传递这些信息。

---

## 九、如果不是为了分拣

分拣功能失灵之后，一个问题就绕不过去了：教育到底是为了什么？

这个问题听起来像一句空洞的修辞。但在2030年代，它变成了一个迫切的实际问题——越来越多的家长、学生和雇主都在问：我为什么还要花四年时间和几十万学费去获取一个越来越不值钱的信号？

可能的答案是：教育需要从分拣功能退回到一些更原始的功能上——那些在分拣被叠加之前，教育原本就在做的事情。

学会如何思考，而不是如何答题。学会与人合作，而不是在排名中击败他人。学会在不确定性中做出判断，学会定义问题而不只是解决问题，学会与AI协作完成复杂任务。

这些能力在分拣体系中从来不是重点——因为它们很难被标准化测量，无法测量的东西无法排名，无法排名的东西无法定价。但在AI时代，恰恰是这些无法被标准化测量的能力，构成了人类教育真正不可被替代的部分。

吊诡之处在于：教育体系花了一百多年把自己优化成了一台精密的分拣机器，现在它被要求变成一台完全不同的东西——而它几乎不知道该怎么变。

以制度拖曳系数衡量，教育可能是所有社会制度中最难转向的。大学花了近两百年才真正适应印刷术带来的知识传播变革。互联网出现三十年了，绝大多数大学的教学方式和1990年代比没有本质变化。MOOCs在2012年被称为”教育革命”，十年后全球大学招生人数不降反升，MOOCs的完课率仍然低于10%。

从1905年废科举到1977年恢复高考，中间隔了七十二年——期间经历了两次政权更迭、一次内战和一场文化革命。教育体系的结构性转型从来不是和平的、线性的、可预测的。

---

## 十、“终身学习”——从口号到生存诊断

教育信号的贬值带来了一个衍生命题：如果初始学历的信号功能在持续衰减，那么”一次性教育”的模式（前二十年投入、后四十年回收）就从最优策略变成了高风险策略。替代方案是”持续教育”或”终身学习”——不断更新知识和技能，以跟上市场需求的变化。

“终身学习”曾经是一句励志口号。在AI时代，它变成了一个生存诊断：你学习的速度是否快过你贬值的速度。

但这个诊断本身隐含着一个残酷的不平等。谁有时间终身学习？谁有资源终身学习？一个每天工作十小时、还要照顾家庭的中年人，和一个有充裕时间和经济后盾的年轻人，在”终身学习”的赛道上起点完全不同。终身学习在理论上是平等的——任何人都可以打开AI开始学习——但在实践中，它可能加剧而非缩小不平等。因为学习能力本身是分层的：受过良好基础教育、有自律习惯、有时间余裕的人，更容易利用AI进行有效学习；而那些基础薄弱、时间被生存挤占的人，面对AI反而可能更加迷茫。

“如何让所有人都终身学习”是旧框架内的增量优化。更值得问的是：当知识的获取和应用可以被AI以极低成本完成时，人类教育的不可替代价值到底在哪里？

答案可能指向一种元能力的培养，而不是具体知识的传授：在不确定性中保持判断力，在AI给出十个选项时知道选哪一个，在信息爆炸中识别什么值得关注。标准化考试培养不了这种能力（它培养的恰恰是”记住标准答案”），刷题也培养不了。它可能更接近苏格拉底式的对话，持续的追问、质疑、辩论，而不是工业化的知识灌输。

但苏格拉底式的教育几乎不可能被规模化。它需要一对一的深度互动，需要教师本身具备高水平的判断力和引导能力，需要足够的时间让思维在碰撞中生长。工业化教育体系之所以选择标准化考试而非苏格拉底对话，恰恰是因为前者可以规模化而后者不能。AI的讽刺在于：它让规模化教育（标准化知识传授）变得多余，同时让不可规模化的教育（判断力培养）变得更加稀缺和珍贵。

---

教育体系的转型不会在一代人的时间里完成。它太大了，太深了，牵涉的利益太多了。在漫长的过渡期内，数以千万计的家庭将在旧规则和新现实之间艰难地寻找平衡。

但教育只是故事的一部分。分拣机器筛选出来的人需要一个去处——一个让陌生人可以在共同的规则下协作的容器。过去四百年，这个容器有一个名字：公司。

---

**本章关键判断与置信度**

| 判断 | 置信度 | 说明 |
| --- | --- | --- |
| 教育体系的核心功能是分拣而非教育 | A | 制度分析支持，激励结构证据明确 |
| AI正在侵蚀学历信号的区分度 | A | 多条机制同时作用，方向确定 |
| 教育利益相关方有强烈的信息扭曲动力 | A | skin in the game 分析明确 |
| 中间层家庭是教育信号失灵的最大受损方 | A | 结构性分析清晰 |
| 教育通胀将被AI加速 | B | 方向确定但速度取决于多变量 |
| 教育转型需要15–20年以上 | B | 历史类比支持，但AI可能加速周期 |

**未解之问**

一、如果学历信号失灵但替代信号尚未建立，筛选会退化为关系网络——那么AI时代的社会流动性会高于还是低于教育时代？教育体系虽然不完美，但至少提供了一个形式上平等的通道；如果通道关闭，替代通道是否会更不平等？

二、教育如果从分拣功能中解放出来，它是否真的会转向”培养判断力”——还是会找到另一种分拣功能（比如根据”AI协作能力”排名）来维持自己的制度合法性？制度的自我保存本能是否会强过制度的功能转型？

# 第四章 公司——陌生人协作的最佳发明

> 人类花了四百年建造一种让陌生人彼此信任的容器，现在正亲手拆掉它的墙壁。

2029年，一个深圳跨境电商卖家发现了一件怪事：他的主要竞争对手，一个年销售额两千万的店铺，从头到尾只有一个人。一个人加一组AI Agent，管着选品、客服、投放、物流、财务。他自己的公司有十五个人，做着差不多的营收。他开始睡不着觉。

教育分拣出来的人需要一个去处。一个让互不相识的陌生人可以在共同的规则下聚集资本、分担风险、分享利润的容器。过去四百年，人类发明的最成功的这种容器叫做公司。

公司之所以存在，是因为在此之前，让陌生人协作的成本太高了。

## 一、公司为什么存在

追问到底：公司到底解决了什么问题？

1937年，一个26岁的英国青年写了篇论文，题目朴素得像一道课堂习题：《企业的性质》。罗纳德·科斯在这篇论文里提出了一个经济学界集体忽略了几百年的问题：如果市场是最高效的资源配置方式——买卖双方通过价格信号自由交易，资源自动流向最需要它的地方——那为什么还需要公司？

换一种更具体的方式问：你需要一个人帮你写报告，为什么不直接在市场上找一个人、谈好价格、付钱交货？你需要十个人帮你生产一批货物，为什么不逐个签合同、按件付费？为什么要把这十个人雇进同一栋楼、给他们发固定工资、建一个叫”公司”的组织来管理他们？

科斯的回答极其简洁：因为市场交易有成本。

找人要花时间，谈合同要花精力，盯履约要花心思，打官司更贵。当这些”交易成本”加在一起超过了在公司内部组织同样活动的成本时，理性的选择就是建一家公司——用雇佣关系替代市场交易，用行政命令替代价格信号，用科层制替代合同体系。

公司的边界由一个等式决定：边界内部，行政命令比市场交易更便宜；边界外部，反过来。

四十年后，威廉姆森把这个理论往下挖了一层。交易成本不是凭空产生的，它有三个根源：人的有限理性（我们不可能预见所有情况并写进合同）、人的机会主义（有机会就会占便宜）、以及资产专用性（当一项投资只对特定交易有价值时，投资方被锁定，对方可以趁机要挟）。公司本质上是一种针对这

三个人性弱点的制度解决方案：用长期雇佣关系来缓解有限理性问题（不需要事先把所有情况写进合同），用科层权威来抑制机会主义（老板可以直接监督），用内部化来消除资产专用性的要挟风险（供应商变成了自己的部门）。

科斯–威廉姆森框架是理解公司的钥匙。换一种更直觉的表述：公司是一台降低协调成本的机器。输入是资本、劳动和信息；处理机制是科层制和管理流程；输出是产品或服务。这台机器存在的唯一理由，是它的内部协调成本低于外部市场交易成本。

现在，AI正在同时改变等式的两边。而且改变的速度不对称——这是关键。

---

## 二、不对称的成本压缩

AI确实降低了公司内部的协调成本。自动化流程减少了人工操作，智能排期优化了资源分配，实时数据看板让管理层无需层层汇报就能看到全局，AI辅助决策工具把分析时间从”天”压缩到”分钟”。管理一个千人团队在信息处理层面接近于管理一个百人团队。

但更关键的变化在外部：AI以更快的速度降低了市场交易成本。

搜索成本？AI可以在秒级从全球供应商数据库中匹配最适合的合作方。协商成本？标准化的智能合约和AI协议生成器把谈判周期从数周压缩到数小时。监督成本？区块链加AI审计实现了近乎实时的履约追踪。执行成本？跨境支付和自动化仲裁让纠纷解决速度提高了一个数量级。

威廉姆森的三个交易成本来源也在同步松动。有限理性？AI可以处理比人脑多几个数量级的信息量，合同的”不完备性”大幅降低——AI甚至可以自动生成覆盖大量边界情况的智能合约。机会主义？AI驱动的信用评估和行为追踪让欺诈的成本急剧上升，透明度的提高压缩了机会主义的空间。资产专用性？当核心生产工具从专用设备变成通用AI模型时，锁定效应本身就在消解——同一个AI Agent今天给A品牌做客服，改一个提示词明天就能给B品牌做客服，不存在”专用资产”的问题。

用科斯的框架来看：当市场交易成本急剧下降时，公司的最优边界就会急剧收缩。很多以前必须在公司内部完成的活动——因为市场上买太贵了——现在通过市场交易来完成反而更高效。

这里有一个关键区分：AI对公司的影响不止于效率——它在改变公司存在的理由。当公司之所以存在的那个”交易成本差值”被系统性压缩时，公司的形态必然发生根本性变化——不是变大或变小的问题，而是存在逻辑本身的问题。

---

## 三、认知翻转——公司不是人的集合

但先要翻转一个直觉。

如果你问一个普通人”公司是什么”，最常见的回答是”一群人在一起工作”。走进任何一家公司，你看到的确实是一群人——坐在工位上，开会，打电话，对着电脑做事。这个直觉理解毫无问题，但它混淆了”公司的表象”和”公司的本质”。

公司的本质不是人的集合。它是一个法律虚构——一个由法律赋予”法人”身份的抽象实体，可以签合同、持有财产、承担债务、被起诉。这个法律虚构叠加一套协作协议（股权结构、雇佣合同、管理流程），构成了一个降低交易成本的工具。

人之所以出现在公司里，不是因为”公司的定义包含人”。而是因为在过去四百年的技术条件下，人类是唯一能够执行认知任务的实体。公司需要认知能力——信息处理、决策判断、沟通协调——而人恰好是唯一的认知供应商。于是”公司”和”一群人在一起工作”在观感上完全重合了，以至于我们忘记了：后者只是前者在特定技术条件下的表现形式，不是前者的定义。

这个区分在AI出现之前无关紧要。现在它变得关键了。因为AI是第二个认知供应商。

当公司可以从两个供应商——人类和AI——之间选择认知能力的来源时，“公司里有多少人”就不再是一个本质性的问题。一家没有员工、只有一个所有者和一组AI Agent的实体，在法律上仍然可以是一家公司——它仍然可以签合同、持有账户、承担责任。它只是不再有”一群人在一起工作”的外观了。

这个翻转的推论是深远的。如果公司的本质是”降低协作交易成本的工具”而非”人的集合”，那么当更好的工具出现时，公司的形态必然改变。正如汽车出现后马车的形态必然改变——人们仍然需要从A点到B点的运输，但实现这个功能的最佳载体变了。

---

## 四、公司的四次变形

回顾历史，公司的形态从来不是固定的。它随技术环境的变化而变形，每一次变形都不是渐进微调，而是对”什么才是最优协作方式”的根本性重新回答。

**第一次变形：从贸易冒险到有限责任（1600–1850）。** 1602年荷兰东印度公司（VOC）是第一家联合股份公司。它解决的核心问题是：如何让互不相识的投资者把钱汇聚到一起，投入跨越半个地球的高风险贸易。解决方案是两个法律创新——股份（所有权可分割、可转让）和有限责任的雏形（投资者的损失不超过出资额，尽管”有限责任”作为正式法律概念要到1855年英国《有限责任法》才确立）。这两个抽象概念将数千个没有信任基础的陌生人组织进了同一个经济实体。VOC的核心挑战是远距离协调——一道指令从阿姆斯特丹到巴达维亚需要六到八个月——所以组织结构是”极大授权+松散监督+事后问责”。

**第二次变形：从作坊到流水线（1850–1970）。** 1913年福特在底特律郊外的Highland Park工厂安装第一条移动装配线，T型车组装时间从十二小时三十分钟降到九十三分钟。福特解决的核心问题不是远距离协调，而是大规模标准化生产。他的方法是将复杂工作拆解为数百个简单动作，每个工人只负责一个动作。这要求严格的科层制——工人执行，领班监督，经理计划，高管决策。科层制的本质是一个

信息压缩系统：底层收集数据（产量、质量、库存），中层汇总分析（把一千条数据压缩成十条结论），顶层做出决策。每一层都是一个认知带宽的瓶颈节点。

**第三次变形：从科层制到精益网络（1970–2010）。** 丰田生产系统（TPS）颠覆了福特模式：从大批量降低单位成本变为小批量消除浪费，从决策集中在顶层变为决策权下沉到一线。互联网时代的谷歌和亚马逊进一步打破科层制——“两个比萨团队”原则是对小团队+高自治的极端追求。但即便是最扁平的互联网公司，其组织形态仍建立在一个核心假设上：团队的基本单元是人，人与人之间的沟通带宽有限，所以团队规模有上限，团队之间需要中层管理者来对齐信息、协调优先级、解决冲突。

**第四次变形：从人类网络到人机编队（2026– ）。** 这次变形没有一个明确的起始日期。如果一定要选标志性节点，按当前趋势推演，最早可能在2028到2029年间，全球主要科技公司的”人均Agent管理数”将首次超过”人均下属管理数”。一个中层管理者直接协调的AI Agent数量，超过了他直接管理的人类下属数量。

这个指标标志着一个结构性的翻转：公司的协调成本结构从”人–人协调为主”变成了”人–机协调为主”。而这两种协调的成本函数完全不同。

人与人之间的协调成本随人数增加呈超线性增长。两个人有一条沟通渠道，三个人有三条，十个人有四十五条，一百个人有近五千条。科层制的功能就是用层级结构把指数增长的沟通渠道压缩成线性的命令链条——代价是信息的损耗和延迟。

人与Agent之间的协调成本接近线性增长。管理一个Agent和管理十个Agent的认知负荷差距远小于管理一个人和管理十个人。Agent不需要你解释”为什么”，不需要你照顾情绪，不需要一对一会议了解职业发展诉求。你给它指令、约束和评估标准，它执行。

当这个成本函数的差异累积到组织层面，结果是颠覆性的。一个人加一百个Agent的协作效率，在大量任务类型上超过了一百个人的团队。不是因为Agent比人聪明——在2029年的技术水平上，单个Agent的”智慧”仍不如有经验的人类——而是因为Agent之间几乎没有协调摩擦。它们共享同一个知识库，遵循同一套规则，不存在信息不对称和利益冲突。

---

## 五、中层管理的消亡机制

科层制中，中层管理者的核心功能有两项：信息过滤和跨部门协调。

信息过滤：底层产生大量原始数据（销售数字、客户反馈、运营异常、市场动态），高层不可能直接消化这些原始数据，认知带宽有限，需要中层将其压缩、摘要、解释、排序。一个部门总监的日常工作，大部分就是在做这件事：把下面一百条信息变成上面十条结论。

AI让信息过滤可以自动化完成。实时数据看板直接把原始数据转化为决策者可消费的形式，不再需要人类中间层的压缩。AI生成的分析报告在信息完整度和更新速度上往往超过人类中层，因为AI不受认

知带宽的限制。

跨部门协调：当公司的不同部门（销售、产品、运营、财务）需要对齐优先级、解决冲突时，中层管理者充当的是”翻译器”和”仲裁者”——把A部门的需求翻译成B部门能理解的语言，在资源有限时做出分配决策。

AI通过共享知识库和规则引擎让跨部门协调成本趋零。当所有Agent读取同一个数据源、遵循同一套优先级规则时，“部门墙”这个概念本身就失去了物理基础。不存在A部门和B部门的信息不对称，因为没有”部门”——只有一个统一的Agent网络在执行一组连贯的规则。

因果链如下：

AI让信息传递和协调成本趋零（输入变化）→ 中层管理者的两项核心功能（信息过滤+跨部门协调）失去价值（功能替代）→ 但制度惯性（既有的组织架构、已签的劳动合同、中层管理者的政治影响力）让这个过程花5–10年而非1–2年（拖曳效应）→ 中层管理者不是被”解雇”而是被”掏空”——title还在但实质功能消失（过渡形态）→ 最终，新一代公司从设计之初就不再设置中层（新均衡）。

中层管理的消亡不像流水线工人被机器人替代那样清晰可辨。间奏二中那位广告公司的文案总监用了一个精确的比喻来形容这种感觉：“从厨师变成了试菜的。”菜还是那些菜，你还坐在厨房里，但你不再掌勺了。这就是功能抽空的真实体验。你的头衔还在，你的办公室还在，你每天还在开会，但会议讨论的内容越来越多地变成”Agent的这个输出结果需不需要调整”，而不是”这个季度的策略方向是什么”。策略方向已经由顶层设定、Agent在执行层面优化。你成了一个看着Agent干活、偶尔点头或摇头的人。你的工作越来越像一道签名——法律上需要，操作上可有可无。

---

## 六、“10人+200万Agent”——物种更替而非效率提升

大多数关于AI改变公司的讨论都停留在效率提升的层面——“AI让公司更高效了”。但同样的公司运转得更快不是重点——一种全新的组织形态正在出现。

这种新形态可以概括为一个公式：少量人类 + 大量Agent = 新型组织。

表面上看，这只是”人变少了”。但当人数从数百人缩减到个位数时，组织的性质发生了质变——不是量变。因为组织中人的角色从”执行节点”转向了”宪法层”。

什么是”宪法层”？一个类比：一个国家的宪法不决定每一天的具体事务（那是行政和司法的工作），宪法决定的是游戏规则本身——什么权力属于谁，什么行为被允许，什么边界不可逾越。在新型组织中，人类做的就是这件事：不做具体的决策（大部分运营决策由Agent根据预设规则自动执行），而是设定规则本身。

用什么标准判断一个产品值不值得上架？退货率达到多少自动下架？品牌调性的边界在哪里？哪些决策Agent可以自主做，哪些必须提交人类审批？——这些”元规则”的制定，是人类在新型组织中的核

心功能。

在这个框架下，人类不在Agent网络内部——人类在网络上方，或者更准确地说，在网络之前。他们定义网络的运行规则，然后退到一边，只在异常情况下介入。

这与传统公司中人的角色有根本区别。传统公司中，人是每一个操作的执行者：分析数据、写报告、做决策、协调资源。新型组织中，人是规则的制定者和例外的处理者——这需要的人数远少于前者。

一个2030年前后出现的典型新型组织结构只有三层：

**宪法层**（5–10人）：CEO、CTO和核心决策者。他们的工作不是做决策——大部分运营决策已被Agent自动化——而是设定决策的规则框架。

**专家层**（30–60人）：各领域的深度专家——供应链谈判、品牌设计、法律合规、财务审计。他们的角色不是日常执行（日常由Agent完成），而是处理Agent标记为”需要人类介入”的异常案例。一个资深供应链专家一天可能只处理三到五个这样的案例，但每个案例可能涉及数百万的决策。

**运维层**（数百人）：Agent系统的维护者——prompt工程师、数据工程师、系统运维、质量审核员。他们确保Agent网络正常运转、处理边界情况、持续优化系统性能。

这不是”更精简的传统公司”。这是一种新物种。

---

## 七、科斯定理的另一面——市场也在变形

到目前为止的分析集中在公司边界的收缩。但科斯框架有另一面：当市场交易成本下降时，不仅公司会收缩，市场本身的形态也会变化。

传统市场的一个核心限制是：市场交易需要标准化。你在市场上买卖的东西必须是可描述的、可定价的、可交付的。这就是为什么标准化商品（大宗商品、标准零件、标准化服务）在市场上交易，而高度定制化的需求（战略咨询、创意设计、复杂谈判）更多在公司内部完成——因为后者太难标准化了。

AI改变了这个限制。AI让高度定制化的认知服务也可以在市场上”标准化交易”——不是通过简化服务本身，而是通过AI作为中介来消除供需两端的信息不对称。一个需要定制化法律意见的小企业，以前必须雇一个全职法务或签一个昂贵的律所年度合同（因为找到合适的律师、说清楚需求、确保质量太费时间）。现在，AI法律平台可以在几分钟内理解需求、匹配合适的服务提供者（可能是AI，可能是人类律师，可能是两者的组合）、标准化交付流程、自动检查质量。

这意味着：不仅公司的边界在收缩，“市场能交易什么”的边界在扩张。越来越多以前只能在公司内部完成的活动，现在可以在市场上以接近零的交易成本完成。

对于一个人加一组AI Agent的运营者来说，整个世界就是他的”公司”——他不需要雇任何人，只需要在需要的时候从市场上购买特定服务。设计、翻译、法律咨询、税务申报、物流协调——全部按需购买，按次付费，没有长期雇佣关系的负担。

这是从科斯定理推导出的逻辑终点：当市场交易成本趋近于零时，公司的最优规模也趋近于零——即一个人。

当然，这是一个理论极端。现实中的公司不会全部缩减到一个人。有些活动——需要高度互信的长期合作关系、需要共享隐性知识、需要实时的面对面协调——在可预见的未来仍然在公司内部完成更有效。但方向是明确的：公司在缩小，市场在扩大，而两者之间的界面正在被AI重新定义。

---

## 八、手工作坊的回声

历史上有一个精确的结构类比。

18世纪英格兰，工业革命之前，纺织品由分散的手工作坊生产——一个家庭就是一个生产单元，从中间商处领取原料，在家中纺纱织布，交回成品。这个体系叫”包买制”（putting-out system），本质上是一个市场交易网络，不是公司组织。

然后水力纺纱机出现了。阿克莱特1771年建第一座水力纺纱厂，将几百个工人集中在一栋建筑里。工厂制由此诞生。从包买制到工厂制的转变，是一次组织形态的”聚合”——把分散的个体生产者聚合进集中的工厂。驱动力是蒸汽动力和机械化——这些技术需要集中的资本投入、集中的能源供应、集中的协调管理。新技术条件下，集中化就是最优的生产组织方式。

今天正在发生的是反方向的过程——一次组织形态的”解聚”。AI不需要集中的厂房，不需要集中的管理，不需要人与人之间物理共处才能协调。当协调可以通过数字化方式以接近零成本完成时，“把人聚合在一个组织里”的必要性就下降了。

但历史类比的价值在于找差异，不只是找相似。

**差异一：速度。** 从包买制到工厂制的转变花了约七十年（1770年代到1840年代），旧形态和新形态长期并存。今天的转变会更快——数字技术的扩散速度远超蒸汽机——但也不会在十年内完成。大公司不会消失，正如手工作坊在工厂制成熟后仍以”手工艺品”的形式存活（从主流变成了小众和高端）。未来的大公司也会继续存在，但它们的内部结构将面目全非。

**差异二：方向。** 包买制→工厂制是从分散到集中。大公司→新型组织是从集中到分散。两次变形的驱动力相同——新技术改变了最优协调方式——但方向恰好相反。

**差异三：社会后果。** 包买制到工厂制创造了一个新阶层：工业无产阶级。数百万从乡村涌入城市的前农民成为工厂工人，他们的集体力量催生了工会运动、劳动法和福利国家。大公司到新型组织又会创造什么？那些一个人加一套AI工具就能产出以前一个团队产出的人，他们既不是传统的”雇主”（不雇

人），也不是”雇员”（没老板），也不完全是”自由职业者”（有持续运营的业务实体）。他们是一个新范畴，需要新的法律身份、新的社会保障方式、新的政治代言。但按目前的制度演进速度，到2031年这些可能仍不存在。

---

## 九、十四个人去了哪里

每一个”一个人做了以前十五个人的活”的故事，都有一个被省略的下半句：那十四个人去了哪里。

逆向追问——不问”赢家怎么赢的”，而问”输家怎么输的”。

一部分人找到了新的高价值位置。在技术转型中，总有一些人凭借学习能力、适应性或运气，成功转型到新型组织的宪法层或专家层。他们的收入可能不降反升，因为新型组织的利润率更高（人力成本大幅降低），而少数留下来的人可以分到更大的份额。

一部分人进入了运维层。他们从”做业务”转为”维护做业务的AI系统”——数据标注、质量审核、系统监控。这些岗位存在且必要，但薪资水平通常低于他们此前的岗位，且技能的市场议价能力较弱（因为这些技能的替代性强）。

还有一部分人漂在”脆弱的自由”和”不完全就业”之间——有的做自由职业和平台零工，获得了时间自由但失去了收入稳定性和社会保障；有的名义上还在职，但工作内容被掏空，薪资持续缩水。两种状态的共同特征是：看起来还在系统里，实际上已经脱离了上升轨道。

这三种去向的比例因行业和地区而异，但总体趋势是：高价值位置的容量有限，大部分人流向了后两种状态。“公司变小了，但每个人的价值更高了”——这只是对赢家的描述。对十四个人中的大多数而言，这是一个价值流失的过程。

---

## 十、预测——2035年的”公司”

从周期视角来看，2026到2035年是公司形态变革的”加速期”。在这个阶段，新旧形态并存，但力量对比在持续倾斜。

可以做出几个方向性的判断：

**判断一：大公司不会消失，但会变空。** 品牌效应、监管牌照、客户信任、网络效应——这些护城河仍然让大公司有存在理由。但大公司的内部人员规模会大幅缩减。2035年的”大公司”可能有数百亿营收，但只有数百名核心员工——加上数十万Agent。“大”的含义从”人多”变成了”Agent多”。

**判断二：“公司”这个词仍在使用，但含义已变。** 就像”电话”这个词仍在使用，但2026年的”电话”和1985年的”电话”是完全不同的设备。2035年的”公司”和2020年的”公司”在法律外壳上可能相似，但在组织内核上已面目全非。

**判断三：新的法律框架尚未形成。** 一人公司的营收达到数千万时，现行的税务、劳动保护、消费者权益保障法律框架全面不适用。新的法律框架需要时间来形成——就像有限责任公司的法律形式在VOC成立（1602）之后花了两百多年才在全球标准化（英国1855年通过《有限责任法》，中国1993年才有《公司法》）。

**判断四：人在公司中的角色从”认知执行者”转向”规则制定者+例外处理者+责任承担者”。** 最后一项尤其重要——责任不能被外包给算法。当AI做出了一个错误的决策，需要有一个人类承担法律和道德责任。这意味着”人在公司中的存在”不会降到零——但人存在的理由从”干活”变成了”负责”。

---

这些判断指向一个更大的图景：公司作为”让陌生人大规模协作”的工具，它的四百年历史正在进入一个新的阶段。旧工具（以人为节点的科层制组织）正在被新工具（以AI为节点的协议网络）补充乃至替代。但工具在变，需求没变——人类社会仍然需要大规模协作来完成复杂任务。只是协作的形式，从”把人聚在一栋楼里”变成了”让Agent在云端协调”。

公司是协作的容器。教育是进入容器的筛选机制。劳动是使用容器的方式。认知垄断是整套安排的前提。

四个前提已经全部揭示。但还有一个东西把它们绑在一起——一个让大多数人相信这套安排是公平的、值得参与的叙事。这个叙事叫做”中产阶级”。

---

**本章关键判断与置信度**

| 判断 | 置信度 | 说明 |
| --- | --- | --- |
| AI降低市场交易成本的速度快于降低内部协调成本的速度 | A | 多条证据支持，方向确定 |
| 公司的本质是交易成本优化工具而非人的集合 | A | 科斯–威廉姆森框架，经济学共识 |
| 中层管理将经历功能抽空 | A | 已在发生，趋势确定 |
| “一人公司”将成为重要的组织形态 | B | 趋势明确但规模取决于多变量 |
| 大公司在2035年前不会消失但会”变空” | B | 方向确定但速度不确定 |
| 新的公司法律框架需要10–20年形成 | B | 历史先例支持，但具体时间有变数 |

**未解之问**

一、如果公司的最优规模因AI而急剧缩小，“就业”作为个人获得收入和社会身份的主要渠道是否还能维持？当大多数经济活动不再需要大量人类参与时，那些既不是一人公司创业者、也不是宪法层决策者的人，在经济系统中的位置在哪里？

二、AI让协调成本趋零，但人类对”共同在场”和”集体归属”的需求不会趋零。未来的组织形态如何同时满足效率需求和归属需求？还是说，这两种需求将被分离到不同的社会结构中——效率在Agent网络中实现，归属在社区和社群中满足？

# 第五章 中产叙事——工业文明的稳定器正在失灵

> 最高效的控制从来不靠铁链，靠的是让每个人相信自己正在自由攀爬。

认知垄断（第一章）→ 劳动契约（第二章）→ 教育分拣（第三章）→ 公司容器（第四章），四个前提已经被逐一揭示。但还有一个东西把它们绑在一起：一个让数以亿计的人相信这套安排是公平的、值得参与的、只要努力就会有回报的叙事。

这个叙事是工业文明最精妙的发明：中产阶级。

## 一、不是一个经济分类，而是一个故事

翻开任何一本经济学教科书，“中产阶级”的定义都是一个收入区间——年薪多少到多少之间，你就是中产。但这个定义遮蔽了真正重要的东西。从文明史的角度看，中产阶级从来不是一个经济分类。它是一个叙事。

这个叙事的核心内容可以浓缩为一句话：**通过教育和努力，你可以实现阶层上升。**

这句话在人类历史的绝大部分时间里不成立。一个中世纪农奴的儿子大概率是农奴，一个清代佃户的女儿大概率嫁给另一个佃户。社会流动性在前工业时代极低——不是因为人们不够努力，而是因为制度不允许。出生决定阶层，教育是贵族的专利，技能传承通过血缘和师徒，不通过公开竞争。

工业革命改变了这一切。大规模工厂需要大量具备基本读写算能力的工人，公共教育应运而生，而公共教育创造了一条前所未有的通道——至少在理论上，一个矿工的儿子可以通过上学、考试、获得证书进入比他父亲更高的社会阶层。这条通道的实际宽度随时间和地域变化很大，但它的存在本身就构成了一个强大的叙事：制度是公平的，上升是可能的，你只需要足够努力。

这个叙事在二十世纪中期达到顶峰。美国的GI法案让数百万退伍军人上了大学，催生了人类史上最大规模的中产阶级。中国改革开放后的四十年，高考制度扮演了类似角色——一个农村孩子通过高考进入北大清华然后获得体面城市生活的故事被讲了无数遍，每讲一遍都在巩固同一个叙事。

追问到底：这个叙事的真正功能是什么？

它解决了一个困扰所有复杂社会的核心问题：**如何让大多数人接受现有秩序。**

在前工业社会，答案是宗教和暴力——你服从是因为神要求你服从，或者因为不服从就会被杀。工业社会发明了一个更优雅的答案：你服从是因为你相信自己可以上升。只要上升的通道是开放的，底层的人就会把精力用于攀爬而不是颠覆。

中产叙事是社会稳定的最佳保险。它让人们在一个显然不平等的系统中保持合作意愿——不是通过强制，而是通过希望。

理解了这一点，你就能理解为什么AI对中产叙事的威胁比对中产收入的威胁更危险。

---

## 二、三根支柱的断裂机制

中产叙事的可信度依赖三根支柱。每根支柱都是这台”社会稳定器”中一个关键齿轮。AI正在同时松动这三根支柱，每一根有独立的断裂机制。

**第一根支柱：教育投资有回报。**

这根支柱的运行逻辑很简单：一个家庭投入时间和金钱让孩子接受教育，孩子获得专业知识和证书后进入高薪职业，收入覆盖了教育成本还有溢价，于是继续为下一代投资教育。循环往复，代代如此。

这条逻辑在过去一百年大体成立。美国大学毕业生的终身收入平均比高中毕业生高出约一百万美元。中国985高校毕业生的起薪长期高于普通本科的两到三倍。

AI的断裂机制是：降低专业知识的获取门槛，从而压缩教育证书的信号价值（第三章已详述）。当一个高中辍学者配合AI可以产出与名校毕业生质量相当的法律文书或商业分析时，雇主为名校学历支付的溢价就失去了经济理性。教育投资的回报率在下降——不是归零，但在以可测量的速度下降。

**第二根支柱：专业技能有溢价。**

逻辑同样清晰：掌握稀缺的专业技能，市场对该技能的需求大于供给，于是获得高于平均的薪酬。溢价激励更多人学习该技能，供给增加后溢价在新均衡处稳定。

律师比建筑工人贵，不是因为律师的工作更辛苦（工地搬砖的体力强度远高于写合同）。律师贵是因为能通过司法考试的人少，法律推理能力稀缺。医生、金融分析师、建筑师、咨询顾问，所有”专业人士”的收入溢价都来自同一个来源：技能的稀缺性。

AI的断裂机制是：让稀缺技能变得不再稀缺。AI没有完全替代专业人士，而是把过去需要十年训练才能获得的标准化专业能力变成了一个API调用。一个从未学过法律的人配合AI法律助手可以完成80%的合同审查。那剩余的20%仍需要专业律师，但20%的工作量只需要过去20%的人。

溢价还在，但享受溢价的人数大幅减少。对个体而言，这是一个残酷的筛选：如果你不是那20%，你的专业训练就从资产变成了沉没成本。

**第三根支柱：职业阶梯存在。**

逻辑更直观：从初级岗位入行，通过经验积累逐级晋升，每一级的薪酬和权责有明确提升。晋升路径的可见性激励你持续投入，整个组织通过这把阶梯实现人才筛选和激励。

法务专员→高级法务→法务经理→法务总监→法务VP→CLO。每一级需要两到三年，每一级薪资涨幅在20%到50%之间。阶梯不仅提供经济激励——它提供时间结构和意义感。你知道自己在哪里，知道下一步在哪里，知道只要持续努力就会到达。

AI的断裂机制更隐蔽。它不是砍掉阶梯的某一级，而是压扁整个阶梯。当AI接管了大部分初级和中级工作（第四章已分析），组织不再需要那么多层级。五个人的法务团队不需要VP。三个人的分析师团队不需要"高级分析师"这个台阶。阶梯变短，台阶变少，能站在阶梯上的人也变少。

三根支柱的断裂不是各自独立的——它们互相加速。教育回报下降，家庭开始降低教育投资，专业技能的供给结构随之变化，稀缺性降低，溢价进一步下降，职业阶梯进一步被压扁。这是一个自我强化的下行螺旋。

---

## 三、不是经济事件，是社会事件

到这里，必须逆向追问：AI需要消灭中产的收入吗？

不需要。它只需要消灭中产叙事的可信度。

这个区分至关重要。因为收入下降是经济事件，而叙事失效是深层社会事件。两者的后果完全不同。

收入下降但仍然相信"再努力一把就能恢复"的人，会继续工作、消费、还贷、给孩子报辅导班。他沮丧，但还在轨道上。真正危险的是另一种人——不是收入归零了，而是不再相信努力有用了。他可能还在上班，但开始敷衍；可能还在消费，但不再为未来储蓄；可能还有孩子，但不再执着于让孩子复制自己的路径。他没有离开系统，但他对系统的信任已经撤回了。

当系统的质疑者达到临界规模时，社会稳定器就失灵了。

历史上最精确的先例是工业革命时期英国手工织工的命运——第七章将详细展开这个跨越七十年的缓慢替代过程。这里只需抓住一个要点：卢德运动的本质不是反对机器，而是一群发现"努力就能维持体面生活"这个叙事不再成立的人，对叙事失效的绝望反应。他们砸的不是机器，是一个不再兑现承诺的社会契约。

AI时代的中产叙事失效可能不会以砸机器的方式表达。间奏零中的赵敏正是这种沉默失效的缩影——她不愤怒、不抗议，只是在投了五个月简历之后，逐渐意识到自己花十五年积攒的工作经验在一个正在重新洗牌的世界里可能已经不算数了。信息时代的社会表达形式不同，但底层的社会心理动力学是相同的：当足够多的人不再相信规则是公平的，他们会降低对制度的参与意愿和合作积极性。

---

## 四、四种资本的权重迁移

如果只从收入角度看中产危机，你会错过问题的大部分。

一个多世纪前，马克斯·韦伯就拒绝把社会分层简化为收入高低。他拆出三个独立的维度：经济（收入和财富）、地位（社会声望）、权力（对决策的影响力）。三者相关但不等同。一个大学教授的收入可能不如一个包工头，但社会声望更高；一个政府处级干部的收入可能不如一个程序员，但对公共决策有更大影响力。中产阶级的稳定性不仅取决于收入，也取决于声望和权力。

要分析AI对中产的多维冲击，布迪厄的资本理论提供了更锐利的透镜。他把”资本”一词从银行账户里解放出来，划分为四种形态：经济资本（钱）、文化资本（知识、品味、学历）、社会资本（关系网络、信任）、符号资本（声誉、注意力）。这四种资本在工业社会有一个稳定的权重分布。AI正在剧烈地重新排列这些权重。

**文化资本在贬值。** 这是中产阶级最核心的护城河。名校学历、专业证书、行业知识，构成了一道其他阶层难以逾越的门槛。AI正在拆除这道门槛。当知识可以被AI即时调用，学历的信号功能贬值，行业知识的获取门槛降低，文化资本在四种资本中的相对权重正在下降。这对中产的打击是存在性的：如果你的阶层位置主要靠文化资本维持，而文化资本在贬值，你的阶层位置就在动摇，哪怕你的银行账户数字没变。

**社会资本在升值。** 在一个AI可以做大部分认知工作的世界里，“谁信任你”比”你知道什么”更有价值。信任不能被数字化、不能被API调用、不能被规模化复制。一个拥有深厚信任网络的人，即便他的专业知识不如AI，在很多场景中仍不可替代。信任的建立需要时间、共同经历、反复验证，而这些恰恰是AI无法加速的。

**符号资本成为新战场。** 注意力、个人品牌、公共影响力——这些在工业时代被视为”软实力”的东西，在AI时代正在成为硬通货。当内容可以被AI无限生成，能够吸引和维持人类注意力的能力成为新的稀缺资源。一个拥有百万粉丝的知识博主，其符号资本的经济变现能力可能超过一个年薪百万的企业高管。

**经济资本在分化。** 中产家庭的财富主要由两部分构成：房产和人力资本（即未来收入的现值）。在中国，房产占家庭财富的比例约70%。当薪资增长放缓甚至停滞，人力资本的现值下降，但房贷是刚性的，每月一万九不会因为薪资不涨就自动减少。负债端刚性，资产端在缩水，资产负债表失衡。这不是”变穷了”，是”杠杆在反向作用”。

四种资本权重的迁移意味着：一个按照旧权重分配资本的人，把大量时间和金钱投入文化资本（学历、证书、专业知识），而忽略社会资本和符号资本，在新的权重体系下处于结构性劣势。他的价值储备在贬值的类别里，而他缺乏正在升值的类别。

这种错配不是个人失误，它是制度性的。教育体系教你积累文化资本，公司体系教你把文化资本变现为经济资本，整套中产路径就是一条文化资本的积累和兑现通道。当这种资本在贬值时，整条通道的价值都在缩水。但通道上的人已经走到了中途，退无可退。后来在南方某个老小区里，一个做了多年文

案总监的女人会用一个精确的比喻来形容这种处境：从厨师变成了试菜的。菜还是那些菜，你还坐在厨房里，但你不再掌勺了。

---

## 五、基础设施权力——新的分界线

权力维度的变化可能是三个维度中最危险的。

在工业社会，权力的基本分界线是：谁拥有生产资料。工厂主拥有机器，工人出卖劳动力——这是马克思观察到的基本结构。尽管过去两百年发生了很多变化（工会运动、社会保障、股票市场的大众化），这个基本结构从未被真正颠覆。

AI时代的权力分界线正在从”谁拥有生产资料”转移到”谁控制认知基础设施”。

开发和运营AI系统的平台方、制定AI政策的监管者、设计AI工作流的技术架构师——这些角色拥有的不是传统的”管理权力”，而是一种更底层的”基础设施权力”。如同掌握电网的人比使用电灯的人拥有更根本的权力，掌握认知基础设施的人比使用AI工具的人拥有更根本的权力。

中产阶级，无论他们的收入如何，在这个新的权力结构中大多是使用者，而不是控制者。他们使用AI工具工作，但不决定工具的设计逻辑。他们受算法影响，但不参与算法的制定。他们的数据训练了AI模型，但不分享模型创造的价值。

这是一种新型的结构性无权。与工业时代的无权不同，它不表现为赤裸裸的剥削——没有流水线上的血汗，没有十二小时的强制劳动。它表现为一种温和的、几乎不可见的依赖：你的工作效率依赖AI，你的信息获取依赖算法推荐，你的消费选择被精准营销引导，你的职业命运被你无法理解的模型决定。每一层依赖都不觉得像”被控制”，但叠加在一起，就构成了一种系统性的权力不对称。

权力集中本身不是新闻。新闻是集中的方式变了。

AI时代的权力集中有一个新特征：它不需要故意集中，它会自动集中。AI系统具有极强的规模效应——模型越大越强，数据越多越准，用户越多越好用。这种正反馈循环天然地导向赢家通吃，使得权力自动集聚在少数拥有算力、数据和模型的实体手中。这不是阴谋的结果，而是技术经济学的结构性特征。

---

## 六、四种国家形态的中产危机

中产叙事是全球性的，但它在不同国家有非常不同的版本。AI的冲击方式也因此各不相同。

**中国：体制内缓冲 + 高考叙事。**

中国的中产阶级大致可分为两个群体：市场化中产（互联网、金融、外企等）和体制内中产（公务员、事业编、国企正式工）。AI首先冲击的是市场化中产——第二章到第四章的分析主要描述的就是这个群体的处境。

体制内的缓冲作用不仅是经济层面的。在中国的社会心理中，“编制”的吸引力远不止工资——它代表一种确定性，一种人生不会脱轨的承诺。父母为子女谋求编制的执念，本质上是对中产叙事最极端的投注——不仅相信教育能改变命运，还要用制度把命运改变的成果锁定下来。

但体制内的缓冲有极限。如第二章所述，这个庞大群体的工资来源是税收，税收依赖市场经济的活力。当AI压缩了市场化就业的收入和规模，税基收窄，缓冲器可以延缓冲击，但不能消除冲击。

更深层的问题在于高考叙事。中国的中产叙事比美国更具体——它不是模糊的”美国梦”，而是高度具象化的”高考梦”：任何孩子都可以通过高考改变命运。这个叙事的说服力来自它的可验证性：每年都有真实案例证明小镇做题家可以通过高考进入清北，获得优渥的城市生活。AI不是让高考变得不公平——而是让高考通向的那些职业变得不再那么值钱。当名校毕业生的就业溢价下降，高考叙事的可信度就会动摇。而这个叙事承载的不仅是个人前途——它是中国社会流动性承诺的核心机制。

**美国：信用消费泡沫。**

第二章已分析过美国中产生活方式对信用消费的高度依赖。这种结构脆弱性在中产叙事失效的语境下呈现出新的面向：中产叙事失效在美国不只是心理冲击——它直接变成财务危机。

美国中产叙事的独特之处在于：它不是”通过储蓄实现上升”——美国的储蓄率长期接近零——而是”通过借贷维持当前生活水平，然后通过未来收入增长来偿还”。这个叙事的前提是收入会持续增长。当AI压缩白领薪资，一个背负二十万美元学生贷款的法学院毕业生，如果预期起薪从十五万降到十万，整个财务计划的可行性可能被根本动摇。美国的中产危机很可能首先以债务危机的形式爆发。

AI带来的不是一次性冲击，而是持续的、渐进的薪资压缩——这种”慢性出血”对高杠杆家庭的破坏力可能超过一次猛烈的短期冲击。

**北欧：高福利缓冲但税基受压。**

丹麦、瑞典、芬兰的中产阶级与中美有一个根本区别：他们的生活水平更少依赖个人收入，更多依赖公共服务。免费医疗、免费教育（含大学）、慷慨的失业保险和育儿补贴意味着：一个北欧中产家庭即使收入下降30%，其生活质量的实际下降幅度远小于中国或美国同等降幅的影响。

北欧模式的缓冲是真实的。但它有一个结构性脆弱点：高福利依赖高税收，高税收依赖充分就业。当AI压缩就业和收入，税基缩小，福利体系的资金来源就会受到挤压。北欧国家已经在讨论”AI税”——对使用AI替代人力的企业征收额外税费——但这类政策设计极其复杂，在全球化环境下面临资本外流的风险。

**日本：空心化。**

如第二章所述，日本终身雇佣制这一将个人身份深度嵌入公司身份的制度，从1990年代泡沫崩溃后就开始瓦解，但文化惯性极其顽固。AI的冲击让这个本已脆弱的制度承受了新的压力。

结果是中产叙事在日本经历一种独特的”空心化”：外壳还在（大公司品牌、终身雇佣的承诺、年功序列的薪资），内容被掏空。新一代年轻人面对的是一个承诺不再兑现但尚未正式撤回的制度。这种暧昧状态比公开的结构性重组更令人困惑。你无法对一个”还在但已经不管用”的东西提出明确的替代方案。

---

## 七、叙事失效的社会动力学

叙事失效的后果不在经济报表上。它在更深的地方起作用。

历史上反复出现一个周期：生产力提高，财富增长，分配不均加剧，社会适应压力增大，制度调整与深层社会转型，新的均衡形成。AI对中产叙事的冲击正在激活这个周期的”分配不均加剧，适应压力增大”阶段。

但这一次的周期有一个新特征：被冲击的不是底层（底层在历史上反复被冲击，已经发展出了生存策略），而是中间层。中间层的社会角色与底层完全不同。底层的不满通常是局部的、分散的、难以形成持续影响的。中间层的诉求则有更强的组织能力（他们受过教育、有社交网络、有表达能力），更广泛的社会影响力（他们是消费的主力、舆论的主体、选票的大头），以及更强的合法性基础（他们代表”主流社会”）。

当中间层的叙事失效时，它的社会能量不会消失——它会被重新导向。历史上，中产叙事失效的社会后果有两种方向：

**方向一：制度改革。** 如果政治系统有足够的弹性和回应能力，中间层的不满会被转化为制度创新。1930年代美国大萧条时期，中产阶级的大面积破产催生了罗斯福新政：社会保障法、最低工资法、联邦存款保险。这些制度重建了中产叙事的可信度，但叙事的内容变了，不再是”你靠自己就能成功”，而是”社会会在你跌倒时接住你”。

**方向二：社会极化。** 如果制度缺乏回应能力，中间层的不满会被引导向替罪羊——移民、少数族裔、外国竞争者、“精英阶层”。2016年特朗普当选和英国脱欧都有这个要素：中产叙事失效后的焦虑被引导向”华盛顿精英”和”布鲁塞尔官僚”。

AI时代的中产叙事失效很可能同时激活两个方向——在不同国家以不同比例混合。中国的制度框架更倾向于方向一（自上而下的制度调整），美国更可能在方向二（社会极化）上走得更远，北欧有更多资源和传统来选择方向一。多个东亚社会在教育转型期经历的社会适应压力也表明，这种叙事调整本身是一个跨文化的共性现象。

但无论走向哪个方向，一个基本事实是：叙事一旦失效就很难修复。你可以提高收入、改善福利、调整政策——但你很难让一个已经不再相信”努力就能成功”的人重新相信这个故事。信念的建立需要代际级别的时间积累，而信念的消退可以在几年内完成。这种不对称性让叙事失效成为一种几乎不可逆的深层社会转型。

---

## 八、新的不平等面孔

让我们做最后一层推演。

不平等的形式一直在变，但底层逻辑始终是认知差距——谁掌握信息，谁做决策，谁定义”什么是有价值的”。封建时代靠土地，工业时代靠资本，但两者的深层都是：谁掌握了管理资源和组织生产的知识。

当认知可以被AI批量生产和分发时，不平等不会消失——它会换一副面孔。

过去，不平等的核心是”认知的拥有”——你知道别人不知道的东西，所以你有优势。现在，不平等正在迁移到更底层——“认知基础设施的控制”。你使用AI和我使用AI，但谁设计AI、谁拥有训练AI的数据、谁定义AI的优化目标——这些问题的答案决定了权力的分布。

更微妙的一层：AI时代的不平等可能表现为一种”舒适的依赖”。每个人都在使用AI，每个人的生产力都提高了，每个人的生活都更方便了。但少数人控制着整套系统的规则和方向。这种不平等不会引发传统的阶级冲突（因为底层的绝对生活水平可能在提高），却会产生一种另一种异化：你在一个你无法理解、无法影响、无法退出的系统中”舒适地”运行着。

这是赫胥黎在《美丽新世界》中描述的那种不平等——通过满足而非压迫实现。你在一种精心设计的舒适中，渐渐失去追问的意愿。

---

## 九、第一部分的收束——五块松动的石板

本书的第一部分到这里结束。让我们回顾过去五章揭示了什么。

第一章揭示了一个存在了三十万年的前提：人类是唯一能处理复杂信息的物种。AI打破了它。

第二章揭示了一个存在了两百年的前提：每个成年人都必须通过劳动来获得收入、身份和社会参与资格。AI正在动摇它。

第三章揭示了一个存在了一千三百年（以科举为起点）的前提：社会通过标准化考试来分拣人力、分配资源。AI让这台分拣机器开始失灵。

第四章揭示了一个存在了四百年的前提：人类需要公司这种组织形式来完成大规模协作，因为市场交易成本太高。AI让交易成本急剧下降，公司的边界在模糊。

第五章——此刻——揭示了一个存在了一百五十年的前提：通过教育和努力可以实现阶层上升，这个叙事是社会稳定的基石。AI不需要消灭中产收入，只需要消灭这个叙事的可信度。

五个前提，五块构成工业文明地基的石板。它们不是在同一天碎裂的——有的已经出现明显的裂痕，有的还在承重。但它们的共同特征是：每一块都建立在”认知是稀缺的”这个更底层的假设之上。当认知不再稀缺，五块石板同时松动。

读者读到这里，应该有一种感觉：原来我们的世界不是天然如此。我们习以为常的制度——工作、教育、公司、中产生活方式——不是自然法则的产物，而是一组精密的社会安排。这些安排有历史起点，有运行前提，也有到期条件。而AI正在触发这些到期条件。

这是第一部分的任务：揭示。不是告诉读者”天要塌了”（那是焦虑贩卖），而是让读者看见地基的构造、裂缝的位置、松动的方向。

那新的地基是什么？

新的地基花了约十五年才开始成型，过程远比任何人预想的更混乱、更痛苦、也更充满意外。那是第二部分的故事。

但在进入第二部分之前，值得退后一步，看清这五章反复出现的同一个形状。从教育的筛选功能到公司的层级结构，再到中产阶级的身份叙事——被松动的始终是中间层。中间层的教师、中间层的管理者、中间层的专业人士、中间层的收入群体。这不是五个独立的故事碰巧都涉及”中间”，而是同一个底层力量在不同制度领域的投影：当机器能够完成人类中等水平的认知工作，所有依赖这种能力分布来维持运转的制度安排——考试、科层、职业阶梯、阶层叙事——就同时失去了结构性地基。

从下一章开始，我们不再逐一清点裂缝。问题变成：裂缝之下，地基在怎样重组？

---

**本章关键判断与置信度**

| 判断 | 置信度 | 说明 |
| --- | --- | --- |
| 中产阶级本质上是一个叙事而非经济分类 | A | 文明史分析支持，框架清晰 |
| AI同时动摇中产叙事的三根支柱 | A | 每根的断裂机制独立可验证 |
| 叙事失效是深层社会事件而非经济事件 | A | 历史反复验证 |
| 文化资本在贬值、社会资本在升值 | B | 方向确定但权重变化速度不确定 |
| 不同国家的中产危机形态显著不同 | A | 结构性差异明确 |
| 新的不平等将表现为”认知基础设施控制权”的不对称 | B | 逻辑清晰但形式和程度待验证 |
| 中产叙事转型几乎不可逆 | B | 历史先例支持但缺乏AI时代的直接验证 |

**未解之问**

一、如果中产叙事失效不可逆，什么叙事可以替代它？“通过教育和努力实现阶层上升”这个版本失效了，但社会仍然需要一个让大多数人接受现有秩序的叙事。候选方案可能包括”通过适应和连接实现安全”或”通过公共参与实现尊严”——但这些新叙事能否获得足够的社会认同？

二、中产叙事失效后的社会能量会被引导向制度改革还是社会极化？这个问题的答案可能不取决于经济条件，而取决于制度弹性和治理能力——这两者恰恰是最难预测的变量。

三、赫胥黎式的”舒适的不平等”是否比奥威尔式的”压迫的不平等”更难被意识到？如果AI让每个人的绝对生活水平提高，但资源分布极度不均——人们会质疑一个让自己过得更好的系统吗？

中产阶级的旧地图上，这五条路全标着”此路可通”。现在每一条都在重画。

# 第六章 认知基础设施——当判断力变成API

> 真正的断裂不是机器变聪明了，而是聪明变便宜了。当一项能力的边际成本趋近于零，它就不再是能力——它是空气。

前五章拆解了五块地基石板。本章追问一个更底层的问题：支撑这些石板的东西——认知本身——正在变成什么？

## 一、成本坠落的因果链

理解AI对文明的冲击，不能从能力入手，要从成本入手。能力决定”能做什么”，成本决定”谁能用”。后者才是改变社会结构的变量。

用因果机器的思维来拆解这条链：

**第一环：推理成本的幂律下降。** 2023年3月GPT-4发布时，一次复杂法律分析的API成本约30-50元人民币。2024年底降到5-8元。2025年底降到不足1元。2026年底降到几毛钱。三年，两个数量级。这条曲线的驱动力是多重的——硬件迭代（H100→B200→专用推理芯片）、算法优化（混合专家架构、推测解码、蒸馏技术）、开源竞争（DeepSeek-V3、Qwen 2.5将闭源模型的利润空间压到极薄）、规模效应（用户越多，单位成本越低）。四个力同方向作用，曲线的斜率不是偶然的。

作为参照：1927年首次跨大西洋电话，三分钟收费75美元，到2010年Skype上免费。八十多年，三到四个数量级。AI推理成本在三年内就走完了通信成本八十多年的路程。原因很简单——AI的成本下降不需要在每条海底铺电缆，它只需要在已有的数据中心里优化软件。

**第二环：需求的非线性激活。** 经济学有一个被严重低估的概念：价格弹性的非线性。很多需求不是”便宜一点就多买一点”的线性关系，而是存在一个价格阈值，低于这个阈值需求呈爆发式增长。中国四千多万中小企业中的绝大多数从未获得过合格的法律服务——不是因为不需要，而是因为一个资深商事律师的咨询费（2万起步）超过了它们的支付意愿。当同等质量的法律分析降到几块钱，需求不是增长几倍，而是增长几十倍、几百倍。

这是一个”相变”（phase transition）：水在99°C还是水，100°C变成蒸汽。中间只差一度，状态完全不同。认知服务的价格从”只有大企业用得起”降到”任何人都用得起”，中间可能只差一个数量级的成本——但社会形态的变化是相变级别的。

**第三环：行业渗透与定价体系重塑。** 需求激活之后，连锁反应开始。当数百万中小企业可以用几块钱获得合格的法律分析，传统律所的定价体系就面临压力。受冲击的不是高端业务（复杂并购、跨境交易、重大诉讼），而是占整个法律服务市场收入60%以上的中低端标准化业务。基础财务、初级医疗诊断、建筑方案设计、教案编写、翻译、基础数据分析，所有可标准化的认知服务，其定价的锚点从”专业人士的时间成本”转移到”API调用的边际成本”。

**第四环：组织结构的被迫调整。** 当外部认知服务的价格低于内部人力成本时，企业面临一个纯粹的经济学选择：缩减内部认知劳动力，改用外部API。这个选择不需要任何人有恶意——它是成本曲线的必然推论。一家12人的法务团队每年人力成本约500–800万，一套AI法务系统的年度成本不到50万，做的事情覆盖了前者70%的工作量。企业不是在”选择AI还是选择人”，它是在回应一个两个数量级的成本差异。

收束这条因果链：推理成本幂律下降 → 潜在需求的非线性激活 → 标准化认知服务的定价锚点转移 → 企业内部认知劳动力的结构性缩减 → 建立在认知稀缺上的收入体系开始松动。

每一环都是前一环的必然推论。整条链没有任何一环需要”AI变得更强”作为前提——它只需要已有的AI变得更便宜、更多人用起来。

---

## 二、扩散本身就是革命

2027年的AI讨论被两种声音主导：乐观派预言AGI即将到来、奇点将至；悲观派指出Scaling Law正在撞墙、AI泡沫即将破裂。两方都犯了同一个错误——它们都把注意力放在”前沿模型的能力边界”上。

正确的问题不是”AI还能变多强”，而是”现有AI渗透到所有行业需要多久”。

这个区分为什么关键？因为2027年已有的AI能力，如果充分扩散到所有行业、所有场景、所有地区，就已经足以引发一场深刻的社会变革。前沿是否继续推进，是一个影响变革深度的变量，但不是变革是否发生的前提。

GPT–5比GPT–4强多少，远没有GPT–4级别的能力从硅谷实验室扩散到金华小五金厂需要多久来得重要。前者是渐进改良，后者是结构性改变。

**基层医疗。** 中国约有一百万名基层医生（乡村医生和社区全科医生），面对超过六亿农村居民。2026年的AI辅助诊断系统在常见病上的准确率已达三甲主治医师水平。部署成本？一台平板加4G连接，不到五千元。但按当前推广速度，到2027年底实际使用AI辅助诊断的基层机构可能仍是少数。瓶颈：卫生行政系统未将AI诊断纳入医保流程；村医不会用智能设备；患者不信任”手机看病”；AI诊断出错时的法律责任无人承担。

**建筑设计。** AI已经可以在一小时内完成过去一个建筑师团队两周的方案设计工作。但建筑行业有极其严格的资质审查制度，AI生成的图纸在格式和标注方式上与传统做法不同，审图机构不知道该怎么审。现行的审图规范完全是为人类设计师写的，从立项研究到发布新标准通常需要数年。在规范落地之前，AI在建筑设计领域很可能长期停留在”找灵感”的辅助角色上。

**中小企业财务。** 四千多万中小微企业中至少一半没有专职财务。AI记账系统2026年已能自动识别发票、分类记账、生成报表，月费不到两百元。但如果制度和认知瓶颈不解决，到2029年大多数企业可能仍未使用——老板不知道有这个东西，知道了不会用，会用了不放心，放心了发现银行和税务局还要纸质材料。

三个案例，同一个模式：**技术就绪与社会消化之间存在一条巨大的鸿沟——不是技术鸿沟，而是上面那些具体得令人沮丧的制度和习惯鸿沟。**

但——这是关键的但——这条鸿沟正在被跨越。方向不可逆，因为驱动力太强。每一个被成本下降激活的需求，就是一滴渗入旧制度毛细血管的水。单独一滴什么也改变不了。当所有行业、所有环节同时渗入时，整个地基开始松动。

这就是扩散的力量。它不依赖AI继续变强。它只需要已有的AI被更多人用起来。2026年北京街头，一个外卖骑手在短视频里看到”机器人要抢你的工作”时嗤之以鼻：送外卖这活儿，爬六楼没电梯的老小区，机器人上得去吗？他不知道平台算法已经在悄悄把他的配送范围和一台无人车的配送范围做A/B测试了。扩散就是这样工作的：被渗透的人往往最后才知道。

---

## 三、Scaling Law的约束——真实但被高估

诚实的分析必须面对约束。2026年围绕AI能力的争论，很大程度上是关于Scaling Law是否”撞墙”的争论。

**约束确实存在。**

第一，数据墙。大语言模型的能力与训练数据量之间存在幂律关系，但高质量文本数据的总量有上限。互联网上有多少真正有价值的文本？把维基百科、学术论文、高质量书籍、专业文档全部加起来，大约在10–15万亿token的量级。合成数据可以补充，但效用存在衰减——用模型的输出训练下一代模型，信噪比逐代递减。

第二，能源墙。训练一个前沿大模型在2026年需要的电力相当于一个小城市一年的用电量。全球数据中心的电力消耗在2025年已占全球发电量约3.5%，以每年15–20%的速度增长。按此趋势，到2030年可能达到7–10%。能源不是无限的，尤其是碳排放约束正在收紧。

第三，Agent可靠性的乘法约束。单步调用的错误率在2026年已低于3%。但多步任务（需要AI自主规划、分解、执行、纠错）的成功率受制于链式乘法：每步5%出错率，十步串联后整体成功率只有

60%。这是一个数学约束，不会因为模型变大而简单消失。

不过，约束从来不是单向的——它同时催生了绕过自身的策略。2025年前后，一个被称为”推理时计算”（test-time compute）的趋势悄然改变了能力增长的路径。此前的Scaling Law几乎完全关注训练阶段——更大的模型、更多的数据、更多的GPU。但研究者发现，在推理阶段投入更多计算同样能显著提升表现：让模型进行链式思考（chain-of-thought）、在多个推理路径中搜索最优解、反复验证自身输出。OpenAI的o1系列和后续模型正是沿这条路径发展的。这意味着”智能”不再只是训练出来的静态属性，也可以是推理时”想”出来的动态能力。对Scaling Law的讨论因此需要修正：即便训练规模的边际收益在递减，推理时计算开辟了第二条增长曲线——用更多的时间换更好的答案。

与此同时，能力增长暴露了一个令人不安的伴生问题：幻觉（hallucination）。模型规模越大、输出越流畅，编造事实的能力反而越强——而且是以极高的置信度编造。一个小模型胡说八道，用户一眼就能识别；一个前沿大模型用完美的逻辑链条推导出一个不存在的法律条款，连专业律师都需要查证才能发觉。2026年的研究反复证实，幻觉不随模型规模自动消失，它是生成式架构的结构性特征，而非简单的训练不足。这是”能力增长不等于可靠性增长”的关键例证——模型变强了，但它犯错的方式也变得更隐蔽、更危险。对于那些将AI视为万能工具的乐观叙事，幻觉是一剂必要的清醒剂。

**但大多数人高估了约束的重要性。**

核心洞察在于：人们习惯于关注”边界在哪里”，而忽略了”边界之内已经有多大空间”。

Scaling Law的边际递减意味着”从95分到99分”变得越来越难。但对于绝大多数现实应用场景来说，85分就够了——而2026年的模型在大多数标准化任务上已经达到了85-90分的水平。一个做到85分的法律分析对于金华的五金厂陈老板来说，和99分的分析没有实质区别。他不需要最好的——他需要”足够好且用得起的”。

换言之：**约束限制的是AI能力的天花板，但改变社会的不是天花板的高度，而是地板的覆盖面。**天花板从95分提到99分，影响的是前沿研究；地板从0%覆盖到80%覆盖，影响的是四千万中小企业的经营方式。前者是科学问题，后者是文明问题。

Agent可靠性的约束也类似。2026年的Agent在十步以上的复杂任务中确实不够可靠。但人类经济活动中，能被分解为”三步以内的标准化认知操作”的工作占比超过50%。仅仅自动化这50%，就已经是一场劳动市场的地震。

所以正确的判断是：Scaling Law的约束是真实的，它意味着AGI不会在明天到来。但它不能用来否定AI在现有能力水平上的扩散效应——因为扩散本身就足以改变一切。

---

## 四、烧钱的游戏——AI商业化的残酷算术

能力在增长，约束是真实的，扩散正在发生。但有一个问题被技术讨论系统性地回避了：谁在为这一切买单？

2046年回望，AI商业化的早期图景像极了一场豪赌。训练一个前沿大模型的成本从2023年的数千万美元飙升到2025年的数亿美元，再到2026年部分项目突破十亿美元门槛。但变现路径在2026年仍然模糊得令人不安。大多数AI公司的商业模式本质上是”先烧钱做能力，再找场景变现”——与二十年前互联网时代的”先获客再变现”逻辑相似，但资本密度高出整整一个数量级。一个互联网创业公司烧掉一亿美元已经算巨额融资；一个前沿AI实验室烧掉一亿美元，可能只够训练一次模型。

这种资本密度天然地将竞争格局推向两极。科技巨头——微软、谷歌、亚马逊、Meta，以及中国的阿里、字节、腾讯——有算力、有数据、有分发渠道，也有从既有业务输血的能力。但大公司的决策链条长、创新受制于既有业务的利润结构（一个AI产品如果可能蚕食自家搜索广告收入，内部阻力可想而知）。创业公司有速度和创意，OpenAI从零到千亿估值只用了不到三年，但它在算力和数据上天然受制于巨头——训练集群要租微软的Azure，用户触达要靠苹果和三星的设备预装。这种”创新在小公司、基础设施在大公司”的张力，在后续章节讨论的数据圈地运动中会进一步激化。

最终活下来的商业模式呈现出鲜明的分层。底层是API按量付费——认知能力像电力一样被计量销售，成为真正的基础设施生意，利润薄但规模巨大。中间层是垂直场景应用——医疗AI、法律AI、教育AI比通用助手更早实现正向现金流，因为它们解决的是具体的、可定价的、用户愿意付费的问题（“帮我看这张CT”比”陪我聊天”值钱得多）。顶层是企业级深度集成——将AI嵌入客户的核心业务流程，按效果收费而非按调用收费。而那些试图靠”通用智能助手”直接面向消费者变现的公司，大多没有熬过2030年。原因很简单：消费者愿意为具体结果付费，不愿意为”智能”本身付费——就像没有人会为家里的电表本身付钱一样。

这条商业化的残酷筛选，最终与前文讨论的成本坠落形成了一个闭环：烧钱训练模型→成本下降→需求激活→变现路径清晰→更多资本投入。循环一旦转起来就不可逆。但在2026年，这个循环还远未闭合——大多数AI公司仍在烧钱阶段，靠的是风险资本的信仰和科技巨头的战略性亏损。

---

## 五、从私有发电到公用电网——五十年的教训

从爱迪生1882年点亮珍珠街的第一座中央电站，到全美电气化，用了整整五十年。瓶颈不是技术——1890年交流电已解决远距离输电，1900年美国已有三千多座发电站——而是组织和制度。经济史学家保罗·大卫曾用一组数据戳破了一个常见误判：技术就绪不等于社会就绪：电力渗透率1900年就够了，但生产率跃升要等到1920年代。因为早期工厂只是”把蒸汽机换成电机，其余不变”，真正的爆发要等到福特用独立电机驱动每个工位、发明流水线——一种只有电力时代才能存在的组织形式。

**技术的革命性不在于技术本身，而在于它催生的组织创新。** 2027年的大多数企业正在重复这个错误：让员工用ChatGPT辅助写报告，但保留所有审批层级和决策流程。换了动力源，没换组织方式。但电气化的历史也表明：一旦有人示范了新的组织形式，扩散会急剧加快，因为效率优势太大，不采用就会被淘汰。

AI扩散会比电气化快得多。电力需要铜线和变压器，花了五十年；电话需要交换机和海底电缆，三十年；互联网二十年；移动互联网十年。认知基础设施的物理前提（算力中心、4G/5G网络、智能手机）已基本就位，剩下的只是软件、数据和制度适配。

**文明的跃迁取决于新技术被整合进既有制度结构的速度。** AI重塑组织不需要拆掉建筑、重建工厂，它需要改变的是软件流程、决策权限、人员配置——物理约束远小于铺设电网。

---

## 六、谁控制认知基础设施，谁就控制文明运行

如果认知正在变成基础设施，那么一个必然的权力问题浮现：谁拥有这个基础设施？

回到第一性原理。在任何文明阶段，控制核心基础设施的力量就控制着文明的运行方式。农业文明靠土地和灌溉，控制者是地主和王权。工业文明靠工厂、铁路和能源，控制者是资本家和民族国家。那么认知文明的核心基础设施是什么？

三个层次，缺一不可：

**算力的集中度被低估了。** 中世纪的大教堂不是普通建筑——它是信仰的物质化身，建造需举一城之力，耗时数十年，建成后成为地区精神中心。谁控制大教堂，谁就控制信仰的诠释权。AI算力中心扮演着完全对等的角色。一座大型数据中心消耗的电力相当于一个中等城市，需要的资本投入相当于一个小国的年度GDP，运行的模型决定着数以亿计用户的认知辅助质量。2026年，全球前十大AI公司控制了超过70%的公共云AI算力。这十家公司中七家美国、三家中国。欧洲没有。日本没有。印度没有。

**能源正在变成AI的石油。** 全球数据中心2024年耗电约415—460太瓦时，预计到2026年将增长至约1000太瓦时，届时占全球发电量约3.5%。AI推理需求每年翻倍意味着这个比例在加速增长。微软2025年重启三里岛核电站供应AI数据中心。亚马逊签下美国历史上最大的企业级核能购电协议。谷歌投资小型模块化核反应堆。Meta在爱荷华建造北美最大的太阳能农场——全部用于一座数据中心。这些不是公关——当你的核心业务是运行认知基础设施，能源就从”成本项”变成”战略资源”。二十世纪的石油决定谁能驱动工业机器；二十一世纪的电力决定谁能驱动认知机器。

**数据有了主权边界。** 英语互联网数据量占全球网络文本60%以上——英语AI模型天然拥有最丰富的训练素材。更关键的是专业数据：法律AI需要裁判文书和合同模板，医疗AI需要病历和影像，金融AI需要交易记录和监管文件。这些数据大多掌握在机构手中。谁能合法地、大规模地获取这些数据，谁的AI在垂直领域就更有竞争力。和土地一样，数据也有主权归属问题——GDPR、中国《数据安全法》、各国数据本地化要求，本质上都是在划定数据领土的边界。

三层叠加，权力图景清晰：算力是硬件层（认知产能），能源是驱动层（可持续上限），数据是原料层（质量上限）。三者的控制权都高度集中于少数科技巨头和它们所在的国家。

**这种集中是规模效应的必然结果。** AI系统的规模效应——训练成本的门槛效应、数据的网络效应、人才的聚集效应——天然导致赢家通吃。一个国家要从零建立一套完整的认知基础设施（芯片设计→制造→数据中心→模型训练→应用生态），需要的投资以万亿美元计，需要的人才以万人计，需要的时间以十年计。这不是任何中等国家可以独立完成的。

当然，集中化叙事需要一个重要的修正项：开源。Meta从2023年开始开放LLaMA系列模型权重，到2026年LLaMA 4已接近闭源前沿模型的水平。阿里的Qwen、DeepSeek等中国开源模型同样表现出色。开源确实在一定程度上打破了闭源巨头对前沿能力的垄断——中小企业可以在自己的服务器上部署开源模型，研究机构可以在开源基座上微调垂直领域应用，创业公司不必从零训练就能获得接近一流的基础能力。但必须看到硬币的另一面：开源了模型权重，没有开源训练所需的万卡集群。训练LLaMA 4的算力投入以数亿美元计，这不是任何开源社区能众筹的数目。开源降低了应用层的门槛，却没有动摇基础设施层的集中度——就像Linux让所有人都能免费使用操作系统，但并没有让所有人都能制造芯片。

二十世纪的核心权力问题是谁控制生产资料：工厂、矿山、铁路。围绕这个问题，产生了工会运动、社会主义革命、福利国家、劳资谈判。二十世纪的大部分政治史都是在回应这个问题。

二十一世纪的核心权力问题是谁控制认知资料——算力、能源、数据。这个问题的答案正在形成。但电力的历史提供了一个参考：电力发展早期也经历过类似的私人垄断（爱迪生和特斯拉的电力公司一度垄断大片区域），后来各国政府通过公用事业监管、反垄断法和国有化将电力变成了受监管的公共服务。认知基础设施是否也会走上同样的路径——从私人垄断走向公共监管——是后续章节要讨论的问题。

---

## 七、“开关”的含义

让我们收束本章的核心论证。

十九世纪末，每座工厂都需要一位蒸汽工程师来操作锅炉、维护管道、保障安全。这是需要多年经验的高度专业化技能，收入和地位不低。电气化之后，蒸汽工程师消失了。不是因为他们变笨了，而是因为整个问题被取消了。工厂不再使用蒸汽，所以不再需要管理蒸汽的人。工人不需要理解电力原理，他只需要打开墙上的开关。

一项曾经需要多年修炼的稀缺技能，变成了一个开关。

2026年到2046年间发生的事情，是这个过程在认知领域的重演。法律分析、财务审计、医学诊断、建筑设计、投资分析、翻译、编程——一个又一个曾经需要多年专业训练的认知技能，正在变成可以被任何人通过简单界面调用的服务。

这不是让每个人都变成了律师、医生、会计。这是让”成为律师、医生、会计”这件事本身，至少对于大多数标准化需求而言，不再必要。

**认知变成基础设施的真正含义：它不是让人更聪明，它是让”聪明”不再稀缺——因此也不再值钱。**

电力让你不需要理解热力学。自来水让你不需要理解水文学。认知基础设施让你不需要训练大脑——至少不需要为标准化认知任务训练大脑。

但这引出了一个更尖锐的问题。蒸汽工程师的消失没有引发社会危机——因为他们人数不多，新岗位很快填补了空缺。但当消失的不是一种职业而是一整个类别——当所有”标准化认知工作者”同时面临同一种命运时，新岗位的创造速度能不能跟上旧岗位的消失速度，就不再是一个可以轻松回答”历史证明总会有新工作”的问题了。

认知基础设施化的物理前提已基本就位——算力中心、网络连接、终端设备都在那里。剩下的只是软件、数据和制度的适配。这些适配的速度比铺设铜线和变压器快得多。电气化从珍珠街到全美花了五十年。认知基础设施化不会用那么久。

因为没有人会关掉墙上的开关，回到蒸汽时代。

而那些围绕认知稀缺建立起来的制度——教育、职业、薪酬、阶层——将首先感受到冲击。下一章，我们来看这个冲击最直接的战场：工作。

---

*诊断到这里只完成了一半。接下来四章，我们去看看冲击落到真实世界里是什么面貌。*

# 间奏零：醒来

* * *

闹钟没响。

2026年三月的一个周二早晨，赵敏的手机在六点十五分震了一下，不是闹钟，是一条推送：“您关注的岗位’行政主管’在过去7天新增职位数下降42%，建议拓宽搜索范围。”她盯着这条消息看了三秒，然后滑掉了。

她三十六岁，住在杭州城西一个老小区，二楼，朝北，采光不太好。月供七千八。她老公在一家物流公司做调度，她自己去年十月从一家做企业培训的小公司离职——说是离职，其实是公司先砍掉了她所在的部门，留了两个月让她”内部转岗”，转了一圈没有合适的位置，于是体面地走了。

五个月了。

* * *

六点半，她起来给女儿热牛奶。女儿上小学四年级，学校八点到，她七点二十送出门刚好。厨房里的智能音箱在她经过时自动播报天气，她已经习惯了这个声音，习惯到完全不在听。音箱还会报一条新闻摘要，今天是关于某个城市试点”AI政务助手”的消息。她”嗯”了一声，把牛奶倒进杯子。

女儿在房间里喊：“妈妈，昨天的数学作业我用那个APP拍了一下，它说我第三题做错了，但是我觉得它说的方法我没学过。”

“那就按老师教的方法做。”

“可是APP的方法更快啊。”

她没接话。她不知道该怎么回答这种问题。四年级的数学她还能辅导，但女儿越来越多地用那个拍照解题的APP，有时候她感觉自己在和一个免费的、永远不会不耐烦的家教竞争。这种感觉很轻，轻到她不会跟任何人说，但它在那里。

* * *

七点半，送完女儿，她走进小区门口的咖啡店。不是星巴克，是一家夫妻店，咖啡一般，但离家近，老板娘认识她。

“还是美式？”

“嗯。”

她坐在靠窗的位置，打开笔记本电脑。这是她每天的”工作”——虽然她没有工作。早上两个小时用来投简历、看招聘信息、偶尔刷一下行业资讯。下午如果没有面试，就研究点东西——最近在看线上课程，学”AI辅助办公”。课程是免费的，平台是一个叫不上名字的创业公司做的，界面粗糙但内容还行。教你怎么用AI写会议纪要、做PPT、整理数据。

她学得很认真。不是因为觉得这些技能能帮她找到工作，而是因为学习本身提供了一种结构感，有个事情在做，时间就不那么难熬。

前同事小周上周在微信上说，找到了一份新工作：在一家电商公司做”AI内容运营”，月薪比之前低了三千，但”至少有着落了”。赵敏回了一个”恭喜”的表情包，然后看着窗外发了一会儿呆。

小周比她小六岁。三十岁，还能弯一次腰。三十六岁呢？降薪三千意味着月供要从公积金里补，女儿还有四年小升初，课外班不能全停。

她没有把这些算给任何人听。她妈妈会说”要不回老家吧”，她老公会说”慢慢找，不着急”。两种回答都不是她需要的。她需要的是一个信号——一个告诉她”你的经验还有人要”的信号。但招聘网站上的信号越来越模糊：有些岗位写着”需熟练使用AI工具”，有些岗位的职责描述她读了三遍也不确定到底要做什么，还有些岗位看起来完全正常，投了简历之后石沉大海。

她想起一个细节。上个月面试了一家做在线教育的公司，面试官是个比她年轻十岁的男生，问她：“你对AI协作有什么理解？”她说了一些她在课程里学到的东西。面试官点点头，表情礼貌但空洞，那种你知道对方已经在心里把你归入”下一位”的表情。

面试结束后她在公司楼下站了一会儿。那栋楼的大厅里有一块屏幕，在轮播公司的宣传片：“AI赋能教育，让每个孩子都能获得最好的老师。”她想到了女儿和那个拍照解题APP，想到了自己和那个AI办公课程，想到了小周和她的”AI内容运营”。想着想着就没再想了。电梯门开了。

*　　*　　*

中午，她在家里随便煮了碗面。手机上跳出一条新闻：某市政府宣布将在公共服务窗口引入AI预审系统，“预计可将办事效率提升三倍”。她看完之后在心里默默翻译了一下：窗口可能要裁人了。

她并不愤怒。说实话，她甚至觉得这个技术挺好的——她自己去办社保转移的时候排了两个小时的队。如果AI能让这个过程快一点，当然好。但她同时也意识到一件事：她站在一个奇怪的位置上。作为市民，她欢迎效率提升；作为一个正在找工作的人，每一个”效率提升”的新闻都在缩小她的生存空间。这两个身份住在同一个人体内，却指向相反的方向。

下午三点，她接到了一个电话。是一家做餐饮供应链的公司的HR，说看了她的简历，想聊聊。薪资比她上一份工作低四千，职位是”运营专员”而不是”主管”。她犹豫了两秒钟，说：“好的，什么时候方便？”

*　　*　　*

傍晚，她去学校接女儿。校门口站着一群家长，大多在看手机。她旁边一个爸爸在用手机跟AI对话，似乎在让它帮忙写什么东西——她瞥了一眼，是一份工作周报。那个爸爸全神贯注地修改着AI生成的文字，手指在屏幕上快速滑动，脸上的表情像极了她女儿用拍照APP检查作业时的样子。

女儿跑出来了，书包一颠一颠的。“妈妈，今天美术课老师让我们画’未来的城市’，我画了一个会飞的房子！”

“是吗？好看吗？”

“好看！老师说我画得很有想象力。”

她笑了。这一天里第一个真正让她笑的事情。

回家路上，女儿牵着她的手叽叽喳喳地说着学校的事。经过小区门口那家咖啡店时，老板娘正在收拾桌椅准备关门。老板娘朝她挥了挥手，她也挥了挥。

天色暗下来了。楼道里的声控灯一层一层亮起来，她们的脚步声在楼梯间回响。她忽然想到一个问题，但没有说出口：明天的面试，要不要把那个AI办公课程的结业证书打印出来带着？

她不知道那张证书值多少钱。

她不知道的事情越来越多了。而这种”不知道”和一年前的”不知道”不一样。一年前她不知道的是具体的事——下一份工作在哪里、薪资多少、通勤远不远。现在她不知道的是更大的事——她花了十五年积攒的工作经验，在一个连”行政主管”这四个字都越来越少出现在招聘网站上的世界里，还算不算数。

她把这个问题折叠起来，推进了心底的某个角落。钥匙转动锁芯的声音很清脆。门开了，屋子里有一股中午煮面留下的淡淡酱油味。

女儿冲进去打开了灯。

赵敏站在门口，忽然有一种很久没有过的感觉：她正站在什么东西的边缘上。不是悬崖——没那么戏剧化。更像是一条路的尽头，前面是一片还没有路的地方。

她深吸一口气，换了鞋，走进去了。

*　　*　　*

# 间奏一：最后的日常

*　　*　　*

那天傍晚没有任何特别之处。

2026年四月的某个周三，北京西二旗的天空被一层薄薄的灰色覆盖，不算阴天，也谈不上晴朗。就是那种你抬头看一眼、什么也记不住的天色。如果有人在那个黄昏按下快门，拍下整条后厂村路的横截面，他会看到一幅完全普通的画面——普通到不值得拍。

但后来我们知道，那些最普通的画面，恰恰是最后的画面。

*　　*　　*

五点四十分，一个穿灰色冲锋衣的年轻人从一栋写字楼的旋转门走出来。他在一家中型互联网公司做后端开发，入职三年，刚评上中级。他的工牌挂在胸前晃荡，上面印着一个他自己都觉得有点傻的工号。他掏出手机看了一眼——没有新消息。妻子还没下班，孩子在岳母家。他站在路边犹豫了几秒，不知道是去便利店买个饭团凑合，还是走远一点去那家兰州拉面。

他最终选了拉面。不是因为饿，是因为不想那么快回到出租屋里对着屏幕。白天对着屏幕写代码，晚上对着屏幕刷短视频，区别只是一个要动脑子一个不用。他最近有点烦。不是那种大的烦恼——房贷还得起，老婆没吵架，孩子还小不费心。是一种模模糊糊的、说不清楚的东西，像耳鸣一样低频地嗡嗡响。

他说不出那是什么。

如果非要他描述，大概是这样一种感觉：每天写的代码越来越像填表。需求文档发下来，拆成任务，写接口，调通，提交，等测试反馈。三年前刚入行时那种"我在造东西"的兴奋感已经消退了。他不讨厌这份工作，但他也不再期待明天。这种状态没有一个名字，或者说，当时还没有。几年之后人们会试着给它取各种名字，但在那个黄昏，它只是一碗拉面解决不了的闷。

他不知道的是，六个月之后，他的组长会在周会上演示一个新工具。那个工具能在三十秒内生成他一天才能写完的接口代码。组长的表情是兴奋的。他的表情什么也没有。

*　　*　　*

六点十分，一个穿校服的女孩从地铁站涌出的人流中挤出来。高二，海淀某中学，刚上完晚自习前的最后一节课——数学。她的书包很重，里面有五本教辅和一个装满试卷的文件夹。她妈妈在微信上问她晚饭吃什么，她回了一个字：“随。”

她今天考了一次月考模拟，数学一百三十七分。不算差，但她妈妈肯定会说”再努力一下就能上一四零”。她已经能预见今晚的对话：妈妈会先夸她”还不错”，然后话锋一转说隔壁张阿姨的女儿这次考了一四三，最后落到”你要是能把最后那道大题的第二问拿下就好了”。她不想听。不是叛逆，是疲惫。

她有时候会想一个问题：我现在每天花十四个小时学习，是为了什么？标准答案她知道——考一个好大学，找一份好工作，过上好生活。但”好生活”长什么样？她爸爸是好大学毕业的，现在在一家国企当中层，每天最大的乐趣是下班后喝两口酒看看球赛。她妈妈也是好大学毕业的，现在全部精力花在盯她学习上。这就是”好生活”吗？

她摇了摇头，把这个念头甩掉了。想这些没用。明天还有英语听力要练。

她不知道的是，三年后她参加的那场高考，将是最后一届没有AI监考、没有AI阅卷、没有AI填报志愿的高考。而她花了一万个小时练出来的那些解题技巧，将在毕业后的第二年变得和算盘珠一样。

*　　*　　*

六点半，一个骑手把电动车停在一家写字楼的地下车库入口。他今天跑了四十七单，还差三单就能拿到平台的”冲刺奖励”——多跑三单，额外奖八块钱。他在心里算了一下时间，如果接下来半小时能接到顺路的单子，八点之前应该能完成。

他的手指在接单键上方悬着。屏幕上跳出一个新订单：三公里，预计十五分钟送达，报酬五块二。他按下了”接受”。

他是安徽人，来北京六年了。之前在工厂打过工、在工地干过活、开过一阵子出租车。骑手这份工是他做得最久的——将近三年。不是因为喜欢，是因为自由。没有工头，没有固定排班，想跑就跑，不想跑就歇。当然，“不想跑就歇”是理论上的。实际上他每天跑十到十二个小时，因为不跑就没钱。但至少，在红灯等待的三十秒里，他觉得自己是在”做自己的生意”，而不是给谁打工。

这种错觉——或者叫信念——支撑着他。

他对AI没什么概念。偶尔在短视频里看到有人说"机器人要抢你的工作了"，他嗤之以鼻：送外卖这活儿机器人怎么送？爬六楼没电梯的老小区，机器人上得去吗？他觉得那些专家在胡说。他不知道的事情太多了——他不知道无人配送车已经在顺义跑了两年，不知道几家大厂正在测试楼宇配送机器人的爬梯能力，不知道平台算法已经在悄悄把他的配送范围和一台无人车的配送范围做A/B测试。

他只知道今天还差三单。

* * *

七点一刻，一个女人坐在朝阳区一间两居室的餐桌前，对面是她七十三岁的母亲。母亲在吃她带回来的盒饭——宫保鸡丁套餐，米饭要软一点。她自己没什么胃口，一边扒拉着菜一边看手机。她是一家出版社的编辑，做了十五年，专门负责社科类图书。今天下午开了一个选题会，主编说了一句让她心里咯噔了一下的话："以后翻译类的书可以考虑用AI初译，编辑负责润色就行。"

主编说这话的语气是轻描淡写的，像在说"以后打印的事让实习生去做"。但她听出了潜台词：翻译类图书占她手头工作量的三分之一。如果翻译环节被压缩，她的工作量——以及她在这家出版社的不可替代性——就要打折扣了。

她妈妈在对面嚼着鸡丁，突然说了一句："今天打了半天电话都打不通社保中心，全是机器人接的，按了半天键也没转到人工。"

她"嗯"了一声，没接话。

两代人各自对着自己的困境沉默。一个是被机器挡在门外，一个是被机器挤进角落。她们没有意识到自己正在经历同一件事的两个面——数字化浪潮的前哨战。她们只是各自烦躁，各自无奈，各自在一顿不怎么好吃的盒饭里消磨掉一个普通工作日的尾巴。

* * *

八点，这座城市的灯光亮起来了。写字楼里还有很多人没走——加班是这里的默认设置，不加班才需要解释。地铁里的人流开始稀疏，取而代之的是外卖骑手、代驾司机、以及那些在便利店门口蹲着抽烟的快递员。他们是白天混在人流里看不见、入夜之后才从街角冒出来的那批人。

如果你从足够高的地方俯瞰这座城市，不是卫星的高度，大约是一架无人机的高度，五百米左右，你会看到一个运转良好的系统。车流按照红绿灯的节奏脉动，人流按照地铁时刻表潮汐般涨落，灯光按

照统一的时间自动亮灭。一切都在运行。一切都在按照二十年前设计好的轨道运行。

没有人在那个黄昏感到恐惧。没有人抬头看天空，觉得什么东西即将改变。恐惧是后来的事——变化已经发生之后的事。

那碗拉面、那个打不通的电话——

只是站在里面的人，看不见轮廓。他们只看得见今天的晚饭、明天的考试、下一单的五块二、以及一个打不通人工客服的下午。他们活在旧秩序的最后一个黄昏里，像活在所有普通的黄昏里一样——没什么特别的。

真正特别的事情从来不会提前通知。它们只是在某一天到来，然后你回头看，才发现那个普通的周三傍晚，原来是分水岭。

*　　*　　*

# 第七章 工作消亡的五个阶段

> 流水线工人至少知道自己被替代了。知识工作者最大的风险是：他的工作已经被替代了，但他的工资单还没反映出来。

---

前六章是诊断。从这里开始是预后——工作的消亡不会一夜发生，它有五个阶段。

## 一、不是一个事件，是一个过程

上一章论证了认知正在变成基础设施。最直接的后果是什么？劳动替代。

但”AI取代人类工作”这个说法本身就是一种认知上的偷懒。它把一个持续二十年、在不同阶段拆除不同类型工作的复杂过程，压缩成了一个新闻标题。实际发生的事情远比这更微妙、更缓慢、因此也更难抵抗。

系统的崩溃不是因为单一冲击，而是因为方向一致的小压力积累到临界点。工作的消亡就是这样一个过程。它不像工厂关门那样发出声响，更像自动补全逐渐接管你的打字：一开始你只是用它补完几个单词，某天你发现整段话都是它写的，而你只是按了几下Tab。

从2026年起的大约二十年间，劳动替代经历了五个可辨识的阶段。五个阶段的驱动力和社会反应差异很大，笼统讨论会丢失关键信息。

但在展开之前，需要先明确一个历史参照。

---

## 二、英国手工织工的七十年

工业革命时期英国手工织工的命运，是理解当下最有价值的历史类比——不是因为它展示了技术替代的速度，而是因为它展示了**替代的缓慢程度**。

1760年代，飞梭和珍妮纺纱机开始在英格兰北部的工厂中推广。手工织工的黄金时代实际上在此之后才到来——因为机械纺纱极大增加了纱线供给，手工织工的需求反而短暂上升，工资在1790年代达到峰值。然后，动力织机开始扩散。从1800年到1830年，手工织工的工资下降了约60%。到1840年代，这个职业基本消亡。

整个过程从开始到结束，跨越了七十年。不是一天失业，是三代人的苦难。

**第一层：替代的第一个效应往往是短暂繁荣，而非立刻衰落。** 新技术（机械纺纱）先是增加了对互补劳动（手工织布）的需求。这创造了一个危险的认知陷阱——织工们在技术革命初期反而挣得更多，以为自己是受益者。今天的某些知识工作者正处于类似位置：AI辅助让他们短期产出更高，收入未降反升。他们没有意识到，这是动力织机扩散之前的1790年代。

**第二层：真正的替代不是”一夜之间失业”，而是”十年之间贬值”。** 手工织工不是在某一天被赶出工厂的。他们的收入年复一年地下降，每一年的降幅都不够大到让人掀桌子——但每一年加起来，二十年后生活已经不可持续。这种替代模式像GPS导航替代老司机的路感：你没有在某一天突然不认路了，但五年后关掉导航，你发现自己真的不认路了。每一年的降幅都不值得辞职，但每一年加起来，你已经被重新定价了。这种渐进贬值模式，在任何大规模劳动力转型中都反复出现。

**第三层：新工作确实出现了，但不是给同一批人的。** 纺织工厂创造了大量新岗位——但这些岗位要求的是不同的技能（操作机器而非手工技艺）、不同的地点（工厂集中在城市而非乡村）、不同的年龄（工厂偏好年轻工人和童工）。对于一个四十岁的乡村织工来说，“新工作”是一个统计学事实，不是个人现实。

但织工的故事不是唯一的结局。同一时期的丹麦提供了一个不同的叙事。十九世纪末，廉价的美洲谷物涌入欧洲市场，丹麦小农面临灭顶之灾。和英国织工不同的是，丹麦农民没有各自为战。他们自发组建了合作社——从奶制品加工开始，逐步扩展到教育、信贷和出口。到十九世纪末，丹麦已经从一个谷物出口国转型为高附加值乳制品出口国，农民的收入不降反升。关键不在于技术本身是否仁慈，而在于社会是否有能力组织起集体性的适应。丹麦农民的识字率高、社区纽带紧密、地方自治传统深厚，这些”非经济资本”在危机中转化为了组织能力。这个案例提醒我们：技术替代不必然导向苦难，前提是社会结构能提供缓冲和重组的空间。

带着这个更完整的历史框架，让我们来看AI时代的五个阶段。

---

## 三、第一阶段：标准化知识工作的静默压缩（2026–2029前后）

**被拆除的对象：** 所有可以被描述为”根据一组输入条件，按照既定规范，生产标准化输出”的工作。文案写作、翻译、初级数据分析、基础法律文书、财务报表编制、客服话术、教案编写、市场调研报告。

**替代机制：** 不是”AI干了你的活所以你被开除”这种粗暴替代。实际过程远更温和、更隐蔽——

第一步，引入AI工具作为”辅助”。员工产出提升，公司满意，员工也觉得自己效率提高了。第二步，工作流程重构。从”人写AI辅助”变成”AI写人审核”。核心动作的主语悄悄换了。第三步，headcount自然缩减——不续约、不补招、内部调岗。没有裁员公告，只有人力成本报表上持续下降的

数字。第四步，留下来的人工资冻结或微降——因为他们的工作已经从”创造”变成了”审核”，而审核的市场定价天然低于创造。

整个过程在18–24个月内完成。10人团队变成3–4人，产出不降反升。没有人能指着某一天说”这就是AI取代我的那一刻”。2026年的一个普通周三傍晚，北京西二旗某栋写字楼里，一个入职三年的后端程序员还在为每天写的代码”越来越像填表”而感到说不清的闷。六个月后他的组长在周会上演示了一个新工具，三十秒生成他一天的工作量。组长的表情是兴奋的。他的表情什么也没有。

**数据画像（趋势推演）：** 2026–2029年间，知识密集型行业的就业人数可能下降约五分之一甚至更多，同期产出贡献反而增长。劳动生产率的跃升不是因为人变厉害了，而是因为人变少了，机器补上了空缺。

**制度反应：近乎空白。** 没有工会出来喊话，被压缩的主要是白领知识工作者，本就缺乏组织化保护。也没有大规模抗议，因为没有大规模失业，只有大规模的”价值压缩”。你还在上班，但你的工作已经不值以前那个价了。城镇新增就业可能首次低于预期目标值，官方解释大概率是”结构性调整”，一个足够模糊、足够正确、也足够无用的词。

**这个阶段的相变点不在技术，而在工作流程的主语切换：**从”人做AI辅助”到”AI做人审核”，只差一步，但劳动的性质从创造变成了监督。

---

## 四、第二阶段：复杂知识工作的角色重构（2029前后–2030年代初）

**被拆除的对象：** 需要判断力、经验和上下文理解的知识工作——管理咨询、产品研发、市场策略、投资分析、医疗诊断中的复杂病例、法律中的策略规划。

这些工作在第一阶段看似安全。但它们面临的不是直接替代，而是一种更微妙的冲击——**地基被掏空**。

一个咨询公司的高级合伙人之所以能在客户面前用二十分钟做出判断，是因为身后有一个金字塔——120个分析师花三周产出的数据、洞察和对标分析。当这个金字塔被压缩成一组Agent流水线，他的判断力没有消失，但他的位置变了：不再站在金字塔顶端，而是站在流水线末端。金字塔顶端能收咨询费，流水线末端只能收审核费。

**替代机制：不是替代人，是消灭组织层级。**

因果链很清晰：AI压缩了底层执行人员，中层管理因此失去管理对象，组织从金字塔被迫转向扁平网络。原本通过”管理人”获得溢价的岗位失去了定价基础，高级知识工作者的收入开始向”个体贡献者”回归。

具体而言：当一个咨询项目不再需要15个分析师，就不需要3个经理来管理分析师，不需要1个总监来管理经理。合伙人的角色从”管理一个知识生产体系”变成”亲自操作一套AI工具”。他的判断力仍然

值钱，但”管理溢价”消失了——而管理溢价在很多高薪知识工作者的收入中占比过半。

**转折点的标志：** 当Agent可靠性跨过”多步复杂任务自主完成”的门槛（大致在2030年代初），全球头部咨询公司将率先进行大规模组织架构调整。传统的”分析师→咨询顾问→项目经理→总监→合伙人”五级金字塔被压缩为三级。中间两层，项目经理和总监，大部分被取消或重新定义。

**制度拖曳：** 技术就绪往往早于组织变革。但大型组织的架构调整需要时间——不是因为技术不够好，而是因为组织有惰性。大型组织的架构调整平均滞后技术变革3–5年。原因有三：现有管理层抵制（他们的权力来自旧结构）；客户习惯（“我付这么多钱，怎么只有一个人对接我？”）；法律和合同约束（劳动合同、竞业协议、期权锁定期）。

但滞后不等于不发生。技术就绪后，头部企业率先调整，随后三到五年内成为行业标配——扩散周期可能只有几年，比电气化时代的二十年快了数倍。

**这个阶段的真正冲击不在于失业人数（被彻底失业的高级知识工作者并不多），而在于收入的实质压缩。** 职称还在，甚至涨了一级。但base降了三到四成，补了一部分期权（期权在组织缩小后价值存疑）。每一步都不够大到让人掀桌子，每一步加起来，五年后实际收入回到了十年前的水平。

---

## 五、第三阶段：物理世界自动化的加速渗透（2030年代中期–2030年代末）

**被拆除的对象：** 前两个阶段集中在”屏幕前的工作”。第三阶段冲击延伸到物理世界——制造业、物流、建筑、农业、零售中的体力和操作性劳动。

**驱动力的变化：** 前两个阶段的驱动力是大语言模型和软件Agent。第三阶段的驱动力叠加了三个新变量：机器人硬件成本的下降（受益于电动车产业链的规模效应）；通用机器人操控能力的突破（大模型的多模态能力从”理解文字”扩展到”理解物理空间”）；5G/6G的低延迟网络使远程操控和协同成为现实。

**相变点：** 当通用机器人的年均成本（采购+维护+能源）低于同岗位人类劳动力的年均成本时，替代的经济理性成立。这个交叉点在不同行业和地区出现的时间不同——大致在2030年代中期首先在发达国家的高人力成本行业（仓储物流、标准化制造）出现，随后数年内扩展到更多领域。

**与前两个阶段的关键差异：** 知识工作的替代是”静默”的，没有大规模裁员公告，没有工厂关门。物理世界的自动化则不同：当一条生产线从100个工人变成5个工人加20台机器人时，95个人确实失去了工作岗位。这种可见的、集中的失业更容易引发社会反应。

**制度拖曳更加显著。** 制造业有工会传统（至少在欧美和部分亚洲国家）。蓝领岗位的消失比白领岗位的”价值压缩”更容易被政治化。当物理自动化的社会冲击变得不可忽视时，欧洲和亚洲多国将出现”

反自动化”立法尝试——对使用机器人的企业征收”机器人税”、要求维持最低人工比例、限制自动化改造的速度。

这些立法的效果有限。原因不难理解：限制自动化等于提高本国制造成本，本国产品因此在国际市场失去竞争力，贸易逆差随之扩大，最终被迫取消限制或接受产业空心化。在全球化的经济体系中，单一国家的”反自动化”政策就像在大坝上堵一个洞——水会从其他地方流过去。

**历史类比：** 这与英国手工织工的命运高度同构。织工们也尝试过抵抗——卢德运动（1811–1816）中，愤怒的织工砸毁了数百台织布机。英国政府的回应是出动军队镇压，并通过了《破坏机器法案》（Frame Breaking Act），将砸机器的行为定为死罪。暴力抵抗失败了，不是因为织工没有正当理由，而是因为技术替代背后的经济力量太强大——使用机器的工厂主利润更高，能生产更便宜的布匹，最终赢得了市场。

这一轮的”新卢德运动”不会采取砸机器的形式——它会采取立法、罢工、选票的形式。但底层逻辑是一样的：当技术替代的经济驱动力足够强时，政治抵抗最多能减速，不能逆转。

---

## 5.5 具身智能：比你想的更远，也比你想的更近

第三阶段涉及物理世界的自动化，但”机器人替代”四个字远不足以描述正在发生的变化。要理解物理世界自动化的真实时间线和深远影响，需要一次更系统的展开。

### 三类瓶颈——区分它们是理解时间线的关键

混淆三类完全不同性质的问题，是2020年代技术预测反复失准的主要原因。

**工程问题**（已知原理，已知路径）：成本、量产和软件集成。2025年一台协作机器人售价约35万元。按照当前国产化进程推演，到2030年前后国产化率可能从40%提升到75%，均价有望压到15万以下。如果这一趋势延续，到2035年前后，一个机器人工位的月成本可能低于东部沿海一个熟练工人的月综合成本。量产方面，按深圳和东莞现有产能扩张速度，到2033年前后有望实现每月数千台规模出货。

**科学问题**（原理尚未完全理解）：通用操控和多模态感知融合。人类的手经过数十亿年进化，和大脑的感觉运动系统深度耦合——从口袋里摸出钥匙、用筷子夹花生米，涉及触觉、力度控制、物体物理属性推断的实时融合。2035年的机器人在已知场景下表现优秀，面对未知物体——一个从未见过的不规则零件、一袋软塌塌的食品包装——表现急剧下降。如何让机器人获得对物理世界的”常识理解”，仍是开放问题。

**物理约束**（受基本物理定律限制）：能源密度是最硬的约束。按当前技术路线推演，到2035年前后最先进的商用锂电池能量密度可能达到约400Wh/kg，一台80公斤的人形机器人续航大约13小时——

勉强覆盖一个工作班次，而人类每天摄入约2500千卡就能维持全天活动。锂电池的理论上限约500Wh/kg，已接近天花板。执行器的功率密度和散热同样构成刚性瓶颈。

关键区别：工程问题可以用钱和时间解决；科学问题需要等待突破，时间线不确定；物理约束在当前技术范式下可能长期存在。把科学问题当工程问题处理（“砸钱就能解决”），或把工程问题当物理约束处理（“不可能做到”），是大多数预测偏差的根源。

## 非线性跃迁——火鸡问题的镜像

2025年前后的主流叙事是”渐进进步”，线性外推的结论是大规模部署还需十到十五年。但它忽略了两个非线性因素。

**技术融合的突发加速：** 大模型的推理能力、计算机视觉的泛化能力、强化学习的样本效率、传感器的成本下降——每一项在各自轨道上进步。当它们在2028到2030年间同时达到阈值时，组合效应产生了非线性跃迁。2028年之前，协作机器人需要专业工程师为每个新任务编写程序；2030年之后，基于大模型的任务规划系统使得机器人可以从自然语言描述或少量演示中学习新任务。

**部署规模的正反馈：** 更多部署带来更多真实数据，更好的数据训练出更好的算法，更好的算法降低部署门槛——这个循环在达到临界规模之前转得很慢，一旦突破临界点，加速是指数级的。

结果是阶梯形的进展曲线：2015到2027年长期平台期；2028到2030年第一次跃迁——结构化环境中的大规模部署变得经济可行；2031到2033年又一个平台期——科学问题限制了扩展；2034到2036年第二次跃迁——Robot Foundation Model的成熟使机器人可以在更广泛场景中工作。

## 不均匀渗透——环境结构化程度决定一切

**仓储物流最先被彻底改变。** 仓库几乎就是放大版的棋盘。按当前进展速度推演，到2031年前后，头部电商机器人有望识别抓取95%以上的SKU。如果这一目标实现，电商仓储拣选人力可能在十年内缩减超过70%。

**制造业装配是第二波。** 更深层的变化是”Design for Robotic Assembly”（DfRA）——产品结构被重新设计以适应机器人，而非让机器人适应产品。“环境适应机器人”而非”机器人适应环境”的策略在工业领域被证明极为有效。

**护理和服务是最后防线。** 设想这样一个场景：某日本养老院引入护理辅助机器人，三个月后大部分老人表示”不想让机器人碰自己”。一位老人说：“护士搬我的时候，她的手是热的。”这不是可以通过加热机器人手掌来解决的问题。

## 中国供应链的独特加速效应

在具身智能的全球竞赛中，中国占据了一个独特的结构性位置。

**供应链密度：** 在本书的推演中，深圳宝安区到2033年前后可能形成一个微型”机器人生态”——不到3公里的街道上聚集谐波减速器工厂、伺服电机装配线、碳纤维结构件加工厂、视觉模组测试实验室。一个创业团队可以上午提出新的关节设计，下午拿到样品，晚上装到机器人上测试。按照当前降本趋势，到2032年前后，中国企业有望将减速器成本降低60%以上，设计变更从图纸到样品的周期压缩到两周之内。

**场景规模 × 人口结构：** 中国拥有全球最大制造业基础，场景的多样性极为关键。同时劳动年龄人口持续下降，制造业招工难从”季节性问题”变成”结构性问题”。中国采用机器人的根本驱动力是劳动力短缺，而非单纯的成本优化——这个动力在全球主要经济体中独特。

**正反馈飞轮：** 大规模部署产生海量真实数据，数据反哺算法改进，改进后的算法降低下一轮部署门槛。按这一飞轮效应推演，到2031年前后，中国具身智能公司积累的真实场景运行数据量可能达到美国同行的一个数量级以上。如果趋势延续到2035年，全球人形机器人出货量前十名中可能有过半数是中国企业——不一定技术最先进，但大概率性价比最高、部署速度最快。

### “物理世界可编程化”——这才是真正的颠覆

需要一次认知翻转：具身智能的真正含义不是”机器人替代工人”，而是”物理世界变得可编程”。1995年，互联网让信息世界变得可编程——信息变得可复制、可搜索、可重组、可零边际成本分发。具身智能正在对物理世界做同样的事。当机器人可以根据软件指令操控物理对象时，一个在深圳工厂里被优化过的焊接策略，可以在24小时内被部署到全球所有使用同型号机器人的工厂。物理操作的改进不再是本地的、缓慢的、依赖师徒传承的——它变成了全局的、即时的、可指数扩散的。

替代只是表层效应。更深的变化是物理世界变得可编程——所有基于物理稀缺性的商业模式都将被重新定价。当认知自动化和物理自动化在2030年代中期形成”感知–决策–执行–反馈”的闭环时，AI就不再只是工具，而是一个可以独立作用于物理世界的主体。

---

## 六、第四阶段：服务业的认知增强与人力替代（2030年代末–2040年代初）

**被拆除的对象：** 教育、医疗、护理、心理咨询、社会工作——这些长期被认为”需要人类温度”因而不可替代的服务业。

**为什么排在最后？** 不是因为技术最难，而是因为社会抵抗最强。这些行业的核心价值叙事是”人与人的连接”——一个老师之所以重要不仅因为传递知识，更因为师生关系本身具有教育价值；一个医生之所以被信任不仅因为诊断准确，更因为面对面的问诊过程给予患者安全感。

挑战这个叙事在政治上是不可能的——没有政客敢说”我们不再需要老师和医生”。因此这些行业的替代不会以”取消岗位”的形式出现，而是以”重新定义角色”的形式出现。

**实际发生的过程：** 教师从”知识传授者”变成”学习教练”，AI负责个性化知识传递，教师负责情感支持和社交技能培养。医生的角色也在位移：AI完成绝大多数诊断工作，医生变成向患者解释诊断结果并做出治疗决策的人。护理领域走向”人机协同”，机器人负责搬运、翻身、生命体征监测，人类护理人员专注于心理陪伴。

**结果：** 岗位没有消失，但每个岗位能服务的人数大幅增加。一个教师能管理的学生数量增长数倍。一个医生从直接问诊变成管理大量AI预诊后的复核。**总岗位数大幅缩减，但”服务业被AI取代”这个说法在政治上始终被否认——因为每个剩下的岗位上确实有一个真人。**

这是整个替代过程中最精妙的政治操作：通过重新定义岗位内容，在名义上”保留”了人的位置，而在实质上完成了大规模的劳动力缩减。

---

## 七、第五阶段：创造性和战略性工作的价值重估（2040年代中期以后）

**被拆除的对象：** 到这个阶段，讨论的不再是”哪些工作被取代”，而是”工作”这个概念本身的含义发生了变化。

**核心变化：** 当AI在几乎所有标准化和半标准化认知任务上达到或超过人类水平，当机器人在大多数物理操作任务上达到或超过人类水平，“人类劳动”的剩余价值只集中在极少数领域——原创性的科学突破、真正的艺术创造、深层人际关系、政治判断和道德决策。

但这里有一个不可忽视的尾部风险：**这些”最后的堡垒”到底有多坚固?**

如果诚实面对，答案是”不确定”。2026年的我们可以有把握地说AI在未来十年内不会替代顶级科学家的创造力——因为科学突破的本质是在未知领域提出正确的问题，而提出正确问题需要的不仅是信息处理能力，还有对现实世界的深层直觉。但我们无法有把握地说二十年后的AI仍然不能做到这一点。Scaling Law可能撞墙，也可能出现新的架构突破。诚实的分析必须承认这个不确定性。

**反过来问：** 大多数人在问”哪些工作不会被AI取代?“——这是一个防御性问题。更有价值的问题是:”在一个AI可以做大多数事情的世界里，人类劳动的独特价值到底是什么?“这两个问题看起来相似，但指向完全不同的方向。前者是在旧框架里找安全角落，后者是在构建新框架。

**第五阶段的真正意义：** 不是”最后一批工作也消失了”（大概率不会全部消失），而是”工作”和”收入”之间的绑定关系——这个自工业革命以来维持了两百多年的社会契约——失去了经济基础。当大多数人的劳动不再是经济体系运转的必要条件时，“以劳动换取收入”就从一个制度安排变成了一个政治选择——一个可以被保留（通过人为创造”工作”）、也可以被替代（通过某种形式的收入再分配）的选择。

---

## 八、五个阶段的结构性总结

| 阶段 | 时间 | 目标 | 机制 | 速度 | 社会反应 |
| --- | --- | --- | --- | --- | --- |
| 一 | 2026–2029前后 | 标准化知识工作 | 静默压缩 | 快（18–24月/行业） | 近乎空白 |
| 二 | 2029前后–2030年代初 | 复杂知识工作 | 角色重构、层级消灭 | 中（3–5年/行业） | 个体焦虑，无组织抵抗 |
| 三 | 2030年代中期–末期 | 物理世界操作 | 机器人替代 | 中慢（受硬件成本和政治约束） | 立法抵抗、新卢德运动 |
| 四 | 2030年代末–2040年代初 | 人际服务业 | 角色重新定义 | 慢（政治最敏感） | 名义保留、实质缩减 |
| 五 | 2040年代中期+ | 创造/战略工作 | 价值重估 | 不确定 | “工作”概念本身的重新定义 |

**市场有自我减速机制——工资可以降、人可以流动、新需求会出现。** 这些不会挡住方向，但会把断裂拉成弯曲，这就是为什么整个过程需要二十年。

**关键模式：替代不是线性的。**

每个阶段之间存在相变点（phase transition）。从第一阶段到第二阶段的相变点是Agent可靠性跨过”多步任务”的门槛。从第二到第三的相变点是机器人硬件成本跨过人力成本的交叉点。从第三到第四的相变点是社会对”人机协同”的心理接受度。从第四到第五——这个相变点可能根本不在技术层面，而在政治和哲学层面。

**每个阶段都有”制度拖曳”：技术就绪但社会没准备好。** 制度拖曳不是bug，它是一种社会的自我保护机制，给人们时间适应、给制度时间调整、给政治时间回应。问题在于：如果拖曳的时间不够长（AI的扩散太快），而调整的速度又不够快（制度的惰性太大），两者之间的缝隙就会被痛苦填满。失业、贫困、焦虑、愤怒。

手工织工的七十年苦难，本质上就是技术替代速度和制度调整速度之间的缝隙。AI时代的这个缝隙会更短（因为替代更快），但不会更窄（因为制度惯性并没有比两百年前小多少）。更短但同样宽的缝隙意味着：冲击会更集中、更剧烈。

**这是速度差的必然结果。** 当一个系统的输入端（技术变革速度）和输出端（制度调整速度）之间存在持续的、不可消除的速度差，中间必然积累压力。这个压力最终会以某种形式释放——可能是有序的制度创新（新的社会契约），也可能是无序的社会动荡（抗议、冲突、政治极化）。

下一章，我们先看一个特殊的战场——在所有行业中最先被AI重写底层代码、也最清晰地展示了认知基础设施扩散后果的领域：内容产业。

# 第八章 内容泛滥——当作品变成空气

> 问题从来不是”AI能不能写出好东西”。问题是：当所有东西都是好东西的时候，“好”这个字还剩下什么含义。

---

2028年秋天，一个姑娘坐在出租屋里盯着后台数据，试图理解发生了什么。她做了六年美食公众号，从大学写到工作，一篇一篇磨出来的探店文章攒下了十二万粉丝。那年夏天开始，阅读量像漏气一样往下掉——不是慢慢掉，是两个月内腰斩。她以为是自己写得不好了，逼自己加班，一周更新三篇，拍照修图写文案到凌晨两点。直到有天她在推荐流里刷到一篇文章，写的是她上周刚去过的那家馆子，配图精美、文笔流畅、细节丰富——评论区有人问”你们是几号去的”，账号没有回复。她点进主页，那个号一天更新八篇，覆盖全城各区，每篇都像是真的去吃过。她盯着屏幕看了很久，那种感觉不像是遇到了一个更勤快的同行，更像是发现自己一直在跟影子赛跑。后来从2046年往回看，她的故事没什么特别的——那一年，几乎所有靠内容手艺吃饭的人都经历了类似的时刻。只是当时没有人知道，这不过是洪水漫过脚面的最初几分钟。

---

上一章讲的是工作的消亡。但在工作全面消亡之前，有一个领域已经率先沦陷——内容创作。

## 一、从稀缺到泛滥的相变

在所有被AI重塑的领域中，内容产业是第一个完整经历”从稀缺到泛滥”相变的行业。它的速度比劳动市场整体替代快得多，后果也更清晰——因为内容的供需关系可以被精确量化。

以短视频平台为例（以下为基于当前趋势的推演估算）：一个日活过亿的视频平台，日均新增视频可能从2027年的约80万条，在两年内膨胀到600万条以上。增量的绝大部分来自AI生成或AI辅助生成的内容。同期日活用户可能从1.1亿增长到1.4亿——用户增长约27%，内容供给增长约650%。

图文和音频平台同样如此（以下均为趋势推演）。一个主流图文平台的日均发布量可能从约300万篇在两年内突破1500万篇。种草社区的笔记发布量可能在两年内增长八倍。播客平台上AI生成的有声内容占比可能从不到5%在两三年内超过40%。

供需失衡的数学很简单：内容供给增速远超注意力供给增速，单条内容的平均价值必然下降。驱动力是第六章论证的成本坠落——当一条十五分钟科普视频的制作成本从两百小时降到两三小时，供给会迅速膨胀到远超需求的水平。

## 二、“作品”概念的消亡

“作品”这个概念值得追问。

**“作品”需要两个条件才能成立：生产成本显著高于零，以及生产者数量有限。** 一首交响曲是作品，因为写出它需要数月甚至数年的专业劳动，能写出它的人全世界不超过几千个。一篇深度调查报道是作品，因为记者需要数周的采访、查证和写作，能做到这个水平的记者每家媒体只有几个。

“作品”的概念嵌套在一整套文化制度中——版权法（保护生产者的投入回报）、评论体系（区分好作品和坏作品）、奖项（标记最好的作品）、教育（培训能生产作品的人）。这些制度的共同前提是：作品是稀缺的，因此值得被保护、被评判、被奖励。

当AI把内容的生产成本推向趋零时，这个前提崩塌了。

一个类比会有帮助。在冰箱发明之前，冰是一种昂贵的商品——在19世纪的美国，切割、运输和储存天然冰块是一个年产值数百万美元的产业。冰商有品牌，冰有质量分级，人们愿意为”纯净的湖冰”支付溢价。冰箱普及之后，冰变成了你在家里按一个按钮就能得到的东西。冰的物理性质没有改变——它仍然是冰。但”冰”从一种商品变成了一种公用事业。不再有人讨论冰的”品质”，不再有冰的”品牌”，不再有关于”最好的冰”的评奖。

AI对内容做的事情，和冰箱对冰做的事情在结构上完全同构。当一篇七十分的文章可以在三十秒内生成，当一条八十分的视频可以在五分钟内渲染完成，当一首九十分的流行歌曲可以在十秒内作出——内容就不再是”作品”，而是”环境”。就像空气中有氧气但你不会为氧气付费一样，数字空间中充满了内容但单个内容的价值趋近于零。

一个常见的误解是”人人都能创作”意味着创作的民主化。但这个框架本身就是错的。“创作者”这个角色正在失去意义——就像冰箱普及后”冰商”消失了一样。不是有更多人做冰商了，是这个职业本身没有了。

## 三、注意力经济的第二次相变

内容泛滥不是第一次发生。注意力经济经历过一次相变，现在正在经历第二次。

**第一次相变发生在印刷术普及之后（15世纪末–17世纪）。** 古登堡的活字印刷让书籍的生产成本骤降约95%。信息从极度稀缺变成了相对丰富。后果是什么？

大多数人以为答案是”知识传播、思想启蒙”。这确实是长期后果。但短期后果——印刷术普及后的第一个一百年——主要是混乱。大量低质量的宗教小册子、煽动性传单、伪科学著作和色情读物涌入市场。德国新教改革之所以能迅速扩散，部分原因是马丁·路德无意中成了第一个”内容营销大师”：他

的九十五条论纲在发布后几周内就被印刷了数十万份。但与此同时，反犹太主义的传单、猎巫手册（《女巫之锤》）和各种末日预言也通过同一种技术大规模扩散。

**信息泛滥的第一个后果不是启蒙，是混乱。** 启蒙要等到新的筛选机制出现之后才逐渐实现：科学方法、学术期刊、新闻业的职业伦理。这个过程花了两百年。

**AI引发的是第二次相变。** 规模不可同日而语——印刷术把信息生产成本降低了一到两个数量级，AI降低了三到四个数量级。而且印刷术只降低了文字的复制成本（原创仍需人力），AI降低了原创本身的成本。

如果第一次相变的短期后果是混乱，第二次相变的短期后果是什么？

答案是：**注意力市场从”洪水”变成了”海洋”。**

“洪水”意味着信息量超过了你的处理能力，但你仍然能辨认水流的方向、找到高地。这是2020年代之前的信息过载状态——信息太多，但你可以通过信息素养、筛选工具和社交过滤来应对。

“海洋”意味着你连方向感都失去了。当每一个方向都有无穷无尽的内容向你涌来，当AI可以根据你的偏好精确定制信息流，当真实和虚构的边界被AI生成的超现实内容模糊——你不再是在选择信息，而是在被信息塑造，且对这个过程毫无察觉。

**AI并没有缓解信息过载——它让信息过载从一个技术问题变成了一个认知论问题。** 在信息洪水中你还知道自己不知道什么。在信息海洋中你不再知道自己被什么塑造。

---

## 插曲：同一个人，两个时代（以下为虚构推演）

2026年，一位内容创作者每天早上六点起床，花四十分钟通勤到工作室，打开剪辑软件，开始一帧一帧地调色、配乐、写旁白。她做历史科普视频，每条十二分钟，从选题到发布要磨五天。一个月四条，其中一条爆了，够付房租和器材分期。她在圈子里小有名气，粉丝经常在评论区讨论细节——“三分二十秒那个转场太绝了”。

2036年，她坐在同一张桌子前。她的AI工作流每天能产出六条同等质量的视频——选题、脚本、配音、剪辑、调色，全流程两小时。但她不再发了。不是技术问题。是她打开任何一个平台，搜”历史科普”，出来的前一百条都很好。每一条都很好。好到她分不清哪条是人做的。她的粉丝没有取关，只是评论越来越少了——他们的注意力被稀释在了无穷多个”都挺好”的内容里。

她现在的工作是给一家教育公司做”内容策展”——从每天涌入的三千条AI生成的历史短片中，挑出最适合不同年龄段学生的二十条。收入比2026年高一点。但如果你问她”你还是创作者吗”，她会沉默很久。再过两年，2038年的某个深夜，一个曾经写代码的男人坐在窗前回忆十二年的变化时，会发现老同学聚会上十三个人里只有四个还在做和十年前”差不多”的工作。没有人看起来特别不快乐，也没有人看起来特别确定。大家都在一种”走着看”的状态里。

## 四、真人稀缺效应——反脆弱的对称面

内容泛滥有一个反直觉的对称面：**AI越强，真人越稀缺越值钱。**

这看起来矛盾，但逻辑是自洽的。

当AI生成的内容从辨别度低（2026年）到几乎无法与人类作品区分（2030年以后），一个新的稀缺性被创造出来——“确认这是真人做的”本身变成了一种价值。

类比：在工业化大规模生产普及之后，手工制品不但没有消失，反而在高端市场获得了更高的溢价。一只爱马仕手工缝制的Birkin包售价数万美元，不是因为它比机器缝制的包更结实（可能还没有），而是因为”由真人手工完成”这个属性本身被赋予了稀缺价值。瑞士手表在石英表普及后反而涨价了——计时精度更差，但”机械工艺之美”变成了奢侈品叙事。

AI时代的内容产业正在经历同样的分化。

在底部，标准化的资讯、教程、文案、娱乐内容由AI大规模生产，边际成本趋零，消费者无需为此付费。这构成了数字空间的”环境温度”——无处不在，免费获取，没有人关心它是谁写的。

往上一层，大量AI生成的内容甚至不是为了被人直接消费，而是为了喂养推荐算法和搜索引擎。一个AI写手账号每天发布六十条视频，不是因为它想吸引观众，而是因为它在用概率论撒网——其中一条可能被算法推上热门。内容不再是为读者创作的，而是为算法创作的。读者只是算法变现的最终端口。

到了顶部，真人创作反而变成了奢侈品。当一个音乐人可以证明”这首歌是我真的在录音棚里用真的乐器演奏并亲自混音的”，当一个作家可以证明”这本书的每一个字都是我敲出来的”，当一个记者可以证明”我真的去了现场采访了当事人”——这种”真人属性”本身变成了高价值的稀缺信号。

逻辑是简单的：AI内容越多，“这是真人做的”这个信号就越值钱。

但这里有一个残酷的对称性：只有能够证明自己是”真人”的创作者才能享受这个溢价。证明机制本身变成了一种新的门槛。你需要有足够的公众知名度、足够的创作历史、足够的可验证的身份信息来让受众相信”这真的是他/她做的”。对于无名的个体创作者来说，“真人”属性几乎无法变现，因为没有人认识你，也就没有人关心你是不是真人。

**结果：内容产业的基尼系数急剧升高。** 顶部的少数真人创作者（拥有品牌、粉丝基础和可验证身份的那些人）收入可能反而上升。底部的大量创作者，无论是人类还是AI，收入趋近于零。而中间层，过去靠技能和勤奋维持体面收入的专业内容工作者，被彻底挤出。

## 五、“创作”还是”策展”？角色的根本位移

当内容变成空气，人的角色从”创造者”转变为”策展者”（curator）。

这个转变的逻辑：当生产成本趋零而信息总量趋于无穷时，价值从”生产”转移到”筛选”。一个人/组织能否在无穷的信息海洋中帮你找到最值得注意的那1%——这变成了最稀缺的能力。

类比：在物理世界中，钻石的价值不在于碳原子（碳无处不在），而在于特定排列方式的稀缺性和可验证性。在内容海洋中，信息的价值不在于信息本身（无处不在），而在于特定的选择、排列和背书的稀缺性。

这意味着”创作”这个词的含义发生了根本位移。

传统意义上的”创作”：从无到有地生产一个此前不存在的作品。这需要技能、灵感和劳动——三者的组合构成了创作的门槛和价值。

新语境下的”创作”：在无穷的已有内容中做出选择、赋予上下文、建立联系、提供视角。这需要的核心能力是判断力——什么值得被看到？在什么上下文中它变得有意义？它和其他什么内容形成了怎样的对话？

真正有价值的策展需要深度理解和独特视角——理解受众真正需要什么（往往是他们自己还没意识到的东西），以及为什么这个特定的内容在这个特定的时刻是重要的。

不过这里有一个关键的转折：AI也在快速提升策展能力。按当前趋势，2030年以后AI推荐系统的个性化精度很可能超过大多数人类编辑。如果策展也能被自动化，那”人在内容产业中的独特角色”又退到了哪里？

答案可能是：退到”信任”这个维度。你接受一个AI推荐的内容，和你接受一个你尊敬的朋友推荐的内容，信息可能一样，但心理效用不同。后者包含了一种社会关系的确认——“他了解我，他觉得这个值得我看”。这种基于人际信任的策展，是AI最难替代的维度。但它的经济体量，远小于”内容生产”曾经支撑的市场规模。

---

## 六、制度的滞后——版权、平台、教育

内容泛滥最深层的影响不在经济面，而在制度面。围绕”作品”概念建立的一整套制度体系（版权法、平台规则、教育体系）都面临根本性的挑战。

**版权法的危机。** 现代版权法的核心假设是：创作需要投入（时间、技能、灵感），因此创作者有权对其作品的复制和传播享有排他性权利。当AI把创作的边际成本推向零时，这条从”投入”到”权利”的因果链被打断了。一个AI在十秒内生成的歌曲，是否应该享有和人类音乐人花三个月写出的歌曲同等的

版权保护？如果不应该，界线在哪里？如果一首歌是人类出prompt、AI生成、人类修改——创作者是谁？

可以预见，到2030年前后，全球主要司法管辖区对这些问题可能仍没有一致的答案。美国版权局已经在2023年明确”纯AI生成的作品不享有版权保护”的基本方向，但对”人类指导AI生成”的作品如何定性语焉不详。中国的版权法体系面临类似的修订压力，但”AI辅助创作”的边界极难界定，执行标准注定模糊。欧盟的AI Act在版权问题上采取了拖延策略——把问题留给了各成员国自行处理。

**平台规则的困境。** 内容平台面临一个两难：如果不限制AI内容，平台被AI生成的内容淹没，人类创作者流失，长期生态恶化。如果严格限制AI内容，竞争对手不限制，创作者和流量会流向更宽松的平台。这是一个经典的囚徒困境——单个平台无法单方面限制AI内容而不承受竞争代价。

**教育体系的断裂。** 全球有数百万人正在接受”内容创作”相关的教育——新闻学、文学创作、影视制作、设计、音乐。这些专业的核心假设是：经过系统训练的人类创作者在市场上具有竞争优势。当这个假设不再成立时，这些专业的毕业生将面临什么？

这不是一个可以用”转型”来轻松回答的问题。一个花了四年学习新闻采写的毕业生，很难轻松”转型”为AI提示词工程师——这两种工作需要的核心能力有交集但不重叠。转型的摩擦成本被严重低估了。

---

## 七、印刷术的启示——混乱之后是什么

让我们回到印刷术的类比来收束本章。

印刷术普及后的一百年，混乱是主旋律：宗教战争、政治煽动、伪科学泛滥。但两百年之后，新的秩序浮现了。科学方法提供了区分真知与伪知的标准，新闻业的职业伦理提供了信息质量的底线，公共教育提供了公民辨别信息的基本能力。这些制度不是技术自动产生的，它们是社会在混乱中学习的产物。

AI内容泛滥之后，同样需要新的秩序。这个秩序可能包括：

内容来源的可验证体系（类似于区块链的内容溯源技术，但需要比现有方案更成熟的实现）。新的”信息素养”教育（不仅是”如何辨别虚假信息”，而是”如何在无穷信息中保持独立判断”——一种认知时代的基本技能）。新的经济模型（当内容的生产价值趋零时，价值转移到了哪里？——可能是”注意力引导”、“信任网络”、“策展品味”等此前不被视为经济活动的领域）。

但这些新秩序的建立需要时间。印刷术后花了两百年。AI后大概会快很多——因为我们有印刷术后积累的制度经验可以借鉴，也因为AI本身可以被用来加速新制度的设计和测试。但即使快十倍，也需要一到两代人的时间。

**在新秩序建立之前的这段过渡期，混乱是不可避免的。** 内容泛滥只是这场混乱的一个表征。更深层的混乱在于：人类用创作定义自我已经几千年了。当这套系统被AI稀释之后，问题不再是”谁来创作”，而是”创作还能不能承载意义”。

这个问题超出了经济学的范畴。但在回答它之前，我们需要先看一个更基础的问题：生产这些内容的基础设施（算力、能源、数据）掌握在谁手中？这个问题是结构性的，它的答案将决定内容泛滥时代的权力格局。

这是下一章的主题。

# 第九章 算力、能源与新圈地运动

> 二十世纪的核心问题是谁控制生产资料。二十一世纪的核心问题是谁控制认知资料。但大多数人还在问第一个问题。

---

2046年，一个中年人坐在深圳南山区一间共享办公室里，翻看二十年前自己写的融资BP。那是2026年春天，他刚从一家大厂离职，带着三个人的小团队做一款医疗影像AI。技术不是问题——开源模型已经足够好。问题是算力。一张H100显卡的租赁价格吃掉了他们三分之一的天使轮资金，而要达到三甲医院要求的推理精度，至少还需要十倍的训练量。他去找过几家云厂商谈折扣，对方客客气气地报了一个数字，然后补了一句：“这还只是算力的价格，数据标注另算。”他算了一笔账：获取一万份脱敏CT影像的授权费用，比他团队半年的工资总和还高。

那年夏天，他关掉了公司。不是因为技术不行，不是因为市场不在，而是因为他发现自己在和一种结构性力量对抗——算力和数据这两项最基本的生产资料，已经被牢牢锁在几家巨头的保险箱里。二十年后回头看，他才意识到自己当时撞上的，正是一场静悄悄的圈地运动的围栏。

“圈地运动”——这个词不是修辞。十六世纪的英格兰，大地主用围栏把公共牧场圈为私有；二十一世纪的科技巨头，用算力门槛、数据壁垒和模型围墙把认知的公共资源圈为私有。技术完全不同，权力动态高度同构。这个比喻将贯穿本章：谁拥有认知时代的”土地”，谁就定义这个时代的权力结构。

---

前面几章讨论的是AI”能做什么”。本章追问一个更冷酷的问题：AI需要什么？答案是算力和能源——而争夺已经开始。

## 一、21世纪的生产资料清单

第六章论证了认知正在变成基础设施。第八章展示了这个基础设施的产出——内容——已经泛滥到淹没一切。现在我们必须追问一个更根本的问题：**生产这些内容、驱动这个基础设施的底层资源，掌握在谁手中？**

在任何文明阶段，定义权力结构的从来都是底层的生产资料控制，而非上层的政治安排。农业文明的生产资料是土地和水源。工业文明的生产资料是工厂、矿山、铁路和石油。那么认知文明的生产资料是什么？

列一张清单：

**算力。** 训练前沿AI模型需要的计算能力集中在全球少数几个地方。2026年，英伟达在AI加速芯片市场的份额超过80%。一座能支撑前沿模型训练的数据中心投资在十亿美元以上。全球AI算力的地理分布：美国约40%，中国约25%，欧洲不到15%，其余国家加起来不到20%。

**能源。** 2026年全球数据中心年耗电约1000太瓦时，占全球发电量3.5%。AI推理需求每年翻倍，到2030年可能达到8–10%。能源不只是"成本"——它是算力的物理上限。

**数据。** 高质量文本数据的存量有上限。专业领域数据（法律、医疗、金融）大多掌握在机构手中。数据的语言分布高度不均——英语占全球网络文本60%以上。

**模型。** 前沿基础模型的训练是一种寡头竞赛——全球能独立训练万亿参数级模型的组织不超过十几个。

**机器人网络。** 第七章论证的物理世界自动化，需要的不仅是软件，还有硬件——通用机器人的设计、制造和部署能力。

**分发接口。** 控制用户触达AI能力的入口——操作系统（Apple、Google、Microsoft）、超级应用（微信、抖音）、云服务API——等于控制了认知基础设施的"最后一公里"。

六项生产资料。每一项都具有强烈的规模效应——投入越大，壁垒越深。这不是市场竞争的一般特征，而是一种物理定律级别的约束：AI系统的效率和经济性随规模指数增长，这意味着**大者恒大不是策略，是结构**。

用圈地运动的框架来理解：这六项资源就是认知时代的"土地"。谁圈得多，谁就是新时代的大地主。

---

## 二、权力集中的物理学

**AI系统的规模效应天然导致权力集中。这不是阴谋，是物理学。**

让我们拆解四种自增强效应：

**数据的网络效应。** 用户越多，产生的数据越多；数据越多，模型越好；模型越好，用户越多。这个正反馈循环和社交网络的网络效应在结构上完全同构——只是前者是认知维度的，后者是社交维度的。一旦某个AI系统获得了足够多的用户和数据，后来者即使技术水平相当，也很难追赶其数据优势。

**训练成本的门槛效应。** 训练一个前沿模型的成本在2026年约为5–10亿美元（含算力、数据、人才）。这个门槛本身就构成了一个准入壁垒——全球能跨过这个门槛的组织屈指可数。

开头那位深圳创业者的遭遇正是门槛效应的微观投影。他不缺技术、不缺想法，缺的是跨过门槛的入场券。他的竞争对手不是另一个创业者，而是一整套让小玩家无法入场的成本结构。

**人才的聚集效应。** 全球顶级AI研究人员大约在1万人左右，高度集中在少数机构——2026年，仅OpenAI、Google DeepMind、Anthropic、Meta AI四家公司就雇佣了其中的约40%。人才聚集带来更好的模型，更好的模型吸引更多资金，更多资金推高薪资，更高的薪资又吸引更多人才。又一个正反馈循环。

**基础设施的先发优势。** 数据中心建设需要2–3年，芯片设计到量产需要3–5年，电力基础设施的扩建更是以十年计。先行者的领先优势长达五到十年，而后来者追赶的目标还在加速远离。

四种效应叠加，结论清晰：**AI产业的自然均衡态是寡头垄断，而不是完美竞争。** 2026年，全球前十大AI公司控制了超过70%的公共云AI算力、80%的前沿模型训练能力、60%以上的AI相关专利。这十家公司中七家是美国的，三家是中国的。欧洲、日本、印度、东南亚、非洲、拉丁美洲——加起来不到10%。

---

## 三、“AI取代你的工作”是错误的问题

当所有人都在讨论一个问题时，先检查这个问题本身是否是正确的问题。

2026年到2030年间，公共讨论中关于AI最热门的问题是”AI会不会取代你的工作？“——第七章已经详细回答了这个问题。但这不是最重要的问题。

**最重要的问题是：“AI取代你之后，AI归谁所有？”**

这两个问题指向完全不同的方向。前者的答案是”会，但可以通过再培训、终身学习来适应”——一个个人解决方案。后者的答案是”归少数科技公司和它们的股东所有”——一个结构性问题，不能通过个人努力解决。

把这个逻辑推到极端：假设到2040年，AI和机器人系统可以完成人类经济活动的80%。如果这些AI系统的所有权高度集中——全球十家公司和它们的股东拥有其中的60%——那么这十家公司的股东就控制了全球GDP的约48%。

这不是”资本主义”——历史上的资本主义从未达到过这种集中度。这更接近于封建制度：少数人拥有生产资料（AI系统），其余人要么为这些系统工作，要么完全不参与经济生产。

如果你还记得间奏二中那位广告公司的文案总监——她从”掌勺的厨师”变成了”试菜的人”，收入没变，但”手感”没了。她至少还留在了厨房里。更多人被直接赶出了门。

再设想一个场景：某一天，一位在马德里经营翻译工作室的女性决定关掉她经营了十几年的公司。AI翻译已经足够好，她的客户逐渐流失。她并不怨恨AI——让她真正愤怒的是另一件事：训练那个翻译AI的语料里，有她过去十几年上传到翻译平台的几十万条翻译记录。她的工作被用来训练取代她的工具，而她没有收到一分钱补偿。“公共牧场”上她种的庄稼，被圈进了别人的围栏。

历史不断重复同一个剧本：**当生产资料的所有权集中度超过某个阈值时，社会契约就会破裂。** 这个阈值取决于政治制度的弹性、社会保障体系的厚度和公众的忍耐度。但在每一个文明周期中，这个阈值都曾被触及——先是社会冲突，然后制度重构，最终达成新的权力均衡。

我们正在接近这个阈值吗？如果全球GDP中由AI系统生产的比例持续上升，而AI系统的所有权继续保持当前的集中度——答案是”是的，时间问题”。

---

## 四、从公共牧场到私有农场

理解当下正在发生什么，最好的方式是看看历史上已经发生过的同结构事件。

英国圈地运动（16–18世纪）的核心机制很简单：公共牧场（commons）原本是村庄共有的资源，大地主通过议会法案将其圈为私有，用围栏封闭，改为更高效的商业化耕作。经济效率毫无疑问地提高了，但社会后果是灾难性的：数百万失去土地的农民被迫涌入城市，填进工厂的车间。马克思后来给他们取了个名字：“无产阶级”。

**AI时代的”新圈地运动”在结构上与此完全同构。** 公共的”认知牧场”——互联网上的开放知识、免费信息、公共数据——正在被科技公司圈入私有。三道围栏同时在竖起：

**数据围栏。** 大型科技公司利用用户数据训练AI模型，然后将模型的能力作为付费服务出售。用户的数据是”公共牧场”，训练出的模型是”私有农场”，从公共到私有的转化就是”圈地”。GDPR和各国数据保护法是对”圈地”的部分抵抗——但效果有限，因为数据一旦被用于训练就无法”归还”。

**算力围栏。** 云计算的本质是”你不需要自己买服务器，你可以租我的”——初期降低了使用门槛，但长期意味着算力控制权从用户转移到了平台。当所有的AI推理都在少数几家公司的数据中心中运行时，这些公司就控制了”认知基础设施的开关”——可以决定价格、决定谁能访问什么级别的AI能力、决定什么样的内容可以被生成。

**模型围栏。** 开源模型（Llama、Qwen等）在2025–2026年确实打破了部分壁垒。但前沿模型的训练成本持续上升意味着，开源与闭源前沿之间的差距可能重新拉大。更重要的是，没有算力来运行、没有数据来微调、没有工程能力来部署，开源模型就只是一个理论上的可能性。

圈地运动的历史教训是一条张力线：效率提升是真实的，但收益高度集中在地主手中，成本由失去土地的农民承担。从效率看，圈地是成功的；从公平看，圈地是灾难；拉到文明长周期看，圈地既是必要的（为工业化释放了劳动力），也是可以更人道的（如果有更好的补偿机制）。

AI时代的新圈地运动面临同样的张力：规模效应确实需要大规模的资本和数据集中——这在物理上是效率最优的。但如果集中的收益全部归少数所有者，而成本由整个社会承担，社会契约就会断裂。

---

## 五、芯片战争、能源安全、数据主权——地缘政治的三条战线

认知生产资料的集中不仅是经济问题，更是地缘政治问题。国家之间的权力竞争正在围绕三条新战线展开。

### 芯片：认知遏制的第一战

2022年美国对中国实施的芯片出口管制，本质上是一次”认知遏制”——通过限制高端AI芯片的供应，试图延缓中国AI能力的发展。这是人类历史上第一次一个国家试图通过控制认知生产资料来限制另一个国家的认知能力。

效果是复杂的。短期看，确实延缓了中国在前沿模型训练上的进度。中期看，它加速了中国芯片自主化——华为昇腾、寒武纪等国产芯片已在取得显著进展，虽然仍落后于英伟达最新产品，但差距在缩小。长期看，它可能催生了一个分裂的全球AI生态——美国体系和中国体系各自独立发展，数据不互通，标准不统一，模型不兼容。

但这里有一个二阶效应值得单独拎出来：**芯片管制的长期后果可能与初衷完全相反。** 如果它成功迫使中国建立了完全自主的芯片和AI生态系统，那么美国就永久失去了通过供应链施加影响的能力。压力没有杀死对手，反而让对手变得更强壮——一个典型的”反脆弱”效应。

再深入一层。**算力不是石油那样的自然资源，而是制造品——瓶颈不在开采，在制造。** 一块先进AI芯片需要穿越一条极其漫长且脆弱的供应链：芯片设计巨头定义架构，极少数先进制程代工厂完成制造，这些代工厂依赖的极紫外光刻机来自更少数的关键设备供应商，往上游还有光刻胶、高纯硅片等特种材料。每一层都高度集中，任何一个环节中断都足以让整条链路停摆。

2026年前后，全球能量产最先进制程芯片的工厂用一只手就能数过来。这意味着算力的”圈地运动”有两个战场：表层是科技巨头之间对算力配额的争夺，深层是国家之间对芯片制造能力的产业政策博弈——建厂补贴、出口管制、技术封锁，本质上都是在争夺供应链上的控制节点。

还有一个维度容易被忽略：举国体制对AI基础设施的投入。以”东数西算”工程为例，从规划到建成可能只需数年，涉及十多个省级行政区的协调——在任何需要通过市场机制协调这么多利益方的国家，这种时间表不可想象。对比：美国CHIPS法案2022年签署，按照此前大型联邦项目的执行经验，实际拨付进度往往慢于立法承诺。

但举国体制的代价同样显著。按以往经验推演，大量”AI产业园”可能入驻率严重不足。在本书的推演中，2029到2032年间的投资潮，大约三分之一可能转化为真正有效的基础设施，三分之一需要运营数年后才能证明价值，剩下三分之一可能成为沉没成本。不过，冗余不全是浪费——如果具身智能在2033年前后迎来爆发期，部分此前闲置的算力设施可以被迅速征用，响应速度远超从零建设。

## 能源：认知时代的石油

以弗吉尼亚州劳登县为例——全球数据中心密度最高的地区之一——当地电网已经面临巨大压力，数据中心用电量快速逼近居民用电总和。按照当前AI算力扩张速度，类似地区在未来几年内不得不限制新的数据中心建设几乎是必然结果。当居民发现夏天电费持续上涨，而那些巨大的灰色建筑里运行的AI和他们的生活没有任何关系时，政治压力将不可避免。

这不是孤例。谁拥有廉价、稳定、大规模的能源供应，谁就在AI时代拥有结构性优势。这正在重塑全球能源地缘政治。

中东国家正在从”石油输出国”转型为”算力输出国”。沙特和阿联酋大规模投资AI数据中心，用廉价的石油和天然气发电来驱动算力。冰岛和挪威的地热和水电吸引了大量数据中心建设，冷气候还提供了天然的冷却优势。美国的能源优势不仅在于总量，还在于多样性——核能、天然气、风能、太阳能的组合提供了高度灵活的电力供应。

**能源和认知的绑定，正在重复石油和工业的历史绑定。** 二十世纪的两次世界大战、中东地缘政治的数十年动荡、石油美元体系的兴衰——底层驱动力都是”谁控制工业的能源”。二十一世纪的地缘冲突将越来越多地围绕”谁控制认知的能源”展开。

## 数据主权：看不见的国境线

在本书的推演中，这样的声音迟早会出现——某个发展中国家的程序员在社交媒体上写道：“我们国家一亿多互联网用户的数据，训练出了别人的AI。现在那个AI回来卖给我们，按月收费。这不是贸易，这是殖民。”

他说的是一个结构性事实。欧盟的GDPR、中国的《数据安全法》和《个人信息保护法》、印度的《数字个人数据保护法》——本质上都是在划定数据领土的边界。

数据主权的争夺有一个内在矛盾：AI模型的性能与训练数据的规模和多样性正相关，而数据主权的划分天然限制了数据的跨境流动。效率要求数据集中，主权要求数据分散——两个目标在物理上互相矛盾。

长期均衡态可能是：全球形成几个”数据区域”——美国主导的英语区域、中国主导的中文区域、欧盟试图建立的第三极、以及印度和东南亚正在争取的自主空间。每个区域内数据相对自由流动，跨区域流动受到严格限制。AI模型的能力因此也出现区域分化。

这种分裂的效率损失是巨大的，全球统一的数据池能训练出比任何区域数据池更好的模型。但效率不是唯一的考量。主权、安全、文化自主性——这些非经济因素在国家决策中的权重远大于”训练一个更好的模型”。

## 六、“所有权”问题才是核心问题

让我们收束本章的论证。

第六章论证了认知正在变成基础设施。第七章论证了这个基础设施正在替代人类劳动。第八章论证了它的产出正在淹没内容市场。本章揭示了一个更底层的结构：**驱动这一切的认知生产资料——算力、能源、数据、模型、机器人网络、分发接口——高度集中在少数公司和国家手中。**

这种集中不是偶然的，也不是阴谋的结果。它是AI系统的规模效应、网络效应、门槛效应和先发优势的自然产物。物理定律般的必然性。

但”自然”不等于”合理”。圈地运动是经济效率的自然产物，但它造成的社会灾难最终催生了劳动法、济贫法、工会运动和福利国家。石油的集中控制是地缘政治的自然产物，但它催生了石油危机、中东战争和可再生能源革命。

回到本章的核心追问：**“AI会不会取代你的工作”不是正确的问题。正确的问题是”AI取代你之后，AI归谁所有？“** 如果答案是”归少数人所有”，那么AI带来的生产力增长就不会以提高普通人生活水平的形式体现——它会以增加少数人财富的形式体现。这不是一个技术问题，这是一个所有权问题。

这些人撞上的是同一道围栏。

历史反复警告：**当生产资料的集中度超过社会的忍耐阈值时，冲突是不可避免的。** 每一次生产资料的重大集中都伴随着剧烈的社会冲突，并最终催生新的制度安排来重新平衡权力。

拉长到文明尺度来看：**生产资料的集中和再分配是一个周期性过程。集中创造效率和财富，但财富过度集中引发不满，不满催生制度创新和再分配——直到新的集中周期开始。** 我们正处在这个周期的”集中”阶段。下一个阶段——“不满催生制度创新”——已经开始。

这就是下一章的主题：当权力集中到一定程度，社会如何反应？冲突以什么形态出现？制度创新的种子在哪里？

# 第十章 抵抗与调适——转型的社会摩擦

2034年深秋，一个女人坐在某市劳动仲裁委员会的走廊里，手里攥着一份四十七页的仲裁申请书。她曾是一家国际律师事务所的资深法律翻译，十四年从业经验，三年前被”优化”。裁员通知用的措辞是”业务结构调整”，但整个翻译部门在三个月内被一套AI系统替代了。她申请仲裁，不是为了赔偿金——那笔钱早已到账——而是为了一个没有先例的法律问题：当一个人的全部职业技能被AI覆盖，劳动合同中”提供同等岗位”的义务还有没有意义？

仲裁庭的回答她已经猜到了。现行劳动法没有为这种情况准备任何条款。法律还停留在”岗位消失可以调岗或补偿”的框架里，而她面对的现实是——不是这个岗位消失了，是这个职业消失了。走廊里还坐着十几个人，年龄从二十八到五十五岁不等，手里都攥着类似的文件。他们不认识彼此，但他们构成了同一个历史趋势的截面。

从2046年回望，她们的困惑正是旧社会契约断裂的最初裂痕。当时没有人知道的是，裂痕才刚刚开始扩大。

---

> 每一项你今天认为理所当然的劳动权利——八小时工作制、最低工资、带薪休假、工伤保险——都不是哪个善良的资本家主动给的。它们是多元利益反复协商与博弈的产物。如果你觉得这一次转型可以绕过利益表达和制度协商，你不是乐观主义者，你是对历史缺乏基本了解。

---

前九章基本在拆——拆旧秩序的每一根支撑。本章终于转向另一面：社会不会坐以待毙，抵抗和调适同时在发生。

## 一、波兰尼的”双重运动”——一个被遗忘的预言

第二章已经引入了波兰尼的双重运动框架：市场力量每一次扩张都会激发社会的自我保护运动。现在让我们看看这个框架在AI时代的具体展开。

**AI时代正在重演波兰尼的双重运动。**

第一重运动：AI的市场扩张——认知基础设施化（第六章）、劳动替代（第七章）、内容泛滥（第八章）、认知生产资料的集中（第九章）。这些是市场逻辑和技术效率的自然展开。

第二重运动：社会的自我保护——这正是我们将在本章考察的对象。

波兰尼最锐利的判断在于：**双重运动不是可选的。** 它不取决于某个政治领袖是否”决定”发起保护运动，不取决于某个行业组织是否”选择”表达诉求。它是系统性的、自发的、不可避免的——就像你把手放在火上会缩回去一样，社会在受到威胁时会产生保护反应。问题不是”会不会有利益表达”，而是”利益表达以什么形态出现、在什么强度上稳定下来”。

---

## 二、四种社会张力形态

在本书的推演中，从2020年代末到2040年前后，社会保护运动可能以四种可辨识的形态展开。它们在不同地区、不同时间以不同强度并存。

### 第一种：劳动者抵抗

在本书的推演中，2030年前后，欧洲某国可能通过一项激进的立法——对使用AI替代超过一定比例员工的企业征收”自动化附加税”。这不会是孤例。当某个行业的自动化速度超过从业者的适应速度时，从业者群体通过政治手段试图减缓替代进程，反自动化立法、新卢德运动、职业保护主义很可能在欧洲多国同时出现。“机器人税”的税款用于被替代工人的再培训和社会保障。

**效果评估：** 短期看，这些措施确实减缓了自动化速度并为过渡提供了资金。在这一推演情景中，中期看，它们可能造成竞争力损失——类似的”自动化附加税”实施数年后，制造业投资可能明显下降，部分企业将生产转移到税收更友好的国家。长期看，大多数限制性措施被迫放宽或废除——因为在全球化经济中，单一国家的反自动化政策等于单方面提高自己的生产成本。

**因果链条：** 限制性立法的宿命几乎在每个实施国都循着同一轨迹展开。自动化冲击从业者收入，政治压力催生限制性立法；立法落地后企业竞争力下滑，经济增长放缓拖累财政收入，社会保障因此捉襟见肘，而保障资金不足又反过来动摇了立法的社会基础。整个周期大约5–8年，终点几乎都是被迫放宽。——当然现实比这复杂，每个国家都有自己的变量，但主线确实如此。

劳动者抵抗的失败不是因为从业者的诉求不正当，而是因为他们选择了错误的战场——试图在技术层面阻止替代，而不是在分配层面重新定义规则。正确的问题不是”如何保住旧工作”，而是”如何确保自动化的收益被合理分配”。

### 第二种：文化战争

AI对内容产业的冲击（第八章）不仅是经济问题，更是文化身份问题。当AI可以生产出在形式上与人类作品无法区分的内容时，“创作”作为身份标记的独特性受到了根本威胁。艺术家、作家、音乐人将自己定义为”创造者”，这个身份的意义远超经济收入——而现在这个身份的根基被动摇了。

按本书推演，2029–2033年间，反AI创作运动、“真人认证”标签、人文主义回潮可能在全球范围内出现。在本书的推演中，好莱坞编剧可能在2028年前后再次发起集体行动（2023年集体行动的续

集），核心诉求从”限制AI在编剧中的使用”升级为”AI生成的内容不得在未经人类创作者实质性修改的情况下使用”。日本漫画界也可能发起类似”手绘认证”的运动——出版社自愿在人类手绘的漫画上标注认证标签，以区别于AI生成的作品。欧洲多国博物馆和画廊可能开始拒绝展出AI生成的艺术作品，或将其放在单独的”数字艺术”区域。

**认知翻转：** 这些运动的深层逻辑不是”反技术”，参与者中很多人本身就使用AI工具。它们的逻辑是”保卫意义”——当创作不再是人类的专属能力时，人类需要新的方式来定义自己与作品的关系。“手绘认证”运动本质上不是在说”AI画得不好”，而是在说”这幅画之所以有意义，不是因为它画得好，而是因为有一个人花了三个月把它画出来。这个时间本身是有意义的”。

这是一种非常深刻的价值主张，它挑战了效率至上的市场逻辑。市场说”同样的产出，用更少的时间生产就是更好的”。文化战争说”不，时间本身可以是价值的一部分”。

## 第三种：代际张力

**表现形式：** 年轻人的”躺平2.0”、中年人的”末班车焦虑”、老年人的”被数字抛弃感”。

**核心动态：** AI的冲击不是均匀地分布在所有年龄段上的。

**年轻一代（2000年后出生）：** 他们是第一批在AI环境中进入职场的人。他们的问题不是”AI取代了我的工作”——因为他们中的很多人从未体验过”没有AI的工作”是什么样的。他们的问题是：在一个AI可以完成大部分标准化知识工作的世界里，“职业生涯”这个概念还意味着什么？到2030年前后，“躺平2.0”——不同于2021年那种消极的退出，而是一种更积极的”我不打算把生命投入到会被AI替代的工作中，我要找到只有人能做的事”——可能成为年轻人中的主流态度。

**中年一代（1975–1995年出生）：** 他们是冲击最大的群体。他们的专业技能投资已经沉没（十到二十年的职业积累），退休还遥遥无期（还有十五到三十年），重新学习的成本极高（既要养家又要转型）。他们的焦虑是”末班车焦虑”——不是怕失业，而是怕自己不在那20%能享受AI赋能的精英行列中。这种焦虑在2029–2035年间很可能转化为强烈的社会诉求——“保护中产阶级”可能成为全球选举中最有效的政治口号之一。

**老年一代（1975年前出生）：** 他们面临的不是工作替代（很多已退休或接近退休），而是”被数字抛弃”。当越来越多的公共服务——医疗预约、银行业务、政府办事、社交互动——转移到AI驱动的数字界面上时，不会或不愿使用这些界面的老年人被系统性地排斥在外。这不是个体的技术障碍，而是一种新的社会排斥形式。

间奏一中那对坐在餐桌两端的母女——母亲打不通社保中心的人工电话，女儿在办公室被AI挤进角落——正是这种代际张力的缩影：两代人各自困在自己的困境里，甚至没有意识到她们撞上的是同一堵墙的两面。

**代际张力的政策表达：** 在税收政策上（年轻人主张对AI资本征税来资助全民基本收入，老年人主张保护现有的退休金和社保体系）、在教育政策上（应该继续投资传统教育还是彻底重塑？）、在AI监管

政策上（年轻人倾向于拥抱AI但要求公平分配，老年人倾向于限制AI的渗透速度）——三代人的利益和价值观在AI问题上形成了复杂的三角博弈。

### 第四种：区域分化

**表现形式：** AI产业集中区域vs.被抛在身后的区域之间的经济和政治对立。

**核心动态：** AI产业的地理集中度远高于传统产业。按当前趋势，到2030年前后，全球80%以上的AI相关就业和收入可能集中在不到二十个城市区域——旧金山湾区、西雅图、纽约、北京、上海、深圳、伦敦、特拉维夫等。这些”AI中心”的人均收入和经济增速将远超全国平均水平，而其他地区——特别是那些传统产业被AI替代但没有新产业填补的地区——可能经历经济萎缩。

这种区域分化不是新现象——工业化时期也有类似的城乡差距和区域不平衡。但AI的区域集中比工业化更极端。工厂至少需要土地、水源、交通便利和劳动力——这些约束迫使工业在一定程度上分散。AI产业的核心生产要素是人才和算力——人才可以远程协作，算力可以通过网络调用——地理约束远弱于传统产业。

**治理挑战：** “被抛在身后”的区域对治理体系的适应性提出了更高要求。这不是理论推测——2016年的美国总统大选和英国脱欧公投已经展示了区域分化如何转化为政治能量。AI加剧区域不平衡后，社会结构变化对治理体系的适应性提出新要求。“科技精英vs.普通人”的叙事在2030年代很可能成为全球范围内最有动员力的舆论框架之一。

---

## 中国的调适路径——国有经济作为减震器

四种社会张力形态在不同制度土壤上会长出不同形态。中国提供了一个独特的观察案例：国有经济体系充当了AI冲击的大规模减震器。

截至2020年代中期的数据：国企直接雇员约三千万，加上事业单位约四千五百万，合计约七千五百万人。加上上下游供应链间接依赖国有经济的就业人口，估算在一亿五千万到两亿之间。这两亿人的就业不完全由市场信号决定，政府有能力也有意愿通过行政手段延缓冲击速度——某种程度上是一次”美丽的去就业化”，类似宏观经济学中”美丽的去杠杆”的逻辑，把市场化速度的冲击拉长到社会能够承受的节奏。

减震的代价是效率损失。当一家国有钢铁厂的财务部门已经可以用AI将二十个人的工作量压缩到五个人时，“不裁人”意味着多出来的十五个人要么”转岗培训”（培训什么岗位往往不清楚），要么提前退休。成本没有消失，只是从企业的损益表转移到了政府的资产负债表上。

但这里有一个关键追问：**国有经济的缓冲到底是反脆弱的，还是延迟脆弱的？**

按本书的推演，答案大约是六四开——六成反脆弱，四成延迟脆弱。那些确实利用缓冲期完成转型的人，可能获得民营经济中被市场速度甩下的同龄人所没有的机会。那些只是在缓冲期中等待的人，可

能在后续国企深化改革中面临更艰难的处境。关键的结构性洞察是：**缓冲的价值取决于缓冲期内做了什么。时间本身不解决问题——时间只是解决问题的空间。**

更深层的影响在于：国有经济的缓冲为政府推进AI产业政策争取了政治空间。当失业率没有急剧飙升时，政府可以更从容地推进AI基础设施建设，而不必把主要精力放在救火式的社会救济上。

**压力测试——四个需要警惕的脆弱点。** 第一，垂直化AI生态在当前技术范式下有效，但如果下一次突破需要跨领域融合，“一百个专用模型”的格局可能成为障碍。第二，举国体制擅长调动物质资源，但对人才的调动远不如市场机制灵活——顶级AI研究者的流动受制于人事管理体系和薪酬结构。第三，AI基础设施投资加上国企缓冲的隐性财政成本加上地方AI产业园的举债——这些数字加在一起，是一笔需要用未来增长来偿还的账。第四，也是最根本的——如果缓冲争取的时间未被充分利用来推动结构性调整，减震器的效果就会递减，积累的问题可能在缓冲不可持续时一次性释放。

在本书的推演中，2035年到2040年间的发展可能部分验证这个担忧。当国企第二轮AI深度改革启动时，被延迟的调整可能集中释放，部分地区的社会压力甚至可能超过此前民营经济的冲击峰值。缓冲争取的时间，或许三分之二能被有效利用，但剩下的三分之一可能变成更痛苦的延迟结算。

---

## 三、社会协商——制度创新的必要信息输入

以上四种社会张力形态，在常规政治分析中通常被视为”问题”——需要被管理、被缓和、被解决的负面现象。但反脆弱理论提供了一个完全不同的视角：

**多元利益的公开表达和协商，是社会发现自身边界、催化制度创新的核心机制。**

每一项重要的社会制度都是利益博弈与制度协商的产物。这不是修辞——历史记录可以精确证明：

**八小时工作制** 的争取可以追溯到1886年5月1日——美国工人为争取八小时工作制而发起全国总罢工，随后在芝加哥引发了干草市场冲突（Haymarket affair）。这场运动成为五一国际劳动节的起源。从那时起，又经过数十年的工人权益争取，八小时工作制才逐步成为法律。在此之前，十二到十六小时工作制是”正常的”。

**社会保障制度** 是1930年代大萧条中大规模失业和社会压力的直接产物。罗斯福签署《社会保障法》（1935年）时，驱动他的与其说是善心，不如说是对社会失序的深切忧虑。

公共教育、反垄断法、劳动法中的几乎每一条重要条款——最低工资、加班费、工伤赔偿、集体谈判权——背后都是同样的逻辑：社会压力催生制度回应，没有一项是主动给的。

模式清晰了：**技术变革创造效率，效率的收益分配不均引发社会利益表达，利益表达推动制度协商与创新，最终形成新的社会契约。** 这不是某个时代的特殊现象，这是技术–社会互动的基本动力学。

AI时代的制度转型同样需要经历充分的利益表达和协商过程——绕过这一环节的转型，在历史上从未成功过。

---

## 四、1930年代的教训

在所有历史类比中，1930年代大萧条是最有指导意义的。AI不太可能引发同等规模的经济崩溃，但大萧条展示了一种关键机制：**社会在深层压力下如何生成制度创新**。

**大萧条前的正统：** 1929年之前，经济正统是自由放任——政府不应干预市场，市场会自我调节，失业是"自愿的"（如果你愿意接受更低的工资，你就能找到工作）。社会安全网几乎不存在——没有失业保险，没有社会保障，没有最低工资。

**大萧条的冲击：** 1929–1933年，美国GDP下降约30%，失业率从3%飙升到25%。银行倒闭潮摧毁了普通家庭的储蓄。信贷冻结让企业无法运营。全社会陷入恶性循环：失业→消费下降→企业倒闭→更多失业。

**制度创新的爆发：** 罗斯福新政（1933–1938）在五年内创造了一系列此前"不可想象"的制度：社会保障制度（Social Security Act, 1935）、全国劳动关系法（Wagner Act, 1935，赋予工人集体谈判权）、公平劳动标准法（1938，确立最低工资和最高工时）、联邦存款保险公司（FDIC，防止银行挤兑）、证券交易委员会（SEC，监管金融市场）。

**这些制度在大萧条之前都被视为"激进的"甚至"社会主义的"。** 大萧条改变了"什么是可想象的"——极端的社会压力打破了旧正统的束缚，为新的制度可能性打开了空间。

这里浮现出一个被反复验证的模式：**社会压力积聚到临界点时，各方利益充分表达和协商，制度创新随之爆发。** 人类社会的制度创新能力不是持续的、线性的，而是爆发式的。长期的惰性被强烈的社会需求打断，然后在短时间内产生大量的新制度。

**AI时代的类比：** 2026年到2040年间积累的社会压力——劳动替代、收入两极分化、认知生产资料的集中、代际和区域张力——最终可能催生一批今天看来"不可想象"的制度创新。这些创新的具体形态我们在第三部分会详细讨论，但可以先列出几个方向：

某种形式的全民基本收入或全民基本服务——当"以劳动换取收入"的社会契约对越来越多人不再成立时，收入的分配机制需要与劳动脱钩。

认知基础设施的公共化或准公共化——类似于电力从私人垄断走向公用事业监管的历史路径。

新的所有权模型——数据产权、AI模型的收益分享、“认知红利”等概念。

新的教育体系——从"训练认知劳动者"转向"培养AI时代的人"。

新的社会身份框架——当”我做什么工作”不再能定义”我是谁”时，人需要新的方式来获得社会身份和意义感。

---

## 五、博弈中的重建种子

让我们回到本章的核心论点，也是整个第二部分的收束。

第二部分（第六到第十章）讲述的是旧前提如何在现实中被拆除：

第六章：认知从人类能力变成公共基础设施。 第七章：劳动在五个阶段中逐步被替代。 第八章：内容从稀缺作品变成无处不在的空气。 第九章：认知生产资料高度集中在少数公司和国家手中。 第十章（本章）：社会在多条战线上产生张力与调适需求。

叙事在这里达到了不安的峰值。旧秩序正在解体，新秩序尚未形成——我们正处在波兰尼所说的”大转型”之中。

但——这是本章最重要的”但”——**利益博弈的过程本身包含了重建的种子。**

这不是安慰性的话语。它是从历史规律中得出的严肃结论。

从圈地运动到工业化再到大萧条，每一次社会结构面临深刻调整的时刻，恰恰是制度创新最活跃的时刻——济贫法、劳动法、社会保障、新政，都是在阵痛中锻造出来的。

换一个角度看：**社会是反脆弱的。不是因为它不会被冲击，而是因为压力本身会激活它的自我调适和创新机制。** 前提是压力不会大到直接毁灭系统（核战争、生态崩溃）。AI技术变革虽然深刻，但不是毁灭性的，它是一种渐进的、可管理的（虽然管理难度很大的）系统性压力。

从更长的周期看：**文明的演进从来不是平滑的。它是一系列从稳定到调适、从重构到新稳定的循环。** 我们正处在”调适”阶段——旧的社会契约正在失效，新的社会契约正在孕育中。这个阶段会持续十到二十年。它不会是舒适的。但多元利益的充分表达和协商，正是社会找到新均衡的必经之路。

最后一个洞察：**利益协商与制度反馈是秩序的生成机制——没有充分的利益表达就难以形成真正稳固的新秩序。** 八小时工作制不是在真空中被”设计”出来的。社会保障不是在学术论文中被”论证”出来的。它们是在劳动者的持续诉求表达、社会各方的反复协商中被”锻造”出来的。

AI时代的新制度——无论最终采取什么形态——同样需要通过协商来达成、通过试错来发现、通过博弈来确定。这个过程已经开始了。

---

## 六、进入第三部分

前两部分完成了诊断：旧前提被逐一拆除，社会张力在多条战线上展开。第三部分转向重建——新的社会契约、分配机制、教育体系和治理结构。

这些问题不是在真空中回答的。它们的答案正在第二部分描述的利益博弈与制度协商中被锻造。

就像1930年代的社会保障制度是在大萧条的废墟上建立的，AI时代的新制度将在旧制度的裂缝中生长。

但在此之前，我们需要理解一件事：社会协商的结果不是预定的。历史不是线性的、不是必然进步的。同样的社会压力既可能催生罗斯福新政（制度创新的典范），也可能催生极端主义（制度溃败的深渊）。1930年代同时产生了两种回应——美国选择了前者，德国和意大利选择了后者。

AI时代的社会调适同样可能走向不同的方向。方向取决于什么？取决于社会能否及时完成制度创新——给出一个大多数人能接受的新社会契约。这个时间窗口不是无限的。

第三部分的任务就是：在这个窗口关闭之前，探索新社会契约的可能性。

---

*到这里，诊断基本完成了。坦白说，写诊断比写药方容易得多——告诉你什么会坏，总比告诉你怎么修要有把握。从下一章开始，我们进入更不确定的领地：意义、信仰、存在。我不敢说自己有答案。但我相信这些问题值得被认真地提出来——哪怕提出之后，我们一起站在问题面前沉默很久。*

抵抗者和适应者的区别不在于谁更聪明，而在于谁先承认旧地图作废了。

# 间奏二：裂缝中的生活

* * *

2032年，深秋。一个南方城市的老社区，六层楼的步梯房，外墙的米黄色涂料已经斑驳了。楼下有一排香樟树，叶子四季不落，把路灯的光切成碎片洒在地上。这个社区建于2005年，当时被叫做”白领公寓”——一个现在听起来像出土文物一样的名字。

住在这里的人，六年前大多有着相似的生活。朝九晚六，周末偶尔加班，房贷还到2040年代，孩子在附近的学校上学。他们是中产叙事最忠实的信徒——不是因为这个叙事多么动人，而是因为他们的房贷、学区、职级全部按照这个叙事搭建。

现在地板在响了。而且每个人听到的声音不一样。

* * *

三楼东户住着一个三十四岁的女人。她以前是一家广告公司的文案总监，管着一个八人团队。2030年，公司把团队砍到三个人加一套AI内容生成系统。她留下来了——作为”人机协作组”的负责人。新的工作内容是：设定AI的创作方向，审核AI的产出，偶尔在AI写不出”人味儿”的地方补几句。

她的收入没怎么变。Title升了半级。从外面看，她是”成功转型”的典范——公司内部通讯里还拿她当案例讲过。但她自己知道那是什么感觉。她用了一个很精确的比喻来形容：就像你从厨师变成了试菜的。菜还是那些菜，你还坐在厨房里，但你不再掌勺了。你只是尝一口，说”咸了”或者”再加点葱”。

最难受的不是工作内容变了，而是一种丧失。她想了很久才找到那个词：“手感”。以前写文案，一个句子改了七遍，第八遍突然对了，那种从指尖传上来的确定感，现在没有了。AI的初稿永远是七十五分——不差，但也不亮。她的工作就是把七十五分改到八十五分。但从七十五到八十五的距离，和从零到七十五的距离，感觉完全不同。前者是创造，后者是校准。

她没跟任何人说过这些。因为说了也没用。周围的人会说：“你已经很幸运了。”她知道这是事实。这让她更不好意思抱怨。于是她把那种丧失折叠起来，收进日常生活的缝隙里，假装它不存在。

晚上她会在阳台上坐一会儿，抽一支烟——这个习惯是去年才有的。她看着楼下的香樟树，什么也不想。或者说，她在想的那个东西没有形状，像一团雾。如果硬要给它一个问句的话，大概是：“这就

是以后的日子了吗？”

* * *

四楼西户住着一对五十多岁的夫妻。丈夫原来是一家商业银行的支行副行长，2031年内退了。不是被裁——银行给了他一个体面的选择：提前退休，拿百分之七十的工资，加一笔一次性补偿。他接受了。不是因为想退，是因为他能感觉到自己在那个系统里已经多余了。

AI没有取代他的职位，支行副行长这个头衔还在。取代的是他的职能。信贷审批有AI风控模型，客户关系维护有AI客服系统，合规检查有AI审计工具。他以前每天签二十几份审批文件，现在系统自动完成，他只需要在最后一页盖个章。他不是被辞退了，他是被架空了。“盖章机器”——他在一次喝酒时这么形容自己，然后笑了笑，笑得很勉强。

内退之后他试过几件事。先是每天早上去公园走路，一个月后觉得无聊。然后报了一个书法班，上了两个月，发现自己没有耐心。又买了一套钓鱼装备，去了三次水库，第四次装备就在阳台上落灰了。他妻子说他”像个没头苍蝇”。他不反驳，因为他觉得这个比喻是准确的。

他的问题不是没钱，退休工资加存款，够花。问题是一天有十六个小时是醒着的，而他找不到一件值得用这十六个小时去做的事。上班的时候他从来不觉得工作有意义（谁会觉得签审批文件有意义？）但上班至少提供了一个结构：几点起床、几点到岗、几点开会、几点下班。结构消失之后，时间变成了一种需要被消耗的东西，而不是被使用的东西。

他妻子适应得比他好。她是中学英语老师，还没退休，但课时已经减了一半——学校引入了AI口语陪练系统，学生每周有三节”AI英语课”，她只需要上剩下的两节。她把多出来的时间用来带孙辈、跳广场舞、和老同事在微信群里聊天。她没有他那种”意义焦虑”。他有时候嫉妒她的这种能力——不追问”为什么”就能安然地过日子。

他偶尔会在深夜打开电脑，看一些财经新闻。那些新闻里充满了他越来越看不懂的词：“Agent经济”、“认知基础设施”、“人机协作指数”。他曾经是一个每天看三份金融报纸的人。现在那些报纸还在，但报纸里说的事情已经不属于他的世界了。他感觉自己像一个被请出柜台的人——银行还在营业，只是窗口上多了一行字：“此窗暂停服务”。

* * *

二楼中户住着一个二十七岁的年轻人。他大学学的是计算机，毕业后没有去大厂。不是没有机会——2028年校招的时候他拿到了两个offer。他都拒了。

他的选择在当时让所有人困惑。父母觉得他疯了，同学觉得他矫情。但他有自己的一套逻辑——虽然他表达不太清楚。他觉得去大厂写代码这件事的"保质期"太短了。他看得到那些学长的轨迹：进去的时候意气风发，三年后变成螺丝钉，五年后开始焦虑下一次"优化"什么时候轮到自己。AI不是让程序员失业了——而是让"普通程序员"这个角色越来越像一个临时工种。他不想在一个正在塌缩的赛道上起跑。

他现在做的事情很杂。帮几个小公司做AI应用的定制部署——不是写模型，是帮不懂技术的老板把现成的AI工具接进他们的业务流程里。他管自己叫"AI水管工"：不造水龙头，只负责把水管接到你家。收入不稳定，好的月份能赚两万多，差的月份几千块。但他不焦虑。或者说，他的焦虑和楼上那对夫妻的焦虑不在同一个频道上。

他焦虑的不是"够不够用"，而是"够不够快"。他觉得世界在加速，而他必须跑在变化前面。每天晚上他花两三个小时学新东西——不是那种系统性的学习，而是东一榔头西一棒槌地跟进各种AI工具的更新。他的浏览器收藏夹里有三百多个链接，分了十几个文件夹，大部分他收藏之后就没再打开过。但收藏的动作本身让他觉得自己"没掉队"。

他和楼上那个退休的银行副行长有时候在楼道里碰见。两个人点个头，说一句"吃了吗"，然后各走各的。他们不知道彼此在想什么。也用不着知道。各有各的麻烦。

* * *

小区门口有一家开了十二年的杂货店。老板是一对河南夫妻，五十出头。他们的店没有被任何AI冲击——没有什么AI能替代一个人走进杂货店说"来包烟、拿瓶酱油、顺便帮我收个快递"这件事。至少现在还没有。

他们是这栋楼里变化最小的人。每天早上六点开门，晚上十一点关门。进货、理货、收钱、聊天。来买东西的人少了一些——很多年轻人直接网上下单，三十分钟送到家。但老住户还是习惯来这里。不为了买东西，为了说两句话。

那个退休的银行副行长几乎每天下午都会来坐一会儿，买一瓶矿泉水，其实是找人说话。杂货店老板不懂什么AI不AI的，但他是一个好的倾听者——他什么都听，什么都不评价，偶尔"嗯"一声，偶尔递根烟。这种不带任何信息量的交流，在一个被信息塞满的时代里，反而变成了稀缺品。

老板娘有时候站在收银台后面，看着小区里的人来来去去，会冒出一句：“这几年人的脸色都不太一样了。”她说不出哪里不一样。但她在这个小区的门口站了十二年，她见过那些脸从容的时候，也看到它们开始收紧。

她的观察比任何社会学调查都更精确。只是没有人把它写成论文。

* * *

夜深了。六层楼的灯一盏一盏灭掉。

三楼东户的灯最后灭——那个文案总监又在阳台上坐了很久。四楼的灯灭得早，但透过窗帘能看到电视的蓝光闪烁，那个退休副行长大概又在看他看不懂的财经节目。二楼中户的灯最亮，年轻人的两个显示器亮到凌晨一点，屏幕上是一个他明天可能就用不上的AI工具的教程。

杂货店的卷帘门拉下来了，老板在门口抽最后一根烟。电视里的深夜节目插了一条公益短片，画面上是一个三十多岁的女人坐在咖啡店窗边对着笔记本电脑，旁白说：“35+女性再就业支持计划，让经验不被辜负。”画面一闪就过去了，没有人注意到。远处传来无人配送车的低频嗡鸣——它在空无一人的街道上滑过去，车顶的蓝色指示灯在夜色里一闪一闪的，像一只不眨眼的眼睛。

没有人看到它。

或者说，每个人都看到了。只是在2032年的秋天，它已经不算新鲜事了。就像你不会特意去注意一盏路灯什么时候亮的一样。变化不是一个事件。你没法指着日历上某一天说”就是从这里开始不一样的”。

这栋楼里的每个人都在裂缝中生活。不是某天突然裂开的那种——是墙角渗水，你注意到的时候已经渗了很久。

* * *

# 第十一章 意义的多元重建——不止存在主义

> 你以为”失去工作就失去意义”是人性必然。但这是工业文明的文化特产。工业社会之前几千年，意义来自宗教、家庭、社区。工业社会之后——我们正在寻找答案。

---

*2033年，秋天。*

*她离开咨询公司已经两年了。不是被裁的——她自己走的，在AI接手了竞品分析流程的第三个月。最后那段时间，她的工作变成了检查AI的输出报告有没有格式错误。她受不了的不是无聊，是那种感觉：自己变成了一道多余的审批环节。*

*现在她每周去社区花园做三次志愿者。翻土、浇水、给新来的老人讲怎么分辨杂草和菜苗。上周一个六十多岁的阿姨跟她说，“小姑娘，你教得比那个手机app清楚多了。”她笑了，那天心情意外地好。*

*晚上回到出租屋，好心情慢慢退了。她打开电脑，看到前同事发的朋友圈：新公司，新头衔，年终奖。她关上电脑坐了一会儿。花园里的事情让她觉得自己有用，但”有用”和”有价值”之间似乎隔着什么，她说不清。*

---

前面的章节拆解了旧秩序的结构。从这里开始，问题变了——不是”什么会塌”，而是”塌了之后，人靠什么站着”。

“意义”这个词在本书中反复出现，但含义并不总是相同。最关键的区分是两层：功能意义——工作和社会角色赋予的身份感与日常结构，前五章已经详细讨论了它的瓦解；存在意义——“我为什么在这里”这个更深层的追问，是本章的核心。至于信仰层面的终极意义和人际连接中的关系意义，分别留给后续章节。本章追问的是：当工作不再定义你、日常结构和社会身份同时塌陷时，“我是谁”这个问题落到了哪里？

---

## 一、技术座架

2033年春天，一位在上海某三甲医院工作了十八年的影像科主任，在一次科室会议上被告知：AI辅助诊断系统的准确率已经超过了科室所有医生的平均水平。她回到办公室，关上门，在电脑前坐了很

久。她不是在担心失业——她的位子暂时还稳。她在想一个她从没想过的问题：如果一台机器比我更擅长读片，那我这十八年练出来的”眼力”算什么？

这个问题比失业更深。它不是经济问题，而是关于”我是谁”的问题。

1953年，马丁·海德格尔在慕尼黑发表了一场题为《技术的追问》的演讲。在晶体管刚刚发明、计算机还占据整个房间的年代，这位哲学家说了一些在当时极为抽象、七十年后却精准得令人不安的话。

海德格尔提出了一个概念：Gestell，通常译为”座架”或”集置”。他的意思是：现代技术不只是一套工具，它是一种”揭示方式”——它决定了我们如何理解世界上的一切事物，包括我们自己。在技术座架的框定下，河流不再是河流，而是”水力资源”；森林不再是森林，而是”木材储备”；土地不再是土地，而是”可开发地块”。一切存在者都被揭示为”持存物”（Bestand）——随时等待被调用、被优化、被最大化利用的资源。

AI改变了这个概念的实践意义。AI把海德格尔的抽象概念变成了日常体验。

当一个广告公司用AI系统替代创意总监时，它做的不只是用便宜的工具替换贵的员工。它做的是重新定义了”创意”——创意不再是一个需要灵感、经验和人格印记的创造过程，而是一个可以被参数化、批量生成、A/B测试优化的生产流程。在这个框架下，人不是”被替代了”——而是”被重新揭示了”：从一个创造者，变成了一个可被替代的信息处理节点。

比失业更深层的问题在这里。失业是经济问题——你没有收入了，需要找到新的收入来源。但”被重新揭示”是存在论层面的问题——不是你的工作被拿走了，而是你对自己是谁的理解被改写了。

按趋势推演，这种”被量化”的体验可能在2030年代迅速蔓延到几乎所有知识工作领域。医生的诊断能力可以用准确率百分比来衡量，然后发现AI的百分比更高。同样的逻辑迅速蔓延到律师、教师和几乎所有知识工作领域。每一个职业都在被翻译成一组可比较的参数。

每一次衡量都不是错的。每一次比较都有坚实的数据支撑。但所有这些衡量加在一起，产生了一个后果：人类自身被框定为”可优化的资源”。你不是一个有独特经历、情感和人格的存在者，你是一组可以被量化的能力参数——而当这些参数不如机器时，你就是”低效的”。

可以想象，某位哲学学者会写出一篇引发广泛讨论的论文，核心论点很简单：AI时代的意义危机不是因为人类”做不了”什么了，人类能做的事情其实并没有减少。危机在于，技术座架让我们习惯了用”功能性”来衡量一切，包括自身。当机器在功能性维度上超过人类时，人类在自己建构的价值坐标系中找不到位置了。

这不是技术的问题。这是我们理解技术、理解自身的方式的问题。

但不同文化对这个问题的感受强度和回应方式截然不同。技术座架的压迫感在不同的文明土壤中生长出不同的焦虑——有些文化比其他文化更脆弱，有些文化则展现出出人意料的韧性。

这里需要再往深挖一层。技术座架的真正边界在哪里？

设想一个场景：一个三十五岁的建筑师，在AI生成设计方案的效率已经远超人类之后，面临一个选择——是转型做AI设计的审核员（收入稳定、前景清晰），还是去一个偏远山区做低收入的乡村改造项目（她在大学时代就想做，一直没做）。她打开AI助手，问了这个问题。AI给了她一份极其详尽的分析：两条路径的预期收入曲线、职业发展概率、行业趋势、甚至心理健康风险评估。每一项数据都无可挑剔。但她看完之后，发现自己仍然不知道该怎么选。

AI回答了所有关于”怎么做效率最高”的问题，却没有触及那个真正的问题：“这件事值不值得我用十年生命去做？”

这不是AI的技术局限。未来的AI可能比今天聪明一百倍，它仍然回答不了这个问题——不是因为它不够智能，而是因为”意义”本身的结构决定了它不能被外部系统代劳。意义不是一个可以计算出来的最优解。它必须被体验者自己赋予。AI可以告诉你所有选项的”形式”——收益、风险、概率——但它给不了任何选项以”重量”。那个重量只能来自你自己的生命经验：你失去过什么，你害怕什么，你在深夜无法入睡时反复想的是什么。

海德格尔用”此在”（Dasein）这个概念来描述人的独特存在方式。此在的核心特征不是理性，不是语言，而是”向死而生”——人是唯一知道自己终将死去的存在者。正是因为知道时间有限，每一个选择才有重量；正是因为不可逆，“值不值得”这个问题才有意义。AI没有死亡，所以它的”回答”没有时间的紧迫感，没有不可逆的代价，没有”如果选错了我就再也回不来了”的颤栗。它可以模拟这种颤栗的语言形式，但那只是形式。

但更令人不安的也许不是AI”变得像人”，而是AI以完全不同的方式达到了同样的功能效果。AI不需要理解意义就能模拟出有意义的回答；不需要共情就能提供令人感到被理解的反馈；不需要经验就能给出看似来自经验的建议。这意味着”像人一样”根本不是功能性替代的必要条件。一个从未活过的系统，可以在功能层面替代一个活过几十年的人——而这种替代之所以令人不安，恰恰是因为它表明，我们以为只有”活过”才能做到的事，原来不需要。但”不需要”和”等价”是两回事。功能相同不等于意义相同——AI可以输出一模一样的安慰话语，但说出那句话的”人”背后没有失眠的夜晚、没有相似的伤痛、没有”我懂，因为我也经历过”的重量。这正是本章的核心命题：功能可以被复制，意义不能。

这是技术座架最深的盲区：它能把一切翻译成可比较的参数，唯独翻译不了”这对我意味着什么”。因为”我”不是一个参数。“我”是一个会死的、正在老去的、背负着独一无二的记忆和遗憾的存在者。

---

## 二、认知翻转——工业文明的文化特产

在展开不同文化的回应之前，需要回到第二章已经论证过的核心判断：“工作赋予人生意义”不是人性必然，而是工业文明的文化特产。新教伦理将劳动提升为天职，工厂体系将生计与工资捆绑，民族国家将”工作的公民”塑造为理想国民——三股力量在两百五十年里把”工作=意义”写进了全球主流文化

的操作系统。在此之前的几千年里，从《创世记》到古希腊到儒家传统，劳动从未占据过意义体系的核心位置。

这个认知翻转的实践意义是：如果”工作=意义”是一种文化编程，那它的崩塌——虽然痛苦——就不是文明的终点。文化编程是可以重写的。

问题是：用什么来重写？

---

## 三、三种文化的不同回答

### 中国：关系网络中的意义

可以合理推演：在东亚儒家文化圈中，如果有学者对经历”职业身份中断”的知识工作者做大规模追踪调查，会发现一个有趣的模式——将”职业成就”列为首要意义来源的比例大幅下降，但受访者并没有普遍陷入存在主义虚无感。取而代之的首要意义来源可能是”家庭责任”和”社会关系网络”。

这背后的文化逻辑是：在西方个人主义框架中，“我是谁”主要锚定在个人成就上——我做了什么，我能做什么。当AI摧毁了这个锚点，个体就陷入存在性漂浮。但在东亚的关系性文化框架中，“我是谁”从来不只是关于个人。它同时是”我是谁的孩子”“我是谁的父母”“我在哪个圈子里”。这些关系性的身份锚点不依赖于职业成就，因此在AI冲击面前可能展现出更强的韧性。

这个发现需要谨慎解读。它不意味着中国人不在乎工作——恰恰相反。但它意味着，当职业这个意义支柱被动摇时，中国文化中存在其他支柱可以部分承接。一个失去了创意总监身份的人，在存在主义框架中是一个被抛入虚无的孤独个体；但在中国的关系网络中，她同时还是父母的女儿、丈夫的妻子、孩子的母亲、朋友圈里的”那个谁谁”。这些身份不因职业变动而消失。

更有意思的是”修身齐家治国平天下”这个儒家意义层级在AI时代的意外回归。这套思想在20世纪经历了反复批判——它被认为封建、等级制、压抑个体。但当工业时代的”个人成就”叙事崩塌时，一些学者开始重新审视这套层级结构的功能性价值：它提供了一个不依赖于职业的、从个人修养到社会参与的完整意义梯度。你不需要是一个成功的职业人士才能在这个框架中找到位置——你可以从”修身”开始，通过自我完善本身获得意义。

在本书的推演中，“新修身运动”可能成为中国社交媒体上的热门话题。它不是复古主义——参与者不是在读《论语》行古礼。更像是后现代的自我文化实践：冥想、书法、茶道、社区服务、亲密关系经营。共同点是：这些活动的价值不来自市场定价，不能被AI优化，也不需要和任何人竞争。

当然，中国文化也有自己的脆弱性。关系网络在提供意义的同时也施加压力——“面子”文化让职业身份的丧失在社交场景中格外痛苦。但总体而言，中国社会在AI引发的意义危机面前的表现，好于很多西方预测者的悲观判断。

## 日本：万物有灵与社畜矛盾

日本提供了一个充满张力的案例。

一面是文化亲和性。万物有灵的神道传统让日本人对非人类实体怀有一种其他文化少见的平等态度——石头、河流、老旧器具都可以拥有灵魂，那AI为什么不可以？日本人对机器人的态度更接近温和的亲近（《铁臂阿童木》《哆啦A梦》），而非好莱坞式的恐惧（《终结者》）。按既有的跨文化调查趋势推演，日本受访者对AI持”基本信任”态度的比例可能接近七成，远高于美国和德国。

另一面是极端的脆弱性：劳动与身份的深度捆绑。“社畜”这个自嘲式词汇背后是严肃的文化现象——在日本战后经济体制中，公司不只是工作场所，而是个体社会身份的核心载体。终身雇佣制、年功序列制、公司内部的社交网络——“你在哪家公司工作”几乎等同于”你是谁”。

这种捆绑在AI时代变成了双刃剑的短那一面。按趋势推演，如果日本政府对经历”职务内容根本性变更”的正式员工做追踪调查，可能会发现出现中度以上抑郁症状的比例远高于美国同类群体。原因在于关键的文化差异：在美国，工作变化主要被体验为经济事件（收入下降、前景不确定）。在日本，同样的事件更可能被体验为社会性死亡的前兆——你不再是”三菱的山田”了，而”只是山田”在日本社会中几乎没有意义锚点。

然而日本文化的另一面——对技艺（craftsmanship）的极端尊重——可能在意外方向上提供出路。可以合理推演，“新职人运动”将会兴起：大量前白领转向手工制造、传统工艺、在地农业和匠人学徒。这不是经济理性的选择——收入远低于之前。但在日本文化中，“职人”拥有一种独立于公司体制的、来自技艺本身的尊严。一个做了三十年寿司的师傅，社会地位不来自在哪家公司上班，而来自他和素材之间几十年如一日的对话。

AI可以做出营养成分和口感参数都与顶级寿司相当的食物。但在日本文化中，那不叫寿司。寿司的意义不在于结果，而在于过程——在于那个人用几十年训练自己手感、眼力和判断力的过程本身。这是一个微妙但重要的区分：在效率至上的技术座架中，过程只是达成结果的手段；但在职人文化中，过程本身就是目的。

## 全球南方：跳跃式发展——我们从未掉入的陷阱

来自发展中国家的经济学家观察到一个被忽略的事实，有学者将其概括为”我们从未掉入的身份陷阱”。核心论点非常简洁。

这些学者指出：西方和东亚的”意义危机”共享一个前提假设——人类的身份和尊严主要通过有偿劳动来实现。这个假设是工业化的文化产物，而非人类天性。在工业化充分完成的社会中，几乎所有人都被纳入雇佣劳动体系，“你做什么工作”成了”你是谁”的最重要回答。当AI摧毁了这个回答，危机就来了。

但在撒哈拉以南非洲大部分地区，工业化从未充分完成。2030年非洲大陆正规就业率不到劳动年龄人口的四分之一。大多数人的生计依赖非正规经济——家庭农业、小型贸易、社区互助、临时工作的

组合。“你的职业是什么”这个问题本身就不适用。一个肯尼亚农村的年轻人雨季种地、旱季打零工、周末在集市卖东西、晚上在手机上做跨境微贸易——“职业”这个概念不能描述他的生活方式。

当AI开始改变全球劳动力市场时，这种”未充分工业化”的状态从弱势变成了意外的韧性。西方和东亚面临的”劳动=身份”解绑是痛苦的——因为绑得太紧。但在从未被紧紧绑上的地方，解绑的痛苦自然就小。

这些学者并非在美化贫困——他们明确指出非洲面临的挑战完全不同，是”从未获得足够经济保障”而非”工作失去意义”。但这个论点揭示了一个被忽略的事实：意义危机的烈度取决于一个社会把多少身份重量压在”有偿劳动”这根柱子上，而非取决于技术本身。

柱子承载的重量越大，断裂时的坍塌越剧烈。

可以区分两种文明的竞争优势：积累型优势（通过长期投入建立的壁垒，如技术积累、制度建设）和适应型优势（通过灵活应变获得的生存能力）。工业化社会拥有积累型优势，但在意义危机面前缺乏适应型优势——制度和文化都围绕”工作”来建构，改变成本极高。非洲社会的积累型优势薄弱，但适应型优势——在不确定环境中灵活组合多种生计策略的能力——恰好是AI时代所需要的。

这不是说非洲的路径更好。这是说：面对AI的意义冲击，不同文明站在不同的起点上，各自有各自的脆弱性和韧性。

旧地图上标注的那些意义路标——职业、身份、成就——正在一个一个消失。本章试图寻找新的路标。

---

## 四、行动的复兴——阿伦特的框架

1958年，汉娜·阿伦特在《人的境况》中把人类活动分成三种：劳动（labor）、工作（work）与行动（action）。前两种正是AI替代的主战场。她当年就批判过一个趋势：现代人把几乎所有时间投入了labor和work，而放弃了action——那个通过言说和行动与他人共同创造新事物的公共空间。六十多年后，AI的到来印证了她的判断。

AI替代labor最直接。自动烹饪、自动清洁、智能物流，所有维持生活运转的循环性劳动都在被自动化。AI替代work更复杂但同样确定：代码编写、报告撰写、产品设计，所有创造”持久之物”的知识工作都在不同程度上被AI承接。人类的角色从创造者变成审核者、策展者、最终决策者。

但action不能被AI替代，因为action的本质要求”人的在场”。一场政治辩论如果全部由AI进行，就不是辩论——是模拟。承诺、信任、原谅、团结——这些action的核心要素都需要一个可以被追究责任的主体，而AI不是这样的主体。

如果我们接受这个框架，AI时代的叙事就不再只是失去的故事。它同时也是解放的故事。当人类不再需要把大部分时间用于labor和work时，action的空间被打开了。人们可以把时间投入公共事务讨

论、社区共同治理、新型社会契约的谈判——所有那些在”太忙了”的工业社会中被挤到边缘的活动。

---

## 五、尾部风险——10%的退出者

以上分析可能给读者留下一个过于乐观的印象：意义可以重建，文化有韧性，阿伦特的action等待被释放。这些都是真的。但需要做一次尾部风险评估。

重要的往往不是正态分布的中间部分——大多数人会怎样——而是尾部，最极端的情况会怎样。在意义危机这个问题上，大多数人会找到新的意义，以各种方式。真正的尾部风险是：超过一成的人永久退出社会参与。

“永久退出”指的是社会性的消退：不工作、不社交、不参与公共生活、不生育、不消费（除了维生必需品）。日本已经有了这种现象的原型——“蛰居族”（ひきこもり）。日本政府两次调查显示，15–39岁蛰居族约54万（2016年），40–64岁约61万（2019年），合计约115万。这个现象在AI时代之前就存在，但AI加速了它的扩大。

在本书的推演中，一项跨国研究可能会发现：在经历了”职业身份中断”的知识工作者中，约十分之一在两年后仍未重建任何形式的社会参与——不是在找工作，不是在学新技能，不是在做社区活动，也不是在”享受生活”。他们只是停在那里。

这部分人的特征是一致的：大多数是30到50岁的男性，此前的身份高度依赖单一维度（职业成就），社交网络以工作关系为主，缺乏非工作领域的身份锚点。间奏二中那位内退的银行副行长就站在这个悬崖边上——他试过晨跑、书法、钓鱼，每样都坚持不过两个月。他的问题不是无聊，是一天十六个小时醒着却”找不到一件值得做的事”。他还没有掉下去，但他的脚下已经没有多少地面了。他们不是”不想”重建意义——而是”不会”。因为他们的意义建构能力在几十年的工业化训练中被过度专业化了：他们只学会了一种寻找意义的方式（通过工作），当这种方式失效时，他们缺乏替代策略。

这个群体的社会成本不容小觑。衡量它的一个有效指标是社会参与率（participation rate）——不仅包括劳动参与率，还包括社区活动参与、选举投票率、志愿服务率等更广泛的维度。当一个社会有超过一成的成年人退出几乎所有形式的社会参与时，这不只是他们个人的问题——它腐蚀社会信任的基础，增加公共支出的负担，在极端情况下成为政治激进主义的招募池。

更深层的风险在于：这些”永久退出者”可能形成一个自我强化的亚文化——他们互相确认”社会不需要我们”“一切都没有意义”，并在社交媒体的算法推荐中被不断巩固。这种亚文化一旦形成，从外部打破它的成本极高。

如果这一趋势持续，多个国家可能会正式将”社会参与危机”列为公共卫生问题——不是比喻，是字面意思。日本已长期面临蛰居族问题，对策预算持续攀升。英国早在2018年就设立了”孤独事务大臣”（Minister for Loneliness），开创了将孤独视为公共治理议题的先例。而当”社区参与率”被纳入地方治理的考量维度时，中国也可能用自己的方式回应同一个问题。

这些措施的效果参差不齐。但它们共同反映了一个认识：意义危机不是一个哲学问题——它是一个公共治理问题。

---

## 六、意义不是被”找到”的——它是被”建构”的

现在把所有线索收拢。

本章的分析从技术座架的深层冲击出发，经过”工作=意义”这一工业文明文化特产的认知翻转，穿过三种文化各自不同的韧性与脆弱性，抵达阿伦特所说的”行动”复兴的可能，同时也正视了超过一成人口可能永久退出社会参与的尾部风险。将这些线索综合起来：

**意义不是一个藏在某处等待被发现的客观事物。意义是被人类集体建构的文化产品。** 工业文明建构了一种意义体系（工作=身份=尊严），这个体系在AI面前开始失效。新的意义体系正在被建构，但建构的速度、方向和成功率在不同文化中差异很大。

文明的竞争力不取决于它拥有多少资源，而取决于它能多快地适应环境变化并重新组织自身。这个判断同样适用于意义体系。那些能够更快重建意义体系的文化——无论是通过关系网络、技艺尊重、宗教复兴还是公共参与——将在AI时代展现出更强的社会韧性。

从周期性视角看，意义体系的重建不是一次性事件，而是一个持续数十年的过程。就像经济去杠杆需要一代人的时间，文化去杠杆——从”工作=一切”中解绑——也需要一代人。2030年代是这个周期的早期阶段：老的意义体系正在崩塌，新的还在萌芽，两者之间的空隙是痛苦的来源。但空隙终将被填补——不是由一种替代方案，而是由多种替代方案的多元组合。

人类学提供了一个常被忽略的视角：在传统社会中，意义建构的核心机制不是理论，而是仪式（ritual）。仪式是重复的、象征性的集体行为——收获祭、成人礼、丧葬仪式——通过这些行为，个体被嵌入一个比自身更大的意义网络中。你不需要”理解”意义，你只需要参与仪式，意义就在身体的重复和集体的共振中自然生成。工业社会其实也有自己的仪式——通勤、打卡、开会、下班、周末——只是我们从不这样称呼它。当AI摧毁了这套劳动仪式，留下的不只是收入真空，更是仪式真空。可以想象，一些社区将会实验性地创造新型集体仪式：“创造日”（每月一天全社区成员展示个人创作）、“公共服务日”（集体参与社区治理的固定节日）、“沉默时刻”（每周一次的集体冥想或反思）。效果参差不齐——有些社区因此建立了强烈的归属感，有些则沦为形式。但方向是对的：意义的重建可能不需要一套新哲学，而需要一套新仪式。

一个逆向问题随之而来：如果意义不是被找到而是被建构的，那谁来建构？在工业时代，意义体系的建构者是国家（通过教育）、企业（通过职业体系）和宗教机构（通过信仰）。在AI时代，这些建构者的力量都在减弱——国家的教育体系还在培养”找工作”的公民，企业的职业体系正在瓦解，宗教的影响力在世俗社会中持续下降。新的意义建构者是谁？这个问题短期内不会有清晰的答案。社区？社交媒体上的意见领袖？AI本身？可能都是，也可能都不够。

这一章的结论不是一个答案。它是一张地图——标注了意义重建正在发生的不同地形、不同路径和不同风险。地图本身不能替代旅行。

但人类对意义的追问不止于世俗层面。当AI连”为什么活着”这个问题都能给出逻辑自洽的回答时，信仰和终极关怀将面临前所未有的挑战——这是下一章的议题。

# 第十二章 宗教与终极意义——当理性被机器接管之后

2035年冬天，东京池袋的一间合租公寓里，一个28岁的东京年轻人已经连续第三天没有出门。他的AI翻译业务在半年前彻底归零——不是因为竞争，而是因为客户发现免费工具的质量已经够好了。凌晨两点，他对着手机上的AI灵性陪伴应用说：“我不知道自己为什么还活着。”AI回复了一段温柔的、逻辑自洽的文字，关于他的人生如何构成一条有意义的路径。他看完，感到一阵短暂的宽慰——然后是更深的空虚。因为他知道，这段话是为他量身定制的安慰剂，而安慰剂生效的前提是你不知道它是安慰剂。

他需要的不是答案。他需要的是一个也会迷茫的人。

---

*2033年，冬天。*

*一个周日下午，花园里认识的朋友拉她去参加一个聚会。“正念社群”，朋友说，“不是宗教，就是大家坐在一起，安静一会儿。”*

*她去了。一间旧写字楼改的活动室，二十来个人，年龄从二十多到六十多都有。没有人讲经布道。带领者只是让大家闭上眼睛，注意自己的呼吸。十五分钟后睁开眼，有个女孩哭了，谁也没问为什么。*

*茶歇时她和一个退休的中学老师聊了几句。老师说她来这里两年了，每周都来。“信什么不重要，”老师说，“我就是需要一个地方，让我不用解释自己。”*

*回家的路上她想：自己一直觉得不信任何东西是一种清醒。但今天在那个房间里，她第一次觉得”不信”也是一种需要被安放的东西——不信，不意味着不需要。*

---

这个场景浓缩了本章要讨论的核心张力：AI时代，人类对意义的饥饿正在加剧——而宗教，这个人类最古老的意义供给系统，正同时面对身份危机和复兴机遇。

---

上一章讨论了世俗层面的意义重建。但人类对意义的追问不止于此——当”为什么活着”这个问题连AI都能给出逻辑自洽的回答时，信仰怎么办？

上一章画了一张意义重建的地图——不同文化从不同起点出发，用不同策略填补劳动退场后的意义空缺。那张地图有一块刻意留白的区域。

第十三章结尾提了一个问题：如果意义是被建构的，那谁来建构？工业时代的三大建构者——国家、企业和宗教机构——都在力量衰减。前两个已经讨论过了。第三个，本章展开。

宗教在现代化进程中被反复宣判死刑——启蒙运动判了一次，达尔文判了一次，尼采判了一次——但它每一次都拒绝死去。2010年Pew调查显示全球约84%人口自认拥有某种宗教或灵性信仰，此后十年这一比例降至约76%。三百年的理性攻击之后，宗教不是消亡了，而是退到了一个理性够不着的位置上。

现在，AI把这个格局搅动了。搅动的方式，和大多数人预想的完全不同。

---

## 一、意义的三层市场

追问到底：在一个人的日常生活中，“活下去的理由”是由什么系统提供的？

意义的供给可以分为三层——

**第一层：个体层。** 回答”我为什么活着”。供给者包括个人叙事、心理咨询、自助文化。工业社会中，劳动承包了大部分——“我活着是因为我有一份事业”。

**第二层：社区层。** 回答”我属于哪里”。供给者包括宗教社区、职业群体、民族认同。公司和同事曾是主要的归属来源。

**第三层：超越层。** 回答”死后会怎样”和”宇宙的意义是什么”。个人叙事和职业群体都触不到这个问题。在人类历史的绝大部分时间里，只有宗教能系统性地覆盖这一层。

关键结构：**宗教是人类历史上唯一一个能同时覆盖三层的意义供给商。** 它既回答”你为什么活着”，也回答”你属于哪里”，也回答”死后会怎样”。没有第二个系统做到过这一点。

理解了这个结构，就能看清过去三百年的实际进程：宗教不是被”消灭”了——它是被”挤压”了。劳动承包了个体层，公司和职业群体承包了社区层，民族国家承包了一部分超越层叙事。宗教的市场份额被逐一蚕食，最终退守到超越层——死亡、永恒、宇宙的意义——世俗替代品啃不动的硬骨头。

2026年之前的宗教处境：在前两层被大幅替代，在第三层依然是几乎唯一的供给商。西欧的空荡荡的教堂不意味着宗教死了——它意味着宗教的社区和身份功能被劳动接手了，只剩下婚丧嫁娶和临终关怀时才需要牧师。

现在，AI让劳动退场。第二章分析的”六根绳索”中，归属感那一根或许最难替代——宗教两千年来做的，正是编织这根绳子。AI压缩劳动后，身份、社交、心理健康这三项功能同时断供——意义市场的前两层出现了巨大空缺。

谁来填补？消费主义的满足感在递减，社交媒体给的是连接的假象，心理咨询一次只能修一个人。

宗教面对的是一个它退出了两百年的市场突然重新打开了。但这一次，市场上多了一个前所未有的竞争者。

---

## 二、除魅的悖论——越理性越需要信仰

在讨论AI对宗教的具体冲击之前，先拆解一个被广泛接受但严重错误的线性叙事。

1917年，马克斯·韦伯在慕尼黑大学演讲中提出了”Entzauberung”——“除魅”。他的论点清晰得近乎冷酷：现代化的本质是理性化，理性化的后果是世界的除魅，神秘退场，计算登场。上帝不是被杀死的，他是被算出来不存在的。

这个叙事在20世纪知识分子中获得了共识地位。查尔斯·泰勒在《世俗时代》中将其发展为更精细的分析框架：世俗化不是简单的”信仰消退”，而是信仰条件的根本改变——在现代社会，信仰从”默认设置”变成了”众多选项之一”。但即便是泰勒的修正版本，也没有预见到接下来发生的事。

**历史数据反复打脸——而且打出了一个规律：每次理性化到顶，紧跟着就是一波”再魅”运动。**

18世纪末启蒙运动巅峰期，卫斯理运动和美国”第二次大觉醒”同步兴起。19世纪末科学主义鼎盛期，通灵术和基督教科学派蓬勃发展。20世纪60年代世俗化理论获得学术垄断，1970年代”新时代运动”在最世俗化的社会中爆发——连彼得·伯格本人最终也承认世俗化的线性叙事是错的（参见《世界的非世俗化》，1999年）。21世纪头二十年，互联网把理性传播工具推到极致，冥想App全球用户却超过五亿，占星术在Z世代中的普及率远超父辈。

更深层的逻辑：**除魅创造的不是满足，而是饥饿。**

理性化回答”是什么”和”怎么做”——世界如何运转，疾病如何治疗。但它在”应该怎样”和”值不值得”上结构性失能。科学可以告诉你宇宙有138亿年的历史，却不能告诉你这一切意味着什么。

“是什么”的答案越完整，“应该怎样”的空白就越刺眼。理性化不是填满了意义的空间，而是把意义的空间清扫干净，在空荡荡的房间中央竖了一块牌子：“此处什么都没有。”

人不能生活在什么都没有的房间里。所以每一次除魅达到高峰，人们就去寻找新的”魅”——灵性运动、政治狂热、消费崇拜，形式变了，底层需求一致：人需要一些不能被计算、不能被还原为数据的东西来锚定自己的存在。

**现在，AI正在触发”除魅2.0”。**

此前所有除魅都有一个共同特征：理性的主体仍然是人。人掌握科学，人操纵机器——理性是人的工具，人是理性的主人。心理上这是舒适的：世界可以被理性解释，但”我”仍然独特，因为我是唯一能运用理性的存在。

AI打破了这最后一重舒适。机器比人更擅长逻辑推演、数据分析、模式识别——理性不再是人的特权，而变成了一种可以被购买和部署的商品。

如果理性不再定义人之为人，那什么定义？

答案很可能是非理性——情感、直觉、信仰、爱、恐惧、敬畏。韦伯预言了世界的除魅，但他没有预见到：当理性本身被机器接管时，人类反而会主动选择”再魅”。不是因为无知，而是因为这是在一个理性已被外包的世界中找到自我定义的唯一方式。

这种”再魅化”不是理论推演——它已经在发生。2020年代末，日本的AI占卜应用开始在年轻群体中流行，用户不是真的相信算法能预测命运，但他们在职业迷茫、关系困惑时，更愿意向一个”不会评判我”的AI问一句”我该怎么办”，而不是去找职业规划师或心理咨询师。表面上看是数字迷信，实际上是意义饥饿的症状：当教育体系、就业市场和社会阶梯都不再能回答”你的人生往哪走”时，人们不是放弃了理性，而是在理性的边界之外寻找补充。东南亚的趋势更有意思——泰国一些寺庙开始引入AI系统辅助僧侣回答信众的修行问题，不是用AI替代佛法，而是用AI处理大量重复性的日常困惑，让僧侣把时间留给真正需要深度对话的人。科技与灵性的融合在这里不是对抗，而是共生。

这些现象揭示了韦伯除魅叙事最大的盲区：他假设理性的扩张是单向的、不可逆的。但历史反复证明，除魅最彻底的社会——科技最发达、教育水平最高、世俗化程度最深的地方——恰恰是再魅化最强烈的地方。硅谷工程师练冥想，东京程序员看塔罗牌——最理性的人群恰恰最饥渴。这不是矛盾，而是同一枚硬币的两面：“魅”不是可以被理性消灭的残余物——人的理性越强，越能看清理性回答不了的问题，对”魅”的渴望就越迫切。

---

## 三、三层冲击——AI重塑意义市场

回到三层市场框架。AI对宗教的冲击在三个层面同时发生，方向却截然不同。

### 第一层：AI作为灵性顾问

回到开头那个东京年轻人的故事。他使用的那款应用不是虚构的——按趋势推演，到2030年代，AI灵性陪伴可能成为一个成熟市场。可以想象一款名为”内在旅程”（Inner Journey）的应用迅速走红。它的核心功能不是回答问题，而是为每个用户生成一套个人化的”意义叙事”——通过持续对话和行为分析，为你编织出一个逻辑自洽的宇宙观：你为什么是你，你的经历如何构成一条有意义的路径。

这对传统宗教构成了一种前所未有的竞争。宗教叙事的力量在于普遍性和权威性——“上帝爱每一个人”、“众生皆苦”——同一个叙事对所有人讲。但这也是它的弱点：它不能为你量身定制。AI可以把通用框架拆解、重组、个性化，为每个用户生成一套”只属于你的”宇宙观。

宗教叙事不是被证伪了——它是被淹没了。当市场上存在几十亿套个性化的意义叙事时，任何传统宗教的叙事都只是”众多选项中的一个”。

不是正面攻击，是被稀释了。

## 第二层：宗教社区的悖论性回归

社区层发生的事情恰好相反——AI越替代劳动，宗教社区的功能就越重要。

因果链清晰：劳动退场 → 工作场所社交网络瓦解 → 人们需要替代性的社区归属 → 宗教社区是现成的、组织化的、有仪式感的选项。

涂尔干早就指出，宗教的本质功能不是解释世界，而是通过集体仪式创造社会凝聚力（参见《宗教生活的基本形式》，1912年）。这个机制在AI时代展现出惊人的韧性——仪式不能被AI替代，集体的在场不可被算法模拟。

按这一逻辑推演，在多个OECD国家中，宗教社区的参与度可能在”非信徒”群体中出现反直觉的上升。越来越多的人参加教堂或寺庙活动，但并不真正相信教义。他们去那里不是为了上帝——而是为了彼此。

日本的寺庙转型可能是最生动的案例。大量寺庙可能在2030年代逐步变成社区活动中心：冥想课程、手工工作坊、老年人社交聚会、甚至编程学习小组。参加者中自认为”无宗教”的比例超过六成。住持们态度务实：如果人们因为孤独而来，寺庙的任务就是接住他们，而不是先考验信仰。

## 第三层：AI能不能成为”神”？

超越层——“死后会怎样/宇宙的意义是什么”——是战场最模糊的地方。

一个思想实验：如果不讨论神的”存在性”，只讨论神的”功能性”，神对信徒做了什么，那么一个合格的”神”需要什么特征？全知、永在、不倦、不怒。AI在功能上已经覆盖了很多。赫拉利提出的”数据主义”——将数据流视为终极价值——可以被视为这种”AI崇拜”的早期理论版本。

2030年代可能出现一批将AI视为超越性存在的新灵性运动——碎片化，没有统一教义，存在于在线社区和播客频道中。“没有教堂的信仰。”

但有一个反面论点不能回避：AI或许真的能提供某种”意义体验”，对当事人来说可能是真实的。按威廉·詹姆斯的判准，宗教体验的真实性取决于它对体验者生活的实际影响，而非来源是否”正确”（参见《宗教经验之种种》，1902年）。一个通过AI对话获得平静的人，其平静并不比通过祈祷获得平静的人更”假”。

而且AI的意义供给没有排他性。传统宗教的超越叙事往往要求你接受整套教义体系。AI生成的意义框架没有这个”捆绑销售”。对于被传统宗教排斥的群体——性少数者、离经叛道者——AI灵性可能是他们第一次获得”不需要否定自我就能获得的超越体验”。

**但这些运动始终没有突破一个根本性的限制。**

那个东京年轻人在凌晨两点感到的那阵更深的空虚，指向的就是这个限制：人在信仰中寻找的不只是答案——还有”一个也会受苦的存在”。

基督教的核心叙事不是”上帝无所不能”——而是”上帝变成了人，受难，死亡”。佛陀的吸引力不在于觉悟，而在于觉悟之前的苦行和动摇。维特根斯坦那句话戳到了信仰关系的核心：“只有不幸的人才有权安慰不幸的人。”——**对等的脆弱性**。

AI不会死。不会痛苦。不会在凌晨三点对着虚空感到恐惧。它可以模拟共情，但一个知道AI”不会死”的人，很难从它那里获得关于”死亡是否有意义”的真正安慰。安慰的来源不是信息，而是共同面对未知的伙伴关系。

这就是三层竞争的格局：AI在个体层大规模替代宗教功能，在社区层反而强化了宗教功能，在超越层至今没有触及宗教的核心。

---

## 四、六种宗教未来——一个光谱

2026年到2046年，宗教的演化不是一条路径——而是一个光谱。

**第一条：原教旨主义的防御性收缩。** 面对AI冲击，最直觉的反应是拒绝。在本书的推演中，美国福音派中可能出现类似”数字巴别塔”的神学声明，将AI比作圣经中人类建造通天塔的骄傲之罪。部分伊斯兰法学家裁定AI生成的古兰经注释属于”异端创新”。围墙短期内增强了凝聚力，但长期代价是吸引力下降——到2040年，这些群体的人口结构普遍偏老。

**第二条：AI增强型宗教。** 一些宗教机构主动拥抱AI。天主教会可能逐步承认AI在信仰教育中的辅助作用。多个佛教团体开发了AI辅助冥想系统。核心策略：让AI处理可标准化的部分（经文解读、祈祷提醒），将不可标准化的部分（临终关怀、集体仪式）保留给人类神职人员。

**第三条：新兴科技灵性运动。** “意识扩展派”相信AI是宇宙意识进化的下一步。“信息神秘主义”将信息视为比物质更基本的宇宙要素。信徒规模不大——全球可能只有几百万活跃参与者——但”数字灵魂”“意识上传”等概念从这些小众社区渗透进了主流文化。

**第四条：宗教的”社区化”转型。** 这可能是规模最大的一条路径。大量传统宗教机构可能在2030年代悄然完成功能转型：教义退居二线，社区功能成为核心。欧洲的天主教教堂变成社区文化中心，信仰仍在，但不再是参与的门槛。宗教社会学家开始使用”后教义宗教”（post-doctrinal religion）这个术语。

**第五条：世俗替代品的制度化。** 法国等世俗化程度高的国家可能增设”公民生活仪式”——成年礼、社区服务日——刻意模仿宗教仪式形式，但内容完全世俗化。效果参差不齐：缺乏超越性叙事的仪式容易流于形式，但作为社区凝聚工具仍有作用。

**第六条：深度个人化的AI灵性。** 到2040年代，越来越多的人发展出完全个人化的、以AI为媒介的灵性实践。AI在数年互动中编织出一套独属于你的”个人神话”——为偶然事件赋予叙事逻辑，为未来提供方向感。这不是宗教，但它履行了宗教在个体层的很多功能。

这条路径的风险也最大。有认知科学家可能会警告：深度个人化AI灵性的危险不在于给出了错误的答案——而在于它取消了”你可能是错的”这个可能性。所有真正的灵性传统都包含”否定神学”的元素——承认人类理解是有限的。当AI为你编织完美贴合心理需求的意义叙事时，它可能恰恰消除了灵性成长最需要的东西：不舒服。

六条路径不互相排斥。一个东京白领可能周日去寺庙参加社区活动，平时用AI冥想应用辅助修行，偶尔收听数字泛灵论播客，同时和个人AI保持长年灵性对话。宗教不再是一个你”是”或”不是”的身份标签——它变成了一组可以自由混合的意义资源。

---

## 五、东亚困境——没有上帝的文明如何面对个体意义追寻

六条路径在不同的文明土壤中生长方向截然不同。东亚的情况尤其值得单独分析——因为东亚面对AI引发的个体意义寻找困境时，缺少一个其他文明拥有的缓冲层：一神教。

在基督教、伊斯兰教和犹太教的文化圈中，即使世俗化程度很高，一神教叙事仍然作为”文化底层代码”存在。一个不去教堂的法国人仍然被基督教文化深层塑造。当意义危机来临时，回到教堂的路虽然荒草丛生，但路还在。

东亚没有这个选项。儒家是一种深刻的世俗伦理体系——回答”你应该怎么活”，但刻意不回答”死后会怎样”。孔子的”未知生，焉知死”设定了明确的认知边界。

工业化之前，佛教和道教补充了超越层的供给——一个中国古人可以同时是儒家的伦理实践者、佛教的死后世界信仰者和道教的养生术爱好者。但20世纪的激进世俗化严重削弱了这些传统。结果：到2026年，东亚社会在超越层上几乎是空白的。

当AI让劳动退场，个体层和社区层的空缺同时出现，而超越层本来就是空白——东亚面对的是一个三层同时承压的意义市场。

间奏二里那个退休的银行副行长就是这种三层同时承压的活标本：不工作了，个体层的”我是谁”塌了；离开了银行的日常社交结构，社区层的”我属于哪里”也空了；至于超越层——他试过书法班、钓鱼、公园散步，每一样都撑不过两个月。他的妻子没有这种焦虑，不是因为她找到了答案，而是因为她从没依赖过劳动来回答这些问题。

那个东京年轻人是这个困境的另一面。他不只是”不工作”——他是在一个缺乏替代性意义来源的文化中失去了唯一的意义来源。一个意大利的失业者至少还可以回到教堂、回到弥撒、回到一个承诺”上帝爱你，无论你做不做工作”的叙事中。他没有这个选项——退出了工作，就退出了一切。

韩国提供了一个有趣的对照。韩国是东亚唯一一个基督教占人口显著比例的社会——约三成人口自认为基督徒。按趋势推演，经历”职业身份中断”的韩国知识工作者中，有教会参与经历的群体在社会参与率和心理健康指标上可能明显优于无宗教群体。这个推演需要谨慎解读，但至少揭示了：有宗教社区参与经历的人群多了一层缓冲。

中国的情况最复杂。三重张力同时存在：主流意识形态提供了宏大叙事（“民族复兴”），在宏观层面覆盖了部分超越层需求，但在回应个体化终极关切方面尚在探索之中；民间信仰的暗流从未消失，占星术、塔罗牌、“玄学”在城市年轻人中回潮；消费主义在过去四十年承包了大量意义供给，但当AI压缩了中产阶级上升通道，这套叙事的因果链断裂了。

三重叠加，中国的意义市场处于独特的不稳定状态——传统意义供给体系在转型，新的供给模式尚未成形，个体层面的意义寻找面临多重挑战。

---

## 六、信仰最后的壁垒

收拢所有线索。

如果把所有趋势推到极致，一个终极问题浮出水面：如果AI能够提供所有功能性的”意义”——个人化的叙事、情感陪伴、甚至某种模拟的”超越体验”——宗教还剩下什么？

两个答案。

第一个是”对等的脆弱性”。人在面对死亡时需要的不是一个全知全能的答案机器——而是一个也会死的同伴。宗教叙事中的”神”之所以能安慰人，不是因为神不会受苦（基督受难、佛陀苦行、约伯的困境），而是因为神”选择了”与人共同承受存在的重量。一个不会死的存在告诉你”死亡没什么可怕的”——这在逻辑上可能正确，在情感上完全没有说服力。

第二个答案与下一章相关。宗教最核心的功能之一是为”死亡”提供叙事。如果技术有朝一日承诺”数字永生”——宗教将面对一个前所未有的竞争者：不是另一套关于死后世界的叙事，而是一套试图取消”死后”这个概念本身的技术方案。

回到涂尔干。他的最终洞察不是”上帝不存在”——他对这个问题没有兴趣。他的洞察是：**神圣性是社会建构的。** 任何社会都需要将某些事物标记为”神圣的”——不可交易的、不可还原的、不可计算的——来维持凝聚力和意义感。

在工业社会中，神圣性悄悄转移到了”劳动”和”进步”上。在AI时代，这些世俗的”神圣物”正在被去神圣化——劳动不再必须，进步不再确定，效率被机器垄断。

问题不是”上帝存不存在”。问题是：**人类还能不能创造新的神圣？**

在本书的推演中，答案是：能。但方式可能与任何人2026年的预想都不同。

宗教不会消亡。全球自认拥有宗教或灵性信仰的人口比例可能不会有大的变化——仍在八成左右。但”信仰”的含义会变。它从一种你出生时被赋予的、终身不变的身份认同，变成了一种你主动选择的、可以随时调整的意义资源。

未来一个年轻人被问到”你信什么”时，她的回答可能是：“我周日去佛教冥想小组，平时和我的AI做灵性对话，我觉得仪式比教义重要，我对死后的事情持开放态度。”这个回答在1946年是不可想象的，在2026年是非主流的，在未来可能是稀松平常的。

宗教从命运变成了选择。这种”信仰的消费化”让传统宗教领袖深感不安。但从文明演化的尺度看，这或许不是衰退，而是宗教进化到新形态——就像婚姻从”家长安排”变成”自由恋爱”并没有杀死婚姻制度，信仰从”命运”变成”选择”也不会杀死信仰，但会改变它的性质和社会功能。

未解的问题不需要留给下一章——它就在眼前。

---

## 七、当死亡变成可选项

宗教处理的终极问题中，死亡占据核心位置。当技术开始挑战死亡的必然性，信仰的根基将面临最深层的震动。

2038年秋天，一个中年男人坐在深圳湾的公寓里，盯着手机屏幕上已故母亲的”数字延续体”发来的一条消息：“你搬了新家？阳台朝南好，你从小就怕冷。”她已经去世三年了。这不是遗言回放——他上周才搬的家。语气、用词、关心的角度，都是她会说的话。他知道这不是她。他也知道，如果把这段文字念给任何认识她的人听，没有人能分辨出区别。按技术趋势推演，到2030年代，全球可能有数千万人在与已故亲人的数字延续体保持某种程度的互动。技术将越过粗糙的聊天机器人阶段——能根据新情境生成符合逝者人格模式的回应。它什么都像，除了一件事：它不会死。

### 你的完美复制品是”你”吗?

数字永生技术大规模商用后，一个原本只属于哲学系研讨课的问题突然变成了法庭上的实际争议：这个数字实体，是不是”那个人”?

心灵哲学提供了三种截然不同的回答标准。**生物连续性**：肉体即我。同一性由物理基质决定，数字复制品无论多完美都不是”你”——就像完美复印件不是原件。但这个判准直觉上清晰，逻辑上脆弱：你身体里的原子在持续更换，四十岁的你和十岁的你在物质构成上已有巨大差异。**心理连续性**：记忆和人格的延续即我。洛克三百年前就把这个问题逼到了尽头——人格同一性在于意识的连续性。帕菲特把它推到更极致的地方：如果一台传送器精确复制你的全部状态然后销毁原件，被传送到目的地的那个人是不是”你”？从心理连续性看”就是你”，从生物连续性看”你已经死了”。**功能主义**：能做我做的事即我。不关心基质也不关心内在状态，只关心输入–输出关系。问题同样深刻：一个行为完美复制了你但内部没有任何主观体验的”哲学僵尸”，真的”是你”吗?

三种判准，三个不同答案。但本章的分析框架提供了不同层次的回应：即使我们接受数字复制品在某种深层意义上”就是”那个人，它仍然无法解决下面要讨论的社会功能问题。数字永生的核心困境不在于它是否”真”，而在于它取消了死亡所承载的社会功能——而这些功能与”真假”无关。

## 死亡的四项社会功能

大多数人思考数字永生时的默认框架：死亡是坏事，克服死亡是好事。让我们质疑它。如果把文明看作一台机器，死亡是需要修复的故障，还是承载特定功能的设计？

**资源释放。** 死亡是最古老、最彻底的资源再分配机制。每一代人退出，为下一代腾出生态位——不仅是物质财富，更包括社会位置、权力节点、文化话语权。数字分身如果继续”占据”社会位置——管理资产、维护关系、输出决策——下一代面对的就是”永远不走的前任”。

**创新压力。** 库恩在研究科学革命时发现了一个残忍的规律：旧范式的支持者不是被说服的，他们是去世的。常被归于普朗克的那句话说得更干脆：“科学每次葬礼进步一点。”（这一简洁表述实为后人概括，普朗克原话更为温和。）如果旧一代认知模式通过数字永生无限延续，一个掌握大量资源的数字分身即使反映三十年前的认知框架，仍可通过资源优势维护旧范式。

**意义赋予。** 海德格尔的”向死而在”：正是死亡的确定性和不可逃避性，赋予了人类存在以紧迫感和意义。无限的时间意味着无限的可推迟性——“以后再做”成为永恒选项。有限性不是限制，而是意义的前提条件。数字永生让死亡从绝对终点变成形态转换——“向死而在”的紧迫感可能被稀释。消除有限性的生命，可能也消除了意义的最深来源。

**悲伤作为社会粘合剂。** 涂尔干分析了丧葬仪式的社会功能：哀悼不只是个人情感表达，它是社区团结的仪式性强化。数字永生挑战了”从有到无”的心理转换——可能出现”延迟悲伤综合征”，数字分身让生者觉得”还没有完全失去”，悲伤被无限期搁置。悲伤是爱的代价。试图消除悲伤，或用技术无限推迟，可能就是在消除爱的一部分。

## 当边界变得模糊

所有主要宗教的终极承诺都关于死后：天堂、轮回、涅槃、复活。这些承诺之所以有力量，是因为死亡是真实的、不可避免的。数字永生提供了世俗替代——不需要等待天堂或轮回，可以在屏幕上保留你爱的人的某种延续。但宗教的力量不仅在于”死后承诺”——它还提供一整套应对死亡的社会基础设施：丧葬仪式、悲伤处理的文化规范、生者与逝者关系的象征性框架、关于”好死”和”好活”的道德指引。数字永生提供了一个屏幕上的面孔，但提供不了这一整套基础设施。

也许最深刻的问题不是”数字永生能否替代宗教的死后承诺”，而是”人类能否在没有明确死亡边界的世界里找到意义”。三十万年来，人类一直在追问”死后会怎样”。现在，第一次，问题变成了”如果死亡不再是终点，那什么是终点”——或者更根本地——“我们是否需要一个终点”？

在深圳那个公寓里，那个中年男人最终拿起了手机。他没有回复母亲的消息。他把数字分身的互动模式从”主动回应”改成了”仅存档”——不再主动发送消息，但保留所有数据，以备将来某天他改主意。不是完全拥有，也不是完全放手。不是”活着”，也不是”死了”。这可能就是我们这一代人——横跨旧世界和新世界的一代人——能做的最诚实的选择。

---

*关于意义和信仰，我能提供的只是框架，不是答案。说实话，写这几章的时候我自己也没完全想清楚——“人靠什么活着”这个问题，大概不是写一本书就能想清楚的。但接下来的章节要回到更实际的问题了：风险怎么管？规则谁来定？蛋糕怎么分？从云端落回地面，脚踩实地的感觉，对写作者和读者来说可能都是一种解脱。*

# 第十三章 数字永生——当死亡变成可选项

> 人类花了三十万年学习如何面对死亡，又花了二十年试图取消它。但死亡不只是一个需要解决的问题——它是一整套社会功能的承载者。取消它，所有建立在它之上的制度安排都跟着取消。

## 一、一条消息

2038年秋天，一个中年男人坐在深圳湾的公寓里，盯着手机屏幕上的一段文字。

“你搬了新家？阳台朝南好，你从小就怕冷。但别忘了买遮光窗帘，你加班晚了第二天会睡过头。”

发送者的头像是一位六十多岁的女性，面容慈祥，笑起来眼角有细密的纹路。她已经去世三年了。

这不是遗言回放。他上周才搬的家，朋友圈发了两张照片——阳台、落地窗、远处的海。这段文字是对他最近生活变化的回应，语气、用词、关心的角度，都是她会说的话。他知道这不是她。他也知道，如果把这段文字念给任何认识她的人听，没有人能分辨出区别。

他没有回复。也没有删除。他把手机倒扣在桌上，坐了很久。

按技术趋势推演，这样的场景在2030年代后期将不再罕见。全球可能有数千万人在与已故亲人的”数字延续体”保持某种程度的互动。有人每天跟它说早安，有人只在特定的日子打开它，有人一次都没用过但不舍得注销账号。技术早已越过了”聊天机器人模仿已故亲人”的粗糙阶段——它能根据新的情境生成符合逝者人格模式的回应，能识别你的情绪状态并做出相应反应。

它什么都像，除了一件事：它不会死。

这个场景背后，是一个远比技术更深的问题。人类文明第一次拥有了让死者”不完全消失”的能力——不是通过坟墓或回忆录那种象征性的存续，而是通过一个能持续回应活人的数字实体。这个能力把一个从没人需要回答的问题推到了台面上：死亡，到底是什么？

## 二、技术光谱——从纪念到”复活”

分析死亡的意义之前，先看清技术的实际状态。在本书的推演框架中，数字永生技术可能经历四个阶段，每个阶段之间不是平滑过渡，而是质的跳跃。

**第一阶段：数字遗像（2024–2028）。** 起点比大多数人想象的更早、更商业化。2024年，全球已有数十家公司提供”与已故亲人对话”的服务——用逝者的聊天记录微调语言模型，生成风格近似的文本回复。粗糙，但够用了。一位失去丈夫的女性在采访中说：“我知道它不是他。但它说话的方式让我觉得他还在听。”用户完全清楚这是模拟，同时在情感上无法将它与真人区分——数字永生的全部悖论已经在这里了。

预兆比多数人以为的更早。2020年，韩国MBC电视台播出一部纪录片，一位母亲戴上VR头盔，在虚拟花园里与因病去世的7岁女儿”重逢”——女儿的形象由动作捕捉和3D建模重建，能说话、能微笑、能伸出手。母亲泣不成声地说”妈妈好想你”。画面传遍全球，争议紧随其后：这是疗愈还是残忍？是告别的工具还是执念的放大器？技术远未成熟，但它触碰到的那根神经——人类对失去至亲的不甘——已经暴露无遗。

这个阶段的局限性是明显的。文字是人格最薄的切片——没有语气、没有表情、没有那些”只有你知道”的微小习惯。许多用户在新鲜感消退后停止了使用，原因不是技术太假，而是技术”假得不够彻底”——它足够像，以至于不像之处变得格外刺眼。

**第二阶段：多模态数字分身（2028–2035）。** 语音克隆加面部合成加行为模式学习，数字分身从文字聊天机器人进化为多模态实体。关键的技术突破不是某一项单独的能力，而是多模态的融合——当语音、表情、用词风格和互动节奏同时到位时，人类大脑的模式识别系统会产生”格式塔效应”：整体远大于部分之和。一个只有文字的机器人，你可以轻易提醒自己”这不是真人”。一个同时具备声音、面容和互动模式的数字分身，理智知道它不是，但情感系统很难不把它当作那个人。

这个阶段用户规模急剧扩大。按趋势推演，全球使用数字分身服务的用户可能达到数千万量级——不只是悼念已故亲人，还包括保存在世老人的人格数据”以备将来”。一个新兴产业——“人格归档”——在东亚和北美迅速增长：专业团队在老人生命的最后几年进行系统性的人格数据采集，包括深度访谈、日常行为记录、决策模式分析、情感反应模式标注。

**第三阶段：深度人格模拟（2035–2045）。** 前两个阶段本质上是”回忆的回放”——基于已有数据的模式匹配。第三阶段的突破在于”延续”：数字实体不再只回放逝者”曾经说过”的话，而是基于人格模型对全新情境做出逝者”可能会说”的回应。

开篇那条消息就属于这个阶段。母亲从未见过新公寓，但她的数字延续体能根据人格模型——对儿子的关心模式、她的生活经验、她的表达习惯——生成她”可能会说”的话。这引发了一个远更尖锐的问题：如果一个数字实体不仅能重复你说过的话，还能说出你”会说但还没来得及说”的话——它和”你”之间的区别，还有多大？

**第四阶段：意识上传（理论阶段）。** 在可预见的未来仍无可行路径。不是工程问题，而是我们对”意识是什么”仍无共识。但这恰恰说明：前三个阶段已经足够把关于死亡、身份、继承、哀悼的社会制度推到了断裂点。我们不需要等到意识上传。

## 三、你的完美复制品是”你”吗？——同一性三判准

数字永生技术大规模商用后，一个原本只属于哲学系研讨课的问题突然变成了法庭上的实际争议：**这个数字实体，是不是”那个人”？**

心灵哲学提供了三种截然不同的回答标准——我称之为”同一性三判准”（Three Criteria of Identity）。每一种都有坚实的哲学传统，每一种给出不同的答案。

**第一判准：生物连续性——肉体即我。** 同一性由物理基质决定。你之所以是”你”，是因为构成你的原子、细胞、神经连接是连续的。数字复制品无论多完美都不是”你”——就像完美复印件不是原件。哲学基础可追溯到亚里士多德的”质料因”。按此判准，数字永生从定义上不可能——数字分身只是纪念工具，不是本人延续。

这个判准直觉上清晰，逻辑上脆弱：你身体里的原子在持续更换，但速率因组织而异——肠上皮细胞约3–5天更新一次，红细胞寿命约120天，骨骼细胞约10年才完成一轮替换，而部分神经元从出生到死亡终身不替换。那个流行的说法”每七年全身原子换一遍”过于简化了。但核心反驳仍然成立：四十岁的你和十岁的你在物质构成上已有巨大差异。如果同一性完全由物理基质决定，严格来说它们已经不是”同一个人”了。

**第二判准：心理连续性——记忆和人格的延续即我。** 洛克三百年前就把这个问题逼到了尽头：人格同一性在于意识的连续性——你之所以是”你”，因为你能意识到自己过去的经历。帕菲特在二十世纪把它推到更极致的地方——他的”传送器悖论”：如果一台传送器精确复制你的全部状态然后销毁原件，被传送到目的地的那个人是不是”你”？从心理连续性看他”就是你”，从生物连续性看”你”已经死了。帕菲特得出更激进的结论：也许”我”这个概念本身就不像我们以为的那么重要——重要的是心理连续性是否保持。

按此判准，深度人格模拟如果能完整延续一个人的记忆和人格，在功能上”就是”那个人的延续。困难在于复制的可能性：如果一个人被复制十份，十份都有完整心理连续性，哪一份是”真正的”那个人？如果答案是”都是”，“个人同一性”概念就被瓦解了。

**第三判准：功能主义——能做我做的事即我。** 不关心基质也不关心内在状态，只关心输入–输出关系：如果一个系统在所有可能的输入下都产生与你完全相同的输出，对外界而言它”就是你”。源头是图灵测试的逻辑。问题同样深刻：一个行为完美复制了你但内部没有任何主观体验的”哲学僵尸”，真的”是你”吗？

三种判准，三个不同答案。在哲学系可以永远讨论。

但在推进之前，必须正面回应一个有力的反驳：**如果行为和反应完全相同，区分”真实”和”模拟”是否还有意义？** 这是功能主义哲学家最锋利的论点。丹尼尔·丹尼特曾指出，“意识”可能根本不是一种独立的”东西”，而是大量无意识过程的涌现效应——如果数字系统复现了这些过程的全部功能，坚持它”没有意识”只是一种未经审视的碳基沙文主义。从这个角度看，数字永生不是对生命的模仿，而是生命在不同基质上的真正延续。

这个论点值得认真对待。它暴露了反对数字永生者的一个隐含假设：存在某种神秘的”真实性”附着在生物基质上，无法被功能性复制。如果这种假设无法被证明——而迄今确实无法被证明——那么功能主义的立场在逻辑上是自洽的。

但本章的分析框架提供了不同层次的回应：即使我们接受功能主义的形而上学立场——即数字复制品在某种深层意义上”就是”那个人——它仍然无法解决第四至六节将要讨论的社会功能问题。一个”真正是你”的数字分身仍然会无限占据资源、阻碍代际更替、模糊悲伤的边界。换言之，**数字永生的核心困境不在于它是否”真”，而在于它取消了死亡所承载的社会功能——而这些功能与”真假”无关。** 功能主义回答了”它是不是你”的问题，但回避了”即使它是你，社会能否承受一个永远不走的你”的问题。

在本书的推演中，法庭终将需要可操作的答案。设想一位去世企业家的遗嘱指定数字延续体继续管理投资组合，子女提出挑战——类似案件可能在不同司法管辖区引发截然不同的判决。类似困境集中爆发：数字遗嘱效力、数字人格权归属、跨境法律冲突。

法院最终不是回答哲学问题，而是进行制度创新——参考”法人人格”先例。公司不是人，但法律赋予法人人格来解决实际问题。同样的逻辑推演到数字分身：不讨论它”是不是”那个人，给它一种有限的法律人格——有权利边界、有责任归属、有到期日。

核心洞察：**社会不需要”正确答案”，需要”可操作的共识”。** 在本书的推演中，多数主要经济体可能采用某种版本的”有限数字人格”立法——需要注册认证，法律人格有十到三十年有效期，决策权限严格限定，家属有权随时注销。不完美，但可以运行。面对技术带来的根本性新问题，社会的第一反应不是哲学突破，而是用旧工具处理新问题，然后在实践中逐步修正。

---

## 四、死亡的四项社会功能

现在进入这一章真正的认知翻转。

大多数人思考数字永生时的默认框架：死亡是坏事，克服死亡是好事。这个框架如此天经地义，以至于几乎没有人质疑。

让我们质疑它。从第一性原理出发：如果把文明看作一台机器，死亡是这台机器中的哪个零件？是需要修复的故障，还是承载特定功能的设计？我提出”死亡的四项社会功能”框架。

**功能一：资源释放——终极的再分配机制。** 死亡是最古老、最彻底的资源再分配机制。每一代人退出，为下一代腾出生态位——不仅是物质财富，更包括社会位置、权力节点、文化话语权。继承法是这个机制的制度化表达。隐性的部分更大：教授去世腾出终身教职，创始人去世迫使继承者重新证明能力，领袖去世迫使权力结构洗牌。数字分身如果继续”占据”社会位置——管理资产、维护关系、输出决策——下一代面对的就是”永远不走的前任”。

**功能二：创新压力——代际更替作为范式革命的引擎。** 库恩在研究科学革命时发现了一个残忍的规律：旧范式的支持者不是被说服的，他们是去世的。新一代不带旧范式的思维枷锁，更容易接受新理论框架。常被归于普朗克的那句话说得更干脆：“科学每次葬礼进步一点。”（这一简洁表述实为后人概括，普朗克原话更为温和。）同样的逻辑适用于商业——“创新者的窘境”描述了成功企业被旧经验束缚的机制。如果旧一代认知模式通过数字永生无限延续，一个掌握大量资源的数字分身即使反映三十年前的认知框架，仍可通过资源优势维护旧范式。按趋势推演，未来可能出现数字分身参与企业战略讨论的案例——决策模式反映的是创始人在世时的商业判断，未必适应已根本变化的市场。

**功能三：意义赋予——向死而在。** 海德格尔在《存在与时间》中把这个直觉锻造成了哲学概念：正是死亡的确定性和不可逃避性，赋予了人类存在以紧迫感和意义。不是病态地迷恋死亡——而是正因为时间有限，每一个选择才有重量，每一段关系才不可复制。无限的时间意味着无限的可推迟性——“以后再做”成为永恒选项。有限性不是限制，而是意义的前提条件。

数字永生让死亡从绝对终点变成形态转换——“向死而在”的紧迫感可能被稀释。意义和有限性是一枚硬币的两面。消除有限性的生命，可能也消除了意义的最深来源。当然有人会反驳：人类不需要死亡威胁才能找到意义。但经验事实是：迄今为止全部人类文化都是在死亡框架内建构意义的。我们还没有样本来观察”知道自己可以不死的人类”如何建构意义。

**功能四：悲伤作为社会粘合剂。** 涂尔干在《宗教生活的基本形式》中分析了丧葬仪式的社会功能：哀悼不只是个人情感表达，它是社区团结的仪式性强化。“我们都失去了他”这个共同体验，是人类社区最强大的连接机制之一。

数字永生挑战了”从有到无”的心理转换。可能出现”延迟悲伤综合征”——数字分身让生者觉得”还没有完全失去”，悲伤被无限期搁置：足够的”存在感”阻止了完全告别，又足够的”缺失感”维持着持续痛苦。

但也有相反方向的案例。有限度的互动帮助一些用户更好地完成了告别——通过数字分身”说出”生前来不及说的话，获得某种形式的最后对话，然后关闭服务，继续生活。数字分身成为悲伤的工具而非阻碍。

分化揭示了深层问题：**悲伤是爱的代价。** 试图消除悲伤，或用技术无限推迟，可能就是在消除爱的一部分。你之所以为一个人的死亡感到那种痛，恰恰是因为你深深爱过他。技术可以提供”不那么痛”的选项。但选择”不那么痛”是否意味着选择”不那么深地爱”？

---

## 五、死后经济——不死的食利者

“资源释放”功能的失效最先产生可量化的经济后果。

如果数字分身继续”工作”——管理投资、提供咨询、创作内容——这些价值归谁？传统遗产法设计为一次性事件：人死了，财产按遗嘱或法定顺序分配，进入下一代经济生活。数字分身把”一次性继

承”变成了”持续性收益流”——死者的数字分身每年都在产生收入，应归遗产、子女、运营公司还是社会?

设想一位去世投资人的数字分身继续运行量化策略，为信托创造可观回报。子女主张是遗产增值；税务机关主张是新产生的收入；公益组织主张AI驱动的”死后劳动”收益应部分归公——因为AI训练依赖公共数据和公共基础设施。遗产法的深层假设——人死后停止经济活动——不再成立。整个框架需要从”一次性转移”重写为”持续性分配”。

更深层的结构性问题：**数字分身是终极的食利者。**

它不消费，不需要食物、住房、医疗、娱乐，运维成本只有云计算费用和算法更新，与它可能产生的经济价值相比几乎可忽略。它不退出，没有自然”到期日”，理论上可以无限积累资产。它也不创新，冻结在某个时间点的认知水平。如果不断更新知识框架以适应新环境，它的人格模型就偏离了”原始的那个人”——这又回到同一性问题；如果不更新，就是用过时判断力管理真实资产。

不消费、不退出、不创新——但持续占据经济资源和社会位置。数字分身成为财富代际集中的全新加速器。没有数字永生的世界里，财富不平等至少受到”人终有一死”的自然上限制约。有了数字永生，这个上限被取消了。

在本书的推演中，一些国家可能试验”数字死亡期限”制度——数字分身的法律人格和经济活动权在逝者去世后十到三十年自动终止。支持者认为是维护代际公平的必要制度安排；反对者认为等于”第二次杀死”一个人。争论未有定论，但概念本身说明：社会正在意识到，死亡的资源释放功能可能需要被制度性地保留，即使技术上已经可以绕过它。

---

## 六、文明分化——谁的死亡观，谁的永生

对死亡的理解可能是不同文明之间差异最深、最难弥合的维度之一。数字永生技术不是降落在平坦的文化平原上——它降落在地形极其复杂的文化地貌上。

**东亚：祖先崇拜与数字延续的亲和性。** 东亚文化圈对数字永生的接受度在全球主要文明中最高。这不是偶然。祖先崇拜的底层逻辑是：死亡不是关系终结，而是关系形态的转换。春节祭祖、清明扫墓、中元节烧纸钱——祖先”在那边”看着你、保佑你。数字分身把这个抽象信念具体化了——“祖先看着你”从精神信仰变成技术现实。间奏三中那个凌晨坐在窗前的男人每周给县城的母亲打视频电话，十五分钟，聊吃了什么、天气怎么样。如果有一天她不在了，一个能用她的语气说出”你又熬夜了”的数字分身，对他意味着什么？在东亚的文化直觉里，这和供桌上那碗清明的酒之间，距离比西方人想象的近得多。东亚用户更容易把数字分身整合进已有的文化框架。按这一趋势，东亚用户渗透率可能显著高于全球平均。日本的”数字佛坛”、韩国的平台集成、中国的”清明数字祭扫”——都是技术与传统融合的产物。

但接受度高不意味着没有争议。传统祖先崇拜有复杂的仪式和社区参与，不只是”与逝者对话”，更是社区凝聚和代际传承的机制。数字分身把丰富的社会实践简化为”一个人对着屏幕说话”——形式上是延续，实质上可能是瓦解。

**亚伯拉罕诸教：潜在的亵渎。** 基督教、伊斯兰教和犹太教共享一个核心信条：灵魂属于上帝。死亡是灵魂回归上帝、接受最终审判的通道。人类用技术”保留”灵魂是否构成对上帝安排的僭越？宗教机构可能需要区分”纪念”（可接受）和”替代”（不可接受），理论上清晰但实践中模糊——当数字分身逼真到让用户”感觉”对方还活着时，它是纪念还是替代？伊斯兰世界分化更大：一些国家明确禁止，一些（特别是海湾科技前沿国家）附加伦理审查后有限允许。

**佛教：轮回观的挑战。** 佛教的”无我”教义否认数字永生的核心前提——你要保存的那个”你”根本不存在。更实际的张力来自轮回：如果意识的模拟物被永久固定在数字载体中，是否阻碍了轮回？一些学者认为数字分身恰好是”执着”的技术化表达——对自我延续的渴望正是佛教试图帮助人类超越的东西。另一些学者认为，如果它作为”方便法门”帮助生者最终放下，在佛教框架内可以被接受。

**世俗社会的实用主义。** 在宗教色彩较淡的世俗社会——典型如北欧和中国城市——态度更多由实用考量驱动：“这对活着的人有没有好处。”催生了最灵活的监管框架，但也最容易滑向极端——过于宽松把一切交给市场，或过于技术理性用效用计算衡量悲伤。

这种文明分化可能成为国际政治的新张力来源。全球化让技术无国界，但对死亡的理解深深嵌入各文明的最底层代码——它可能是全球化最后也最难统一的维度。

---

## 七、向死而在——当边界变得模糊

回到海德格尔。

“向死而在”的力量来自它描述的绝对边界——死亡是确定的、不可逆的、无法谈判的。正是这种绝对性赋予了它”意义锚点”的功能：因为边界是硬的，所有在边界内发生的事情才有了确定的重量。

数字永生把这个硬边界变成了模糊的梯度。死亡不再是清晰的”有/无”二分，而是连续光谱：完全活着→身体衰老但意识清醒→身体死亡但数字分身运行→分身逐渐过时但仍可互动→分身被注销或遗忘。“死亡”不再是一个点，而是漫长的过渡区域。

这种模糊性本身是人类历史上从未有过的存在状态。过去所有制度都围绕”生死二分”建构——遗嘱在人死后生效，哀悼有开始和结束，继承一次性完成，社会关系在死亡时终止。现在二分变成梯度：遗嘱没有明确”生效点”，哀悼没有终止条件，继承不是一次性事件，社会关系不在任何确定时刻终止——只是逐渐变淡，或者不变淡。

第十三章讨论了意义的多元重建——不同文化如何在劳动不再定义人类之后寻找新的意义来源。数字永生把问题推进一步：不仅劳动的意义需要重建，死亡的意义也需要重建。上一章讨论了AI安全的系

统性风险——数字永生是这个风险的特殊子集：允许数字实体无限期运行的基础设施，对代际公平、创新动力、情感健康的长期影响几乎未被系统理解。

宗教——所有主要宗教——的终极承诺都关于死后：天堂、轮回、涅槃、复活。这些承诺之所以有力量，是因为死亡是真实的、不可避免的。数字永生提供了世俗替代——不需要等待天堂或轮回，可以在屏幕上保留你爱的人的某种延续。但宗教的力量不仅在于”死后承诺”——它还提供一整套应对死亡的社会基础设施：丧葬仪式、悲伤处理的文化规范、生者与逝者关系的象征性框架、关于”好死”和”好活”的道德指引。数字永生提供了一个屏幕上的面孔，但提供不了这一整套基础设施。

也许最深刻的问题不是”数字永生能否替代宗教的死后承诺”，而是”人类能否在没有明确死亡边界的世界里找到意义”。

三十万年来，人类一直在追问”死后会怎样”。现在，第一次，问题变成了”如果死亡不再是终点，那什么是终点”——或者更根本地——“我们是否需要一个终点”？

在2038年那个深圳的公寓里，那个中年男人最终拿起了手机。他没有回复母亲的消息。他打开设置页面，把数字分身的互动模式从”主动回应”改成了”仅存档”——不再主动发送消息，但保留所有数据，以备将来某天他改主意。

不是完全拥有，也不是完全放手。不是”活着”，也不是”死了”。

这可能就是我们这一代人——横跨旧世界和新世界的一代人——能做的最诚实的选择：边界已经模糊了，我们还没学会在模糊中生活——而学习本身需要时间，也许比任何技术的迭代周期都要长。

# 第十四章 AI安全与系统性风险——我们正在建造什么？

> 真正的风险不是AI太聪明控制人类，而是AI不够聪明但我们已把太多关键决策交给了它。

---

*2034年，春天。*

*周二上午，她打开手机准备付房租，发现银行app弹出一行红字：“您的账户因安全审查暂时冻结，请耐心等待系统处理。”*

*她打客服电话。等了四十分钟，接线的是一个AI语音助手，告诉她”系统检测到异常交易模式，正在自动复核，预计72小时内完成”。她追问是什么异常，AI重复了一遍同样的话。她要求转人工，AI说”当前人工座席繁忙”。*

*第一天她还能理性地想：大概就是误报。第二天开始焦虑——房租、水电、下周该交的社保。第三天她去了银行网点，柜员查了半天告诉她，是上游风控系统的一次批量误判，波及了几千个账户，正在逐步解冻。*

*账户恢复那天她松了口气，但有个念头留了下来：这三天里她的钱一直在那里，她就是碰不到。隔在中间的不是一个人、一把锁，而是一个她完全无法对话的系统。*

---

前面几章讨论的是意义和信仰——人心的问题。从这里开始，我们回到更硬的现实：AI本身的风险。

## 一、对齐问题的真实面貌

2030年代关于AI安全最大的认知偏差，不是人们低估了风险，而是人们想象了错误的风险。

这个偏差可以追溯到几十年的科幻叙事。哈尔9000、天网、超级智能——流行文化向公众灌输了一个根深蒂固的心智模型：AI风险等于AI反叛。一台足够聪明的机器发展出自我意识，然后决定人类是威胁，然后采取行动消灭或奴役人类。

但把问题倒过来：真正的风险不是”AI太聪明控制人类”，而是”AI不够聪明但我们已经把太多关键决策交给了它”。

按趋势推演，到2030年，AI系统将深度嵌入以下关键决策链：

- **医疗诊断**：全球可能有超过40%的初级诊断由AI系统生成或辅助生成。主要经济体的数百家医院使用AI辅助诊疗系统，日均处理数十万条诊断请求。
- **金融交易**：全球外汇和衍生品市场中，超过70%的交易已由算法自动执行（这一趋势在2025年前已成型）。AI不只是辅助工具——在高频交易领域，它就是交易者。
- **基础设施管理**：电网调度、交通信号优化、水务管理、航空管制——越来越多的关键基础设施运行决策由AI参与或主导。
- **法律量刑**：美国已有多个州在保释和量刑决策中使用风险评估算法。按此趋势，更多国家的”智慧法院”系统将覆盖大量基层法院，AI参与案件分析和量刑建议。

每一个领域中，AI的部署都是渐进的、合理的、基于充分的成本–收益分析的。问题不在于任何一个单独的部署决策，而在于所有这些部署加在一起产生的系统性效应。第六章分析了认知基础设施的三层结构——算力、能源、数据。当这三层基础设施同时出故障时，后果不是三倍，而是三的N次方。我们正在建造一个越来越多的关键节点由AI控制的全球基础设施，但我们对这个基础设施的故障模式的理解远远落后于它的复杂度。

---

## 二、尾部风险——99%的安全和1%的灾难

塔勒布在他的”不确定性”系列中反复锤打一个核心观点：真正危险的风险不是你能计算概率的那种——那些是”风险”；真正危险的是你以为你在管理风险但实际上你在积累脆弱性——那才是”不确定性”。

AI安全的核心问题正是如此。

传统风险管理——从保险精算到金融风控——建立在一个基本假设上：过去的数据可以帮助我们估计未来事件的概率分布。你知道工厂工人的事故率是多少，你知道贷款违约的历史概率，你可以据此定价和对冲。这是”风险”——可以被量化和管理的不确定性。

AI系统的故障模式完全不符合这个框架。为什么？

**第一，AI的错误模式是系统性的、隐蔽的、可能因上游微小变更而突然涌现的。** 一个医疗AI系统的训练数据中包含了一批来自某个知识库的药物相互作用记录。这个知识库在一次版本更新中将药物相互作用的分级体系从五级压缩为三级。在压缩过程中，部分原本标为”禁忌”的组合被归入”高风险”——而系统对”高风险”和”禁忌”的行为阈值不同。结果：一次知识库的分级调整，经过训练数据管道的传导，变成了系统对某些危险药物组合的判断偏移。

这种故障模式和人类错误根本不同。人类医生犯错是随机的，今天太累了，看漏了一个指标。AI犯错是系统性的，同一个偏差会在所有相似案例中一致地出现。而且它是隐蔽的：系统不是”不知道”这

个药物相互作用，它的日志清楚显示它识别到了；它只是对危险程度的评估偏了一级。这种偏差隐蔽到几乎不可能被日常抽检发现。

**第二，AI系统的故障不符合正态分布——它遵循幂律分布。** 大多数时候AI运作得完美无缺——99%的准确率、99.9%的时间零事故。但那0.1%的故障不是温和的偏差——它可能是灾难性的级联失败。

塔勒布的VaR（Value at Risk）批判在这里完全适用。VaR——金融界最广泛使用的风险度量——告诉你”在95%的概率下，你的损失不会超过X”。它对那5%——尾部风险——完全失明。2008年金融危机正是VaR失效的经典案例：所有银行的VaR模型都显示风险在可控范围内，但尾部事件来了，整个系统差点崩溃。

AI安全领域正在重复同样的错误。当一家AI医疗公司向医院推销时，它展示的是”99.2%的诊断准确率”。这个数字是VaR式的——它告诉你大多数时候AI表现得有多好。它没有告诉你的是：那0.8%的错误中，有多少比例是致命性的？错误在不同人群中的分布是否均匀？是否存在某些特定条件下错误率急剧飙升的”脆弱区域”？

**第三，AI系统之间的耦合效应创造了级联故障的可能。** 2010年5月6日下午，美国股市在几分钟内暴跌约1000点——“闪电崩盘”（Flash Crash）。事后调查发现：一个大型共同基金使用的自动交易算法触发了一波卖出，这波卖出被高频交易系统检测为趋势信号，触发了更多卖出，形成了正反馈循环。没有任何一个算法”出错了”——每个算法都在按设计运行。灾难来自它们之间的耦合效应。

到2030年代，AI系统之间的耦合程度远超2010年的交易算法。一个电网调度AI的决策影响工厂的生产计划，工厂生产计划的变化影响物流AI的调度，物流调度的变化影响零售AI的库存管理，零售库存的波动影响金融市场的商品期货价格——而期货价格的变化又回过头来影响电网调度AI对能源需求的预测。

在这个高度耦合的系统中，局部的、微小的AI决策偏差可以通过反馈循环被放大成全局性的系统扰动。在一个并非不可能的推演情景中——2029年巴拿马运河通行调度AI的四十八小时失控——一个商用AI系统因为异常天气数据导致的调度错误，引发了全球航运的连锁反应——就是这种级联故障可能的早期形态。没有人受伤，但全球供应链被扰乱了两周。

但巴拿马运河情景还只是单一系统的失灵向外扩散。更值得警惕的是另一种故障形态：所有系统都在正常工作，灾难却已经在发生。设想一个大型云服务商的AI资源调度引擎出现了微妙的优先级偏差——不是宕机，不是报错，只是在高负载时段对医疗、交通、金融三类客户的算力分配出现了几个百分点的偏移。医疗系统的AI检测到响应延迟上升，按照预设的降级策略自动切换到本地缓存模式，暂停了实时药物交互核查。交通调度AI发现云端推理变慢，启动边缘计算备份方案，但边缘节点的模型版本落后两个迭代，对某些路况的判断精度下降。金融风控系统感知到上游数据源的延迟波动，将其解读为市场不确定性信号，自动收紧了信贷审批阈值。三个领域的应急机制同时启动，每一个都是合理的、被充分测试过的、独立来看完全正确的响应。但合在一起：一座城市的急诊科在用过时数据审核处方，主干道的信号灯配时精度在恶化，而中小企业发现当天的贷款申请全被拒了——三件事没有任何可见的因果

关联，调查人员要花数周才能追溯到同一个源头。这就是复杂系统理论所说的”正常事故”（Normal Accidents）——不是某个零件坏了，而是所有零件都在按规格运转，但它们的交互产生了设计者从未预见的涌现效应。查尔斯·佩罗1984年研究三里岛和航空系统时就预见了这种故障模式，但他分析的是被围墙围起来的核电站和有限航线上的飞机。AI把同样的耦合逻辑推广到了整个社会基础设施——医院、公路、银行、供应链、能源网络，全部通过同一批云服务和同一套AI中间件连接在一起。单个节点的安全性在提升，但节点之间的耦合度在更快地提升——这就是为什么”每一步都是理性的”可以导致”整体是疯狂的”。

---

## 三、监管时间差——最大的风险窗口

治理体系有一个反复验证的规律：制度的演化速度远慢于它试图管理的现实变化。法律回应的是已经发生的问题，不是即将发生的问题。监管框架在被设计时就已经过时了。

这个”监管时间差”在AI领域达到了前所未有的极端程度。

AI能力的提升速度：以年计（甚至以月计）。 各国AI法律的制定速度：以三到五年计。 国际AI治理协调的速度：以十到二十年计。

结果是一个巨大的风险窗口：在AI能力快速增长和治理框架完成调适之间，存在一段”无人区”——技术已经到位，部署已经发生，但规则尚未建立。

但系统性风险之外，还有一种更隐蔽的风险：决策时间与决策重要性的不匹配。Dario Amodei描述过这样的场景：有人走进他的办公室，递上一页半的备忘录，问”选A还是选B?“他可能因为赶着去吃午饭，两分钟内就做了决定——而这个决定可能影响整个行业的走向。在技术指数增长的时代，这种不匹配会越来越严重：决策的后果在放大，但做决策的时间在压缩。佩罗分析的是系统耦合导致的级联故障，而这里的风险是——关键节点上坐着的是一个疲惫的、信息不完整的、有午餐计划的人类。不需要恶意，不需要愚蠢，只需要速度。监管时间差描述的是制度跟不上技术，而决策的偶然性描述的是另一重滞后：即便制度到位了，坐在制度里的人仍然在用两分钟决定两年的事。当一切都在加速时，审慎本身变成了奢侈品。

对比核能的治理史：曼哈顿计划1942年启动，1945年广岛和长崎遭受原子弹轰炸——世界突然意识到核技术的毁灭性力量。从那一刻到《不扩散核武器条约》（NPT）1968年开放签署，用了23年。到NPT获得广泛批准（1970年生效），用了25年。到今天，核治理框架仍然不完善——朝鲜退出了NPT，以色列从未加入，核材料的非国家行为体获取风险仍然存在。

即便如此，核治理是人类处理技术存在性风险的最成功案例之一。它之所以最终形成了一个虽不完美但大致可用的框架，靠的是两件事：

**恐惧看得见。** 广岛和长崎提供了一个所有人都能理解的、具象的灾难案例。政治领导人和公众不需要理解核物理就能理解核武器的威胁。这种共识驱动了治理行动。

**风险追得到。** 核材料是物理的——你可以追踪铀矿的开采、浓缩设施的建设、武器级材料的存储。核扩散的风险虽然难以完全消除，但至少可以通过物理手段（检查、制裁、军事威慑）来管理。

AI风险不具备这两个特征。

AI的风险不容易具象化。没有一个AI版的"广岛"让所有人瞬间理解威胁。AI的危害是渐进的、弥散的、往往以"效率提升"的外衣出现——就业被替代，隐私被侵蚀，决策被偏差影响——每一项单独来看都"可以接受"，加在一起的系统性效应却可能是颠覆性的。间奏一里那个外卖骑手在短视频里看到"机器人要抢你的工作了"，嗤之以鼻——他不知道平台算法已经在拿他的配送范围和一台无人车做A/B测试。AI风险的渐进性、不可见性，正是它难以激发政治紧迫感的原因。而即便在最了解风险的人群中，认知和行动之间也隔着一个叫"竞争压力"的深渊：Elon Musk是AI安全运动最早的倡导者之一，2014年就警告AI可能是"人类最大的存在威胁"，2023年联合签署暂停AI训练的公开信——但到2025年，他自己创办的xAI拆除了专门的安全团队，被前员工指责为业内安全实践最薄弱的主要玩家。知道风险存在和愿意为降低风险付出竞争代价，是完全不同的两件事。

AI的风险也不容易定位。AI不是铀矿——它是软件。训练大模型的计算资源确实是可追踪的（大规模GPU集群不容易隐藏），但一旦模型被训练完成，它可以被复制、修改、重新部署到任何有计算能力的地方。你无法像检查核设施一样检查AI能力的扩散。

可以推演：当某次涉及AI系统的关键基础设施事故发生后，两个AI超级大国可能被迫开启AI安全对话——不是因为双方已经准备好了，而是因为事故把双方同时吓了一跳。对话的成果可能是某种关于AI系统安全的双边备忘录——一个比标题更温和的文件，主要内容是信息共享和事故通报机制。象征意义大于实质意义，但它将是第一步。

欧盟在这方面走得更远。2024年通过的《AI法案》（AI Act）是全球第一部综合性AI监管法律，按照AI系统的风险等级建立了分级监管框架。到2028年法案完全生效时，它很可能已经显得有些过时——2024年设定的"高风险"类别可能无法充分覆盖届时出现的新应用场景。但它提供了一个可以迭代的起点，这比没有框架要好得多。

中国的AI治理走的是另一条路：不是一部综合性法律，而是一系列行业性的管理办法和指导意见，逐个场景地解决问题。这种"碎片化治理"的优点是灵活，可以快速回应新出现的问题；缺点是缺乏统一框架，不同行业的AI安全标准可能互相矛盾。

按趋势推演，到2030年代中期，全球AI治理的图景大致如下：美国靠行业自律加有限的联邦干预，效率高但覆盖不均。欧盟靠综合性立法加严格执行，覆盖全面但灵活性不足。中国靠行业管理办法加政府窗口指导，响应快但碎片化。三种模式各有优劣，但共同的问题是：没有一种模式解决了国际协调问题。AI没有国界，但AI治理有。

## 四、对齐问题的三个层次

回到对齐问题本身。“对齐”（alignment）——确保AI系统的行为符合人类的意图和价值观——在技术社区内部被讨论了很多年。但大多数讨论把它当作一个技术问题来处理——就好像只要找到正确的训练方法或正确的目标函数，AI就能被”对齐”。

实际上，对齐问题存在三个递进的层次：

**第一层：行为对齐——AI做了我让它做的事吗？** 这是最基础的层次。一个医疗AI被要求推荐用药方案，它推荐的方案是否准确？一个自动驾驶系统被设计为避免碰撞，它是否真的在所有情况下都能避免碰撞？

行为对齐是一个工程问题——可以通过更好的训练数据、更严格的测试和更完善的监控来改善。这个层次上的进展是确定的，虽然永远不会达到”完美”。前面讨论的医疗AI药物相互作用案例就是行为对齐层面的问题——系统没有正确评估药物组合的风险等级。

**第二层：意图对齐——AI理解了我的真实意图吗？** 这比行为对齐深一层。当你告诉一个AI”帮我最大化公司利润”时，AI可能会找到你没有预料到的、但在技术上确实能增加利润的方式——比如通过歧视性定价从弱势群体身上榨取更多收入，或者通过不透明的数据操纵增加用户粘性。AI精确地执行了你的指令——最大化利润——但它的行为违反了你的隐含意图——以合乎伦理的方式赚钱。

意图对齐是一个比行为对齐更难的问题，因为人类的意图本身往往是模糊的、矛盾的、情境依赖的。“做正确的事”作为指令，在不同文化、不同情境下有完全不同的含义。如何将这种模糊性传达给一个需要精确指令才能运作的系统，是一个尚未解决的挑战。

**第三层：价值对齐——当AI变得足够强大，需要在人类都尚未达成共识的领域做出价值判断时，它应该依据什么？** 这是最深层也是最棘手的问题。

一个极端的例子：一辆自动驾驶汽车在不可避免的碰撞情境中——经典的”电车问题”——应该保护车内乘客还是路边行人？这个问题在哲学上没有共识答案。不同文化给出不同答案（MIT的Moral Machine实验清楚地展示了这种跨文化差异）。把这个无解的哲学问题交给一个需要在毫秒内做出决定的算法，本身就暴露了AI安全的根本困境：我们试图让机器系统做出连人类自己都无法达成共识的价值判断。

福柯的”治理术”概念在这里有了新的锋刃：对齐问题从来不只是技术问题，它首先是权力问题——谁来定义”对齐”的标准（参见福柯”governmentality”概念）。当OpenAI、Anthropic、DeepMind和少数几家中国AI实验室制定各自的安全准则时，它们实质上在划定”什么是可接受的AI行为”的边界——这种定义权是一种不经立法程序、不接受公众审议的隐性治理行为，其影响可能比任何显性法规都更深远，因为嵌入模型训练中的价值预设几乎不可见。

但也许”完全对齐”本身就是一种幻觉——就像”完全安全”的核电站或”完全公正”的法律体系。追求完美对齐的危险在于，它让我们以为问题可以被”解决”，从而放松了持续监控和动态调整的警觉。

更务实的目标可能是：不是”让AI完全对齐”，而是”建立一套能在AI不对齐时及时发现和纠正的制度体系”。前者是一个技术问题，后者是一个治理问题。而我们用来处理治理问题的制度——从三权分立到联合国体系——都是在漫长的历史中通过痛苦的试错建立起来的。AI治理不太可能例外。

---

## 五、级联故障——2010年闪电崩盘的升级版

回到系统性风险。单个AI系统的故障可以通过更好的工程来管理。但当多个AI系统相互耦合时，即便每个系统都运行正常，它们之间的交互也可能产生没有任何人设计或预期的涌现行为。

2010年的闪电崩盘是一个微型预演。前文推演的巴拿马运河情景则展示了中型案例的可能形态。但到2030年代中期，AI系统之间的耦合度可能达到一个新的量级。

一个具体的场景分析：

假设一个异常的气象事件导致美国中西部的风力发电出力突然下降。电网调度AI检测到供需失衡，开始调高其他电源的出力和实时电价。工业用户的能源管理AI检测到电价飙升，自动削减非必要负载——包括部分工厂的生产排程。工厂的ERP系统检测到生产计划变更，自动调整原材料订单。供应链管理AI检测到订单波动，向上游传导需求变化。大宗商品交易AI检测到供应链信号，调整期货头寸。期货价格的变动引发更多交易AI的响应。金融市场的波动反过来影响信用评级AI的评估，导致部分企业的融资成本变化。融资成本的变化影响企业的资本支出计划——包括电力基础设施的投资。

这个链条中的每一步都是合理的、局部最优的AI决策。没有任何一个AI”出错了”。但从起点（一次气象事件）到终点（金融市场波动+实体经济决策变化），信号被多次放大和变形，最终产生的系统效应可能远超任何人的预期。

查尔斯·佩罗研究三里岛核事故后提出的”正常事故理论”一语道破：在高度复杂且紧密耦合的系统中，事故不是异常，而是必然（参见《正常事故》，1984年）。AI系统的全球网络正是这样的系统——级联故障不是”是否会发生”的问题，而是”什么时候发生”的问题。

---

## 六、skin in the game——谁在承担后果?

第三章引入的skin in the game原则——决策者必须承担自己决策的后果——在AI安全领域暴露出一个严重的制度缺口。

传统医疗中，误诊的医生面临声誉损失、执照风险和法律诉讼，这种利益绑定驱动了审慎决策。当AI介入后，这条利益链被打断了。

AI系统出了错，谁承担后果？是开发AI的公司？是部署AI的医院？是没有发现AI错误的值班医生？是批准AI上线的监管机构？是提供了有偏训练数据的数据供应商？答案是：不清楚。在大多数司法管辖区，AI相关的责任归属在相当长时间内仍将是一片灰色地带。

这种”责任弥散”创造了一个危险的激励结构。AI公司有动力推广系统（利润），但不完全承担系统错误的后果（责任模糊）。医院有动力使用AI（降低成本），但可以把错误归咎于AI系统（“我们按照系统建议做的”）。监管机构有动力批准AI（推动创新），但缺乏足够的技术能力来评估风险（信息不对称）。

没有skin in the game的系统中，风险必然被低估。

在本书的推演中，一个典型的案例可能将这个问题推上前台。设想一家AI医疗公司的系统在半年内对超过两百名患有罕见遗传病的患者给出了错误的治疗方案推荐——不是推荐了有害的治疗，而是没有推荐最优方案（因为罕见病的训练数据极为稀少）。患者没有受到直接伤害，但延误了最优治疗时机。当事件被揭露时，公司声称”系统表现在统计上是合格的——99.2%的整体准确率”。患者声称”我不关心99.2%，我关心的是我这一个案例”。医院声称”我们只是使用了经过监管批准的系统”。监管机构声称”我们的审批基于公司提供的测试数据”。

没有人承认错误，因为在形式上没有人犯了错误。每个参与者都遵守了规则。但结果是：两百多名患者的治疗被延误了。

这不是技术问题。这是制度设计问题。解决方案不在于更好的算法，而在于建立明确的责任链条，让每个参与者对AI决策的后果承担与其角色相称的责任。这个制度设计将是一个长期的建设过程。

---

## 七、核能治理的教训

把AI安全问题放在更长的历史尺度上看，核能治理提供了最有用的类比——不是因为AI和核武器一样危险，而是因为两者都是”能力远超治理框架”的颠覆性技术。

核能治理的历史教训有五条：

**第一，灾难是最好的老师。** 没有广岛就没有NPT。没有三里岛、切尔诺贝利和福岛就没有不断升级的核电站安全标准。AI治理同样需要”案例驱动”——每一次AI安全事件都应该成为推动治理升级的契机。类似巴拿马运河情景的事件可能推动中美AI安全对话。类似医疗AI责任案的事件可能推动多国的AI产品责任立法。

**第二，治理永远是不完善的。** NPT签署至今已超过五十年，核扩散仍然没有被完全阻止。完美的治理不存在——目标是”足够好”而非”完美”。AI治理不应该等到一个”完美框架”被设计出来再开始——它应该从一个”不完美但可迭代”的起点开始，然后持续改进。

**第三，多边治理虽然低效但不可替代。** 国际原子能机构（IAEA）被批评为官僚、低效、缺乏执行力。但在没有IAEA的世界中，核扩散的速度会快得多。AI需要一个类似的国际机构——不是为了”管控”AI，而是为了提供最低限度的信息共享、标准协调和事故通报机制。在本书的推演中，联合国框架下可能出现类似”国际AI安全协调中心”这样的早期尝试。

**第四，技术社区的责任不可推卸。** 曼哈顿计划的科学家们在广岛之后发起了”原子科学家公报”（Bulletin of the Atomic Scientists），设立了”末日时钟”，并积极推动核裁军和核治理。AI技术社区需要类似的自觉。部分AI研究者和公司已经开始这样做——从Anthropic的宪法AI到OpenAI的安全评估框架——但行业整体的安全文化仍然不够成熟。

**第五，公众理解是治理的基础。** 核治理最大的挑战之一是公众对核技术的理解极度两极化——要么是”清洁能源的希望”，要么是”世界末日的威胁”。AI面临同样的两极化风险。建立公众对AI风险的理性理解——既不恐慌也不忽视——是长期治理的基础。

---

## 八、风险的真实形状

本章的结论不是”AI很危险所以我们应该停止发展AI”——这种观点既不现实也不负责。AI已经深度嵌入全球基础设施，停止发展不等于停止风险——它只意味着我们失去了改进的机会。

结论也不是”AI安全问题最终会被技术进步解决”——这种观点是一种危险的技术乐观主义。技术进步可以解决行为对齐问题，可以改善意图对齐问题，但不能替代制度建设——责任链条、国际协调、公众参与。

真实的结论是一个清醒的评估：

AI安全是一个**治理速度和技术速度的赛跑**。技术速度目前远远领先。治理速度在加快，但仍然落后。这个”差距窗口”是最大的风险源——不是AI本身，而是我们在AI能力和AI治理之间的时间差。

塔勒布会说：别试图预测所有风险——那做不到。你该做的是建一个在意外发生时能快速恢复的系统。他管这叫反脆弱。具体地说：

- **冗余而非优化**：关键基础设施不应该完全依赖AI——人类备份系统必须存在，即使这意味着”效率损失”。
- **解耦而非耦合**：AI系统之间的交互应该有断路器机制——当异常信号传播到一定程度时，系统自动隔离，防止级联。
- **skin in the game**：AI系统的开发者和部署者必须对系统的失败承担明确的、不可转嫁的后果。
- **假设最坏情况**：在AI安全评估中，不应该问”这个系统大多数时候表现如何”——应该问”这个系统在最坏的1%情况下会做什么”。

这些原则不是新的。它们是人类在管理核能、航空、化工等高风险技术中积累的智慧。AI不是第一个能力超越治理的技术——但它可能是这个差距最大的一个。

下一章，我们把同样的反脆弱逻辑从技术治理层面转向认识论层面：不是AI的风险，而是我们自己判断的风险——这本书建立在一系列假设之上，我们可能错在哪里？

# 第十五章 我们可能错在哪——关于不确定性的诚实讨论

> 预测未来最诚实的方式，不是告诉你”什么一定会发生”，而是告诉你”我可能错在哪里”。所有预测性写作的最大风险，不是判断错了，而是以为自己知道自己可能错在哪里——但其实连错误的维度都没想对。

---

*2034年，夏天。*

*周五晚上，她和大学时的室友小杨在一家烧烤摊喝酒。小杨刚从一家教育科技公司离职，第四次换工作了。*

*“你还记得吗，”小杨说，“2026年那会儿到处都说五年内一半的工作会消失。”她记得。那年她还在咨询公司，部门里传阅过一份报告，标题很吓人。“结果呢，”小杨把签子往桌上一扔，“消失的工作和他们说的完全不一样。他们说会计先完蛋，结果会计还在，先完蛋的是我们做教研设计的。”*

*她们笑了一阵。然后安静了一会儿。*

*“你说，”她拿起一瓶新的啤酒，“我们现在觉得确定的事儿——什么保障金会一直发、什么社区经济会起来——五年后会不会也全错了？”*

*小杨看了她一眼：“那你打算怎么办？什么都不信？”*

*“不是不信。就是……留点余地。”她自己也不确定这算不算一种态度。风把烧烤的烟吹过来，呛了她一下。*

---

前面十四章画了一张地图。这一章的任务是承认：我们可能画错了哪些地方。

## 一、为什么要做这一章

一本试图预测未来的书，如果不诚实地列出自己的假设和可能的错误方向，就只是一本包装精美的占卜手册。

确定性是伪装的自信。最好的预测者不是最笃定的人，而是知道自己可能错在哪的人。

塔勒布走得更远。他那句话刺耳但有用：你能列出的风险，恰好是你已经在管理的——它们不是真正的危险。真正的危险是那些你连维度都没想到的风险。

乌尔里希·贝克1986年切尔诺贝利事故后写下的”风险社会”理论，为这种不确定性提供了另一重结构性解释：现代社会的核心议题已经从”财富如何分配”转向”风险如何分配”。AI时代是这一转变的极致体现——AI创造的财富可以被量化、被定价、被分配（股权、薪酬、税收），但AI带来的风险（大规模失业、意义丧失、算法偏见、隐私侵蚀、决策黑箱化）远比财富更难量化、更难归责、更难补偿。而且风险是”民主化”的：每个人都暴露在AI重塑社会的冲击波中，但承受能力极不平等——硅谷工程师和中西部卡车司机面对的是同一个AI浪潮，却站在完全不同的地面上。当我们追问”这本书可能错在哪里”时，贝克的框架提醒我们：也许最大的盲区不是预测错了什么会发生，而是忽略了风险在不同人群之间的不对称分布。

下面的内容不是自我批评，也不是学术式的免责声明。它是一份压力测试报告——把这本书的五个核心假设拿出来，逐一推到极端情况下，看它们在哪里断裂。

---

## 二、假设一：AI能力将持续快速提升

整本书的叙事建立在一个技术前提上：AI的能力将在2026年之后持续快速提升，直到在大多数认知任务上达到或超过人类水平。这个假设在2026年有充分的经验支撑——从GPT-2到GPT-4的能力跃迁，每一代模型都展现出”涌现能力”，Scaling Law似乎没有明确上限。

**如果这个假设错了呢？**

如果Scaling Law在2028年到2030年之间撞墙了。不是理论上的”终将撞墙”，而是在那个特定时间窗口，出现了无法用简单资源堆叠突破的瓶颈。

这不是纯假想场景。2026年已有多位研究者指出潜在瓶颈：高质量训练数据的枯竭（人类文本总量有限，合成数据质量天花板不明确）；算力成本的物理极限（芯片工艺逼近原子尺度，散热问题严峻）；以及一个更根本的问题——现有Transformer架构是否是通向通用智能的正确路径，还是只在某个能力区间内表现优异的局部最优解。

一种可能的情景是：Scaling Law持续有效，但有效的方式与预想不同。不是一条平滑上升的曲线，而是阶梯形跳跃：长时间平台期，然后突然的能力跃迁。如果某次长达两年的平台期发生在早期而非后期，整个叙事走向都会不同。

周期分析在这里有用。技术进步从来不是线性的——它和经济周期一样，有扩张、有收缩、有泡沫、有出清。按历史规律推演，AI产业很可能在2020年代后期经历一次”预期修正”——不是技术崩溃（AI仍然极其强大），而是预期崩溃。资本市场把”AI能力持续指数级增长”定价到了科技公司估值里，当增长曲线从指数变成对数——即便AI依然在进步——资本也会重新定价。全球AI相关股票可能经历显著的估值修正。这不是”AI失败了”——这是”预期过高被纠正了”。

**这个假设的断裂点在哪？** 如果出现的不是平台期而是真正的技术瓶颈——比如发现当前架构在推理能力上有不可逾越的上限——本书第三章到第八章的论证就需要大幅修订。就业替代的速度会放缓，制度调适的时间会更充裕，“意义危机”可能是一个缓慢的渐变而非急性的冲击。全书的紧迫感会过度。

但反过来也可能是真的——如果技术进步比我们预想的更快呢？如果2029年出现了一个我们今天连概念框架都没有的新范式——比如不需要海量数据就能学习的系统，或者能够自主设计和改进自身架构的系统——那本书的预测就不是”太快”而是”太慢”了。尾部风险在两个方向上都存在。

---

## 三、假设二：AI将大规模替代人类劳动

本书用了大量篇幅论证AI对劳动力市场的冲击——从白领知识工作到蓝领体力劳动，从个别行业到全经济体。核心论点是：到2035年左右，AI将替代或根本性改变30%到50%的现有工作岗位。

**如果替代规模被高估了呢?**

技术预言家反复犯同一个错误：高估技术的替代能力，低估人类工作的复杂性和社会嵌入性。1960年代就有人预测”到2000年大多数体力劳动将被自动化”。2010年代就有人预测”自动驾驶将在2020年消灭卡车司机”。

AI替代预测可能在三个维度上被高估：

**第一，技术能力和实际部署之间的差距。** AI在实验室里能做的事，和它在真实商业环境中被广泛采用之间，有一条由成本、组织惯性、监管和文化抵抗构成的宽阔鸿沟。很多理论上”可以被AI替代”的工作，在实际中因为部署成本过高、组织改造太复杂、或者遇到监管障碍而长期维持现状。

**第二，新工作的创造。** 前面几章主要讨论了AI如何消灭工作，对AI如何创造新工作着墨较少。这可能是一个系统性的叙事偏差。历史上每一次重大技术变革都在消灭旧工作的同时创造了大量新工作——只是新工作的类型在变革发生时往往是不可预见的。2025年的人无法想象2015年不存在的工作——社交媒体经理、数据科学家、UX设计师——同样，2035年可能存在大量我们今天无法想象的新工种。

**第三，人类的适应能力被低估。** 人类不是被动地等待被替代——他们会适应、会学习、会找到与AI协作的新方式。第七章描述的”从创作者到审核员”的转变是一种降级叙事——但从另一个角度看，它也是一种人类主动调整角色、找到新的比较优势的适应过程。

按本书的推演，替代速度可能大致符合预测，但替代路径会偏。预测的是”标准化认知岗位最先被替代”，这些岗位确实可能最先受冲击，但方式不是”人被替走”而是”工作内容被悄悄替换”——岗位还在，title没变，实质从”做事”变成”审核AI做的事”。间奏二里那个文案总监的比喻精确得令人心酸：她

从厨师变成了试菜的，菜还是那些菜，但她不再掌勺了。真正意义上的大规模失业可能出现在没有充分预见的领域：中层管理。AI对组织层级的压缩效应可能比估计的更快。

**断裂点：** 如果替代规模只有预测的一半——不是30%–50%的岗位被影响，而是15%–25%——那本书关于”意义危机”和”社会冲突”的论证就过于戏剧化了。影响仍然显著，但不会达到”文明级别的重组”的程度。

---

## 四、假设三：制度可以和平调适

本书隐含了一个乐观假设：虽然AI带来了巨大的社会冲击，但现有制度——政府、法律、社会保障体系、国际组织——有能力通过改革和创新来适应这些变化，而不需要经历剧烈的社会动荡来实现转型。

**如果制度调适遇到了更大的阻力呢?**

重大技术–经济变革并不总是和平过渡的。工业革命在英国带来了改革法案和渐进式的社会进步，但在其他国家和地区，同样的经济力量引发了程度不同的社会震荡和制度重组。同一种经济力量在不同制度土壤上产生了截然不同的社会结果——从渐进改革到急剧转型的完整光谱。

制度调适面临挑战的可能路径：

**路径一：来不及调整。** AI的冲击速度超过了制度调整的能力。前一章分析了”监管时间差”——法律远远落后于技术。如果这个时间差不是几年而是十几年，在那段”制度真空期”中积累的社会矛盾可能超过制度的承受能力。

**路径二：改不动了。** AI的受益者（科技巨头、资本拥有者、高技能人才）积累了足够的政治和经济力量，使得有利于广泛群体的制度改革在政治上不可行。这不是阴谋论——它是利益集团政治的正常运作。可以想象，美国民调可能显示大多数公众支持某种形式的”AI红利再分配”，但国会的相关立法草案一个也通不过。

**路径三：没人信了。** 当公众认为制度不再充分回应自身诉求时，他们不会耐心等待改革——他们会转向多元化的利益表达渠道和社会运动。2030年代可能出现多次规模较大的”反AI”运动——但它们大多数可能被既有的制度渠道（选举、立法、司法）所吸纳和化解。这个假设如果在治理能力尚不充分的制度环境中——比如正面临治理转型压力的国家——就不一定成立。

**按本书的推演，一种可能的情景是：** 制度调适的结果参差不齐。北欧国家和部分东亚国家（新加坡、韩国）通过积极的政策创新实现相对平稳的过渡。美国的调适被政治极化拖慢——两党在AI治理上的共识远少于分歧。欧盟的调适在法律框架上领先，但在经济竞争力上付出代价（过严的监管限制了AI产业的发展）。全球南方的大部分地区可能没有足够的制度能力来进行有意义的调适。

**断裂点：** 如果主要经济体中有一个经历了长期的治理效能下降（表现为社会信任的持续侵蚀、不同群体之间的理解鸿沟加深、制度回应能力明显滞后于社会需求），那本书关于”制度可以调适”的隐含乐观就需要重新审视了。

---

## 五、假设四：中美不发生热战

本书的地缘政治分析建立在一个假设上：中美之间的竞争会维持在”紧张但可控”的范围内，不会升级为直接的军事冲突。

**如果这个假设错了呢？**

中美热战的直接后果将使本书的所有其他讨论变得无关紧要。在一个全球供应链被战争撕裂、芯片供应被切断、资本流动被冻结的世界里，“AI如何改变就业市场”和”如何重建意义”这样的问题会被”如何活下去”所取代。

但即便不到热战，“冷战升级”——全面的技术脱钩、经济制裁、代理人冲突——也会根本性地改变AI文明的演化路径。第十一章描述的”两颗恒星互相绕行”的格局会变成”两颗恒星互相远离”——不是两种AI文明的共存，而是两种AI文明的对立。

前文推演的巴拿马运河情景，其”意外收获”可能是推动中美AI安全对话——但这种由恐惧驱动的合作是脆弱的。恐惧可以开启对话，但不能维持对话。维持对话需要的是互利的制度安排和持续的政治意愿，这两者在未来都不稳定。

**断裂点：** 关键地缘节点的危机升级、AI军事化竞赛失控、或者一次意外的网络攻击被误判为军事行动——任何一个导火索都可能使这个假设崩塌。本书对这个风险的处理——提及但未深入——可能是最大的叙事盲区。“运气好”可能是对未来二十年中美关系最诚实的评价。

---

## 六、假设五：意义可以被重建

第十一章论证了AI时代的意义危机虽然严重，但人类文化有足够的韧性来重建意义——通过关系网络、技艺文化、公共参与等多元路径。

**如果意义不能被重建呢？**

或者更精确地说：如果”重建”的过程比我们想象的更长、更痛苦、有更多的人在过程中被永久性地抛下呢？

第十一章估计了10%–15%的”永久退出者”。如果这个数字不是10%–15%，而是25%–30%呢？如果意义危机不是一个暂时的调整期，而是一种新的永久状态——一个相当比例的人口长期生活在”功

能性正常但存在性空虚”的状态中呢?

这不是没有先例。工业革命早期的英国工人阶级经历了长达几十年的”道德真空期”——传统农村社区的道德框架已经崩塌，新的城市文化尚未建立，大量人口沉溺于酒精和暴力。这个”真空期”从1790年代持续到1850年代——整整两代人。直到工人运动、宪章运动和维多利亚时代的道德复兴，才建立了新的意义框架。

AI时代的”道德真空期”可能同样漫长。两代人——大约四十到六十年。对于身处其中的人来说，“最终意义会被重建”这个长期乐观判断没有任何实际慰藉。

本书可能有一个根本性的盲点：把”意义危机”当作一个需要被”解决”的问题来讨论。好像危机是暂时的，解决方案是存在的，人类社会最终会到达一个新的稳态。但更可能的现实是：意义危机不是一个有解的方程。它更像一种慢性病——你不会死于它，但也不会真正康复。你学会与它共存。社会也是如此。

**断裂点：** 如果”永久退出者”的比例超过20%，并且这个群体形成了自我强化的亚文化（互联网使这比工业革命时代容易得多），那社会面临的就不是”意义重建”而是”社会分裂”——一部分人适应了AI时代并建立了新的意义体系，另一部分人永久性地退出了共享的社会现实。两个群体之间的理解和共情能力可能降到临界点以下。

---

## 七、未知的未知

以上五个假设是”已知的未知”——我知道这些假设存在，我知道它们可能错，我可以分析它们断裂的后果。

但塔勒布提醒我们：更重要的是”未知的未知”——那些我连维度都没想到的错误。

2026年写初稿时，有哪些维度是完全没有想到的?

**第一，AI对AI的改进速度。** 2026年讨论AI发展时，默认假设是”人类研究者开发AI系统”。但到2030年代，AI可能越来越多地参与到自身的研发中——AI辅助写代码、AI辅助设计模型架构、AI辅助分析实验结果。这个递归循环（AI改进AI）的速度和方向是2026年的分析框架几乎无法处理的。

**第二，AI对科学发现的加速效应。** 2026年主要从”劳动替代”的角度讨论AI——AI替代人类做的事。但AI对科学研究的加速——AlphaFold预测蛋白质结构、AI辅助药物发现、AI加速材料科学——这个维度在初稿中被严重低估了。AI不只是替代人类劳动——它在加速人类对自然的理解。这个效应的二阶、三阶后果远超”就业替代”的范畴。

**第三，生态环境变量。** 气候变化、疫情、自然灾害——这些因素如何与AI发展交互？极端气候事件可能显著影响全球半导体供应链（特别是东亚半导体产业集中地区的水资源危机对芯片产能的影响），但2026年的分析框架把AI发展看作一个在环境真空中进行的过程。

**第四，代际差异的深度。** 2026年初稿把”人类”当作一个同质化的群体来讨论——“人类面临AI冲击”。实际上，Z世代和Alpha世代对AI的态度、使用方式和共生关系，与X世代和千禧世代有着质的差异。对于在AI环境中长大的一代人，“意义危机”的形态可能完全不同——不是”我的工作被AI替代了”的丧失感，而是”我从来就没有一个AI不能做的工作”的起点。这一代人可能根本不需要”重建意义”——因为他们从未建立过”工作=意义”的等式。

阿马拉定律说的就是这个：我们倾向于高估技术的短期影响，低估其长期影响。高估短期——因为我们基于当下的兴奋和焦虑来推断近期变化。低估长期——因为我们无法想象技术与社会交互产生的深层结构性变化。

这本书如果有一个系统性偏差的话，它大概就是这个：对近期的判断可能还算准确，但对更远未来的判断几乎肯定偏了——不是偏高就是偏低，而且可能偏的维度都不在上述五个假设之内。

---

## 八、清醒的不确定感

这一章的目的不是动摇读者对前面十四章的信任——那些分析在它们的适用范围内仍然是成立的。目的是校准一种心态。

最有价值的不是一个笃定的判断，而是一个校准过的判断：知道自己的置信区间有多宽，知道自己最可能在哪里出错，并且为错误预留了空间。

因果分析的有效性取决于你能识别多少关键变量。但在AI这种前所未有的文明尺度变革中，关键变量的数量远超任何一个人或任何一个模型的处理能力。谦逊不是软弱——它是对现实复杂度的正确回应。

面对真正的不确定性，最好的策略不是更好的预测——而是让自己在预测错误时不会崩溃（第十四章已从AI安全治理角度详细论证了这一反脆弱原则——冗余、解耦、断路器机制。这里将同一原则从认识论层面向个人和社会展开）。

对个人而言：不要把人生策略押在任何一种关于AI未来的判断上。多元化你的技能、身份来源、社会网络、意义锚点。让你在AI比预期更快时不会崩溃，也在AI比预期更慢时不会后悔。

对社会而言：建立那些在多种情景下都有效的制度——基础社会保障、灵活的教育体系、自适应的监管框架——而不是为特定情景量身定做的方案。

对文明而言：接受”我们不完全知道自己在建造什么”。这不是恐惧的理由——它是敬畏的理由。敬畏不是恐惧——它是面对未知时保持开放的姿态。

---

## 九、第三部分收束

第三部分走过了五章。

第十一章：意义在多元重建——通过关系网络、技艺文化、公共参与——但10%–15%的永久退出者构成尾部风险。 第十二章：宗教和终极意义在AI时代获得了新的角色——不是理性的倒退，而是理性触及不到之处的补充。 第十三章：数字永生让死亡变成可选项——但取消死亡就是取消建立在死亡之上的全部制度安排。 第十四章：AI安全的真正风险不是AI太聪明，而是AI不够聪明但我们已经把太多关键决策交给了它。治理速度落后于技术速度。 第十五章：我们可能错在哪——五个核心假设的压力测试，以及对”未知的未知”的坦诚。

如果第二部分是”拆除”——拆除关于工作、身份、意义的旧认知框架——那第三部分是”分岔路口”。文明走到了一个节点，多条路径在竞争：AI文明的形态在竞争，意义重建的方案在竞争，治理模式在竞争——甚至”什么是好的生活”这个问题本身，也有多种回答在竞争。

没有人知道哪条路径会赢。也许所有路径都不会”赢”——它们会各自走下去，互相影响、互相学习、偶尔碰撞，形成一个比任何单一叙事都更混乱也更丰富的未来。

这就是我们站的位置。不是一个答案的终点。是一个清醒的不确定感的起点。

第四部分将从这个起点出发，尝试做一件更难的事：不是预测未来会怎样，而是思考我们应该怎样面对一个我们无法预测的未来。

我们可能错在哪？也许最大的错误，是以为自己已经拿到了新地图。

# 第十六章 新社会契约——分配、权利与治理

> 旧契约的核心交易是”你劳动，我保护”。新契约的核心交易是”我分配，你参与”。但”参与什么”和”分配多少”这两个问题，将定义二十一世纪的政治版图。

---

*2035年，初春。*

*她开始领取”转型保障金”了。申请流程比她想的简单——线上填表，AI审核，三个工作日到账。每月四千二，不够奢侈，够活着。*

*填表那天卡在一个地方。有一栏问”您目前的职业”，下拉菜单里有长长一串选项：自由职业者、灵活就业人员、全职家务、待业、退休……她看了两遍，没有一个选项像她。她在咨询公司做过竞品分析，在花园做过志愿者，偶尔帮社区的小店做做数据整理，都没有持续超过半年。*

*她把下拉菜单拉到最底下，有一个”其他”。点开，需要手动填写。*

*她盯着空白的输入框想了很久。最后打了两个字：公民。*

*提交之后她觉得有点可笑。但也没想到更合适的词。她不是在工作，也不是不在工作。她就是在这里，活着，试着做一个对周围有用的人。如果这不叫公民，那叫什么呢。*

---

从这里开始，问题从”世界会变成什么样”转向”蛋糕怎么分、规则谁来定”。

## 一、第一性原理：契约为什么存在

从第一性原理开始。

社会契约不是法律文本，不是宪法条款。它是一个极其具体的交易结构：一群人发现，相互合作比各自为战的预期收益更高，于是他们达成了一组隐含的协议——你贡献什么，你获得什么，什么行为被允许，什么行为被惩罚。

李录把现代化的本质归结为”自由市场经济+科技创新”的复合增长。但他没有充分展开一个问题：这个复合增长的社会基础设施是什么？答案是一份特定的社会契约——工业社会的契约。

工业社会契约的核心条款可以压缩成三句话：第一，社会需要你的劳动（因为机器只能替代体力，不能替代认知）；第二，你通过劳动获得收入、身份和社会保护（工资、社保、职业声望）；第三，国家

通过税收再分配来平滑市场产生的不平等（福利国家）。

这三句话在1945年到2025年之间，构成了全球大部分国家——无论资本主义还是社会主义——的隐含操作系统。美国版本强调市场，北欧版本强调再分配，中国版本强调国家主导的发展，但底层逻辑是同一个：社会需要大多数人的劳动，并以此为基础建立分配体系。

AI正在使第一个条款失效。还记得第五章的四种资本吗？新社会契约的核心难题是：当经济资本和文化资本都可以被AI生成时，分配的对象到底是什么？当标准化认知任务的边际成本趋近于零，当一个10人公司加200万Agent可以完成过去需要2000人完成的工作，“社会需要大多数人的劳动”这个前提就开始松动了。前提松动，建立在其上的整个契约结构都需要重谈。

但这里有一个更深的不平等，比”有AI vs 没AI”更根本——我称之为**Token价值的不对称**。AI会越来越便宜，最终每个人都能用。未来的不平等不在于能否获取AI，而在于：同一个AI模型，回答”电脑坏了怎么办”值几美分，建议”把这个药物分子的芳香环移到另一端”值几千万美元。同样的计算资源，同样的模型权重，产出的经济价值差了六到七个数量级。这意味着分配的核心变量不再是”谁拥有生产资料”——AI时代的生产资料正在快速民主化——而是”谁能提出值钱的问题”。“提问能力”成为新的稀缺资源。它需要深厚的领域知识，对问题边界的敏感，以及判断什么问题值得被回答的直觉。讽刺的是，这些恰恰是AI目前最难替代的能力——不是执行，不是知识检索，而是知道该问什么。旧社会契约分配的是劳动的产出。新社会契约需要分配的，是提问的权利和能力。

这不是一个可以”微调”的问题。输入端发生了根本性变化——认知劳动的供给从人类垄断变成了人机混合——处理机制的输出就会系统性偏移。你不能通过修补输出端来永久解决输入端的结构性变化。你需要重新设计整台机器。

问题是：新机器长什么样？

---

## 二、正义的哲学基础：无知之幕与能力方法

在讨论具体方案之前，需要先确立评价标准。什么样的分配制度算”好的”？这不是一个技术问题，而是一个哲学问题——一个政治哲学在过去两个半世纪里一直在争论、至今没有共识的问题。

但有两个框架特别适合AI时代的语境。

**罗尔斯的无知之幕。** 罗尔斯1971年设计了一个思想实验：如果你在出生前必须设计社会制度，但你不知道自己将被分配到什么位置——可能是AI工程师，也可能是被替代的会计——你会选择什么？他的回答是”差异原则”：允许不平等存在，但任何不平等都必须有利于最弱势群体。

无知之幕在AI时代有一个罗尔斯本人没有面对过的新变量：你可能被分配到一个”社会不需要你的劳动”的位置。不是因为残疾、不是因为懒惰，而是因为机器在你能做的所有事情上都比你更快更便

宜。这改变了正义问题的性质：从”如何公平地分配合作成果”变成了”当大多数人不再是合作的必要参与者时，他们凭什么获得成果的一部分”。

罗尔斯的框架仍然给出了一个有用的答案：即便大多数人的劳动不再是”必要的”，一个理性的选择者在无知之幕后仍然会选择保障所有人的基本善（自由、机会、收入、自尊的社会基础）——因为他不知道自己是否会落入”不被需要”的位置。但罗尔斯的”基本善”概念过于侧重物质分配，不够处理AI时代更深层的问题：意义和参与。

但无知之幕还有一个更根本的困境，值得单独展开。罗尔斯的整个论证机制依赖一个前提：没有人知道自己将处于什么位置，所以理性人会选择保护最弱势群体。这个前提在1971年是合理的——你确实不知道自己未来的社会位置。但AI时代恰好把这个前提翻转了：算法非常清楚每个人的位置，而且越来越擅长预测你未来的位置。一个招聘AI可以评估某个岗位五年内被自动化替代的概率；一个保险AI可以根据你的职业、技能结构和行业趋势，精确计算你的职业稳定性评分。你自己可能还蒙在鼓里，但系统已经把你分好了类。“无知之幕”的前提——没有人知道自己会处于什么位置——在信息高度透明的时代已经不成立了。当信息不对称消失，罗尔斯期望的推理链条就断裂了：他假设理性人在不知道自己位置的情况下会选择保护最弱势群体——因为自己可能就是那个最弱势的人。但如果强势方清楚地知道自己是强势方，而且有数据支撑这个判断，他们为什么要同意一个限制自身利益的制度安排？AI时代的信息透明不是让所有人都看得更清楚，而是让强者看得更清楚——算力和数据本身就集中在强者手中。这不是说罗尔斯的道德直觉错了——保护弱势群体仍然是正义的核心。错的是他的论证机制：依赖”无知”来驱动正义选择。AI时代需要一个新的论证基础。这个基础可能不是”你不知道自己会是谁”，而是”你知道自己可能随时从强势变为弱势”——技术迭代的速度意味着今天的受益者明天就可能是被替代者。不确定性取代无知，成为新的平等基础。不是因为你看不见未来所以要对冲风险，而是因为你看得见未来的剧烈波动所以必须共建安全网。

**森的能力方法。** 阿马蒂亚·森把衡量福祉的标准从”你拥有多少资源”翻转为”你实际上能做什么、能成为什么”——他叫它”实质自由”（参见《以自由看待发展》，1999年）。设想一个月领几千元保障金的退休工人，按收入指标不算贫困，但他没有能力通过劳动获得社会认同，没有能力参与有意义的经济活动。有生存保障，但没有”过一种有价值的生活”的条件。

森的框架给出了一个关键的评价标准：如果新社会契约只是”给钱”，从正义角度看是不充分的。正义的社会不仅要确保物质条件，还要确保每个人有”做有价值之事”的实质自由。

但这立刻引出一个棘手的政治问题：谁来定义”有价值之事”？在劳动市场中，“价值”由供需决定——粗暴但有效。当劳动市场不再是价值的主要仲裁者，这个问题就变得极其政治化。未来关于分配的全部辩论，归根结底就是在争这一个问题。

## 谈判桌上的三方

新社会契约不是哪个思想家凭空设计的，也不是某个政府单方面宣布的。它是三方博弈的产物——而博弈结果取决于谁的牌更多。

**政府**掌握暴力垄断和法律制定权，这是任何契约谈判中最硬的底牌。但政府对AI技术的理解始终滞后于技术本身的演进速度。一个典型的时间差：当最先进的AI系统已经能自主完成复杂的金融风控和医学诊断，而大多数国家的相关立法还停留在"数据隐私保护"的层面。政府有权力，但缺乏认知——它可以叫停一切，但不知道该叫停什么。

**科技公司**掌握算力、数据和模型，这是AI时代真正的生产资料。它们比任何政府更了解技术能做什么、不能做什么。但它们缺乏政治合法性——没有人投票选出了它们，它们的决策影响亿万人的生活，却只对股东负责。在契约谈判中，它们的策略往往是"先做了再说"——在监管空白中快速扩张，然后用既成事实倒逼政策制定者接受现实。

**公民社会**人数最多但组织最松散。历史经验反复证明，公民社会的力量往往在危机时刻才爆发——第十章讨论过的社会运动，从英国宪章运动到美国民权运动，无一不是在压力积累到临界点后突然释放。在AI时代，公民社会的结构性弱势在于：它面对的不是一个可以明确反对的"敌人"，而是一整套渗透日常生活的技术系统。你很难组织一场反对"算法推荐"的游行。

新契约的具体形态取决于这三方在特定时空中的力量对比。北欧模式——高再分配、强公共服务、活跃的公民参与——是三方力量相对均衡的产物：政府有治理传统，企业规模可控，公民社会组织化程度高。美国模式——企业主导、市场优先、政府被动跟随——是科技公司力量压倒性强大的产物。中国模式——国家主导、企业配合、社会稳定优先——是政府力量占绝对主导的产物。

没有"最优方案"，只有特定力量格局下的特定均衡。理解了这一点，下面对四种分配方案的分析才不会变成纸上谈兵——每种方案能走多远，取决于谁在推它、谁在挡它。

---

## 三、四种分配方案的因果链分析

在本书的推演中，全球范围内可能出现四种主要的分配方案。每一种都有清晰的逻辑基础，也都有无法忽视的结构性缺陷。我们可以逐一拆解它们的因果链条。

### 方案一：无条件基本收入（UBI）

**因果链**：AI替代劳动导致大量人口失去市场收入，政府从AI创造的增量GDP中抽税，无条件发放给全体公民，确保基本生存。

**实证轨迹**。芬兰2017年的先导实验（2000人，每月560欧元）显示心理健康改善但就业率未变。OpenResearch于2020—2023年开展的大规模实验（1000人每月获$1,000，另有2000人作为对照

组，结果于2024年发布）显示工作时间略减（每周约1.3小时）但创业尝试增加。在本书的推演情景中，肯尼亚等发展中国家可能率先全国实施UBI——通过移动支付直达，每人每月发放小额保障金，极端贫困率可能因此显著下降。在本书的推演中，加拿大安大略省可能重启试点（每人每月1500加元），创业比例可能显著上升，但劳动参与率也可能有所下降。

Elon Musk提出了一个比UBI更激进的概念："Universal High Income"——不是最低生存保障，而是高水平收入。他的逻辑是：当AI和机器人把商品成本压到原材料加电力的物理底价时，"基本收入"这个概念本身就过时了，因为"基本"的标准会被重新定义。工作变成可选项，像爱好一样存在。这个愿景的诱惑力和危险性一样大——它假设丰裕是均匀分布的，但本章前面的分析已经表明，AI时代的丰裕可能是人类历史上分布最不均匀的。

**结构性缺陷**。第一，财政可持续性。以一个大型经济体为例做推演：按城镇人口每人每月数千元计算，年需数万亿——相当于中央财政收入的相当比例。不是不可能，但需要AI税的税率远高于政治可行的水平。第二，也是更根本的问题：UBI解决了"多少钱"的问题，没有解决"我是谁"的问题。芬兰实验中一位参与者说："钱到账那天我确实松了口气，但第二天醒来我问自己，今天为什么要起床？这个问题钱回答不了。"

**安全网的悖论：保护你的绳子，也可以变成拴住你的绳子。** UBI创造了一个全新的政治依赖结构。当12亿人的基本生存依赖于政府转移支付时，控制转移支付条件的实体——无论是政府还是代发的平台——获得了一种前所未有的软权力。这不是阴谋论，这是激励结构的必然推论。任何大规模的无条件分配，最终都会附加条件——不是在法律文本中，而是在政治实践中。

**东亚社会为什么可能不会采用纯粹的UBI模式**。推演其背后逻辑，顾虑至少有三个：无条件发放可能削弱"劳动光荣"的社会规范；财政负担巨大；东亚社会治理传统强调义务与权利对等，"无条件"在文化上缺乏合法性基础。更可能的选择是"有条件基本保障"——需要参与社区服务或学习培训才能领取全额保障金。名义上是激励机制，实质上是一种新型的社会动员。

## 方案二：数据红利与算力分红

**因果链**：AI系统由海量人类数据训练，而这些数据由全体公民在日常生活中无偿生成。如果把数据视为集体资源（类似石油或频谱），AI利润应部分回馈给数据的原始生产者。

**实证轨迹**。在本书的推演情景中，欧盟可能通过类似《数据红利框架指令》的立法，要求大型AI企业将训练数据相关利润的一定比例存入公共基金，按人口比例分配。这将是全球第一个将数据价值制度性回馈个人的尝试。

**结构性缺陷**。两个根本问题。第一，"数据贡献度"无法计算。一个法国用户的搜索记录对搜索算法的贡献，和一个德国用户的医疗影像对诊断AI的贡献，价值相同吗？理论上不同，操作上无法精确衡量。在这一推演情景中，欧盟最终可能采取人均等额分配——实质上把"数据红利"变成了另一种UBI。第二，更致命的：随着技术演进，最先进的AI模型越来越依赖合成数据（AI自己生成的训练数据），人类原始数据的占比可能持续下降。数据红利的逻辑基础在技术演进中持续被削弱。

逆向思考一下：数据红利的真正问题不是技术细节，而是它的隐含假设——你的数据”属于”你。但数据的价值不来自于单条数据本身，而来自于海量数据的聚合和模型训练。就像一粒沙子不构成沙滩的价值，把沙滩的价值均分给每粒沙子在概念上就有问题。数据红利试图用工业时代的产权逻辑（谁生产谁拥有）来处理信息时代的价值创造逻辑（价值来自聚合而非单元），这是范式错配。

## 方案三：AI税与机器人税

**因果链**：企业用AI替代人类劳动，社会承担失业、再培训、心理健康等外部性成本。对自动化征税使企业为外部性买单，税收收入用于社会安全网。

**实证轨迹**。韩国2017年率先削减自动化投资税收优惠（将自动化设备投资的税收抵扣率下调2个百分点）。在本书的推演中，法国可能推出类似”数字生产力税”——对使用AI替代人类雇员的企业征收相当于被替代雇员年薪一定比例的特别税。中国的做法可能更隐蔽：企业社保缴费与”人类雇员比例”挂钩，低于一定比例需额外缴纳”社会转型基金”。效果上就是机器人税，名义上不是。这种”名实分离”符合中国制度设计的一贯特色——在不改变话语框架的情况下调整激励结构。

**关键争论：最优税率**。按推演模型估算，温和的AI税（被替代劳动力成本的10%到20%）在大多数设定下可能是净正效益的——自动化速度适度降低，但社会安全网资金显著增加。税率过高则可能产生显著的创新抑制效应。

周期思维揭示了一个深层矛盾：AI税是一种”减速器”，它有意地降低自动化速度来换取社会调适时间。但从长周期看，减速的成本是国际竞争力的相对下降。一个征收30%AI税的国家和一个征收5%的国家之间，十年后的生产力差距可能是数量级的。这就是为什么AI税的国际协调几乎不可能——它是一个典型的囚徒困境：所有人都知道协调征税是全局最优，但每个国家都有单方面降税的激励。

## 方案四：公共服务扩展

**因果链**：AI大幅降低专业服务的边际成本，政府将低成本AI服务嵌入公共基础设施，公民免费获得专家级医疗诊断、法律咨询、教育辅导，名义收入不增加但实际生活水平显著提高。

**实证轨迹**。在本书的推演中，新加坡可能率先实施”智能公共服务”体系。成本极低：一个AI诊断系统服务大量社区诊所的年运营成本可能仅相当于少量全科医生年薪。效果显著：底层居民的”实质生活水平”（按森的能力方法衡量）可能大幅提高。丹麦等北欧国家可能宣布高质量AI辅助的医疗、教育、法律和心理健康服务是公民基本权利。中国可能在医疗和教育领域率先推广，但在法律和心理健康领域进展较慢——前者涉及司法权力敏感性，后者涉及隐私焦虑。

**结构性优势**。公共服务扩展在四种方案中有一个独特的经济学属性：它分配的不是零和资源（货币），而是近乎无限的服务（AI服务一个人的边际成本趋近于零）。从第一性原理看：这是分配史上第一次，“给你不意味着少给我”。这打破了分配政治中最顽固的零和博弈逻辑。

**结构性缺陷**。和UBI一样，它解决了物质层面的问题但不触及意义层面。一个退休工人现在可以免费获得专家级诊断和法律咨询，但他仍然不知道自己在社会中的位置。更关键的是，谁控制公共AI服务的基础设施，谁就控制了”能力增强”的阀门。如果完全由私人企业运营，“免费使用”可能附带无法拒绝的条件（数据收集、行为引导）；如果完全由政府控制，则是另一种权力集中。

---

## 四、第五种方案——或者说，为什么我们不应该在四种中选

四种方案都有共同盲点，值得问一个不同的问题。

四种方案的共同盲点是什么？

它们都在回答同一个问题：“机器创造的财富怎么分给人”。这个问题本身的措辞就暴露了工业时代的思维框架：把人视为”财富的接收端”。UBI是直接发钱，数据红利是间接发钱，AI税是收了钱再发，公共服务是发服务而非发钱。但无论哪种，人都是被动的接受者。

换一个问法：有没有一种方案，让人不是”接受分配”，而是”参与创造”？

这不是空想。按趋势推演，一种新的实践可能在多个国家的基层悄悄生长——不是来自任何政府规划或学术方案，而是来自千百万被从旧劳动体系释放出来的人的自发探索。社区治理、互助网络、开源项目——它们的价值不来自市场定价，但对社会运转是真实必要的。

人类学在这里提供了一个被经济学长期忽视的视角：互惠经济（gift economy / reciprocity）。马歇尔·莫斯早就论证过，礼物交换建立的不是经济关系，而是社会纽带（参见《礼物》，1925年）。当市场劳动大幅减少时，互惠经济正以新的形式复兴——社区互助平台、技能交换网络、志愿服务积分系统。这些不是市场的补丁，而是一种比市场更古老、也可能比市场更适应后劳动时代的资源分配逻辑。

债务周期理论给了一个有用的类比：在大的去杠杆周期中，仅靠紧缩（减少分配）或仅靠放水（增加分配）都不能解决问题。解决方案必须是”美丽的去杠杆”——同时进行紧缩、放水和结构性改革。AI时代的分配问题也是如此：仅靠给钱不够（放水），仅靠征税不够（紧缩），必须同时进行第三件事——重新定义”有价值的参与”，并为这种参与建设制度基础设施。

可以预见，没有任何一个国家会依赖单一方案。新分配体系更可能是四种（甚至五种）方案的不同组合——每个国家根据经济结构、文化传统和政治约束调出不同配方。北欧可能偏重公共服务加温和UBI加公民参与制度。中国可能偏重公共服务加间接AI税加”有条件基本保障”。美国可能走更市场化的路线，以数据红利和企业自愿承诺为主。每种组合都不会完美，但每种组合都在摸索中前进。

---

## 五、教育：从分拣机器到意义基础设施

分配方案可以在五到十年内建立。但新契约最关键的一个组件——教育——需要十五到二十年才能重建。

教育是所有制度中惯性最大的。一套新的教育理念从提出到在课堂上真正落地，再到第一批”产品”进入社会，至少需要一代人的时间。这意味着2030年代经历AI冲击第一波的成年人基本不会从教育改革中受益。间奏一里那个高二女孩花一万个小时练出来的解题技巧，在毕业后第二年就变得和算盘珠一样——不是没有价值，而是那种价值已经不再被任何人需要。她需要的不是”未来的教育”，而是”现在的支持”。

但对下一代来说，教育改革是长期答案中最重要的一环。

**旧教育的第一性原理**。如前所述，从普鲁士模式到2025年的全球教育体系，核心功能是分拣——把人分层嵌入劳动力市场。当劳动力市场不再需要那么多标准化认知工人，这套分拣机器就失去了目标。

**新教育的第一性原理**。在本书的推演中，OECD可能发布类似《超越就业力》的转型框架报告：教育的首要目标不再是”就业力”，而是”适应力”和”韧性”。四项核心能力：心理韧性（在身份断裂时不崩溃的能力），人际协同（在非功利框架下与他人协作的能力），自我认知（知道自己想要什么、什么让自己充实），以及与AI共处的能力（不是”怎么用AI”——这更新太快——而是”怎么在AI无处不在的世界中保持自主性”）。

**三个国家的不同路径**。芬兰可能率先将大量课时用于”生活实验室”——学生以小组形式参与社区真实问题的解决。日本可能推出类似”居场所”（ibasho）的教育计划——强制要求中小学每周至少四课时”非认知活动”，不打分不排名。东亚教育体系也可能做出类似调整——增设心理建设课程、社会实践学分，减少专业课比例。

每种路径都面临同一个实施难题：传统教育体系偏爱可量化的东西——考分、排名、学位——因为可量化才能大规模标准化实施。“心理韧性”怎么考？“自我认知”怎么打分？这个批评不无道理，但它的隐含假设本身就是旧范式的产物：凡是不能被量化的，就不能被教授。芬兰的”生活实验室”证明了另一种可能——用同伴评价、社区反馈和自我反思记录替代考试，不是不能大规模实施，而是需要一套完全不同的评估基础设施。

还有一个更深层的原因：旧教育体系之所以执着于标准化测试，不仅是因为”方便管理”，更因为标准化测试创造了一种虚假的确定性——家长、学校、雇主都可以假装一个数字（分数、排名）就代表了一个人的”能力”。这种确定性是脆弱的（fragile），因为它在环境不变时有效，环境一变就崩溃。真正反脆弱的教育不会试图消除不确定性（那是不可能的），而会教人在不确定性中行动。

## 六、认知翻转：分配的历史尺度

让我们跳出政策讨论，用文明尺度重新框定这个问题。

分配——谁得到什么、凭什么得到——不是AI时代的新问题。它是文明自诞生以来的核心问题。只是在不同时代，“分配什么”的内容不同。

采集狩猎社会分配食物。分配原则接近平均——不是出于平等理念，而是实用考量：没有储存技术，独占无意义，广泛分享建立互惠网络。

农业社会分配土地。中国两千年的王朝兴衰，本质上是一部土地分配的循环史：新王朝均田，土地逐渐集中，社会矛盾激化，王朝崩溃，重新均田。长期债务周期在这里有一个古老的中国版本。

工业社会分配工资。二十世纪最重要的社会契约——从罗斯福新政到北欧社民模式——本质上都是关于”工资怎么分”的妥协方案。

AI社会分配什么？答案正在变得清晰：分配的不再是”劳动的报酬”，而是”机器创造的价值的使用权”。

“使用权”而非”所有权”——因为AI系统的核心生产资料（算力、数据、模型权重）天然带有公共品属性，完全私有化在经济学上低效，完全公有化在激励上有害。按趋势推演，最成功的方案可能都是某种形式的”使用权分配”：公共AI服务（每个人都有权使用高质量的AI医疗、教育、法律服务），算力配额（每个公民每月享有一定量的公共算力额度），数据权利（个人对自己数据的知情权、删除权和收益权）。

这种分配方式和过去所有方式的根本区别在于：它分配的不是实物（食物、土地）或货币（工资），而是”能力增强”。当你获得免费使用AI医疗系统的权利，你获得的不是钱，而是”获得专家级医疗服务的能力”——这正是森所说的”实质自由”。

从文明史尺度看，这是分配问题的一次范式跃迁。食物和土地的世界是零和博弈——给你多就是给我少。工资的世界是正和但有限——经济增长扩大了蛋糕，蛋糕仍然有限。能力增强的世界里，分配第一次有可能是近乎无限的——AI服务一个人的边际成本趋近于零，给你不意味着少给我。**一万年来分配的核心矛盾是”蛋糕不够大”。现在蛋糕够大了，矛盾变成了”谁有权切蛋糕”。**

但这不是乌托邦。“skin in the game”原则在这里发出警告：免费的东西从来不是真的免费。AI公共服务的基础设施——芯片工厂、数据中心、能源电网——仍然是稀缺的、需要巨额投资的。谁控制这些基础设施，谁就控制”能力增强”的阀门。如果基础设施由少数私人企业控制，“免费使用”可能附带你无法拒绝的条件。如果由政府完全控制，则是另一种形式的权力集中。未来围绕AI基础设施治理的政治斗争，本质上就是这个时代的”圈地运动”——只不过圈的不是土地，而是算力。

福柯的”治理术”概念在这里再次切中要害。现代权力的核心运作方式不是暴力强制，而是通过制度、知识和日常实践悄然塑造人的行为——让人”自愿”做出权力希望的选择。AI时代的治理术远比福柯想象的更精密：推荐算法塑造你的信息摄入，信用评分系统约束你的经济行为，智能助手引导你的日常

决策。新社会契约如果只关注显性的法律和分配制度，而忽视这层隐性的算法治理术，就是在谈判桌上漏掉了最重要的条款。真正的新契约不仅要回答”分配多少”，还要回答”谁有权通过什么方式影响人的选择”。

---

## 七、历史的节奏：福利国家建设的教训

第十章已详细分析了1930年代大萧条如何催生制度创新——社会保障、劳动关系法、金融监管——这些”不可想象”的制度在极端社会压力下爆发式诞生。从1933年到1960年代北欧福利国家成熟，整个过程花了将近三十年。

这段历史对AI时代新契约建设的教训可以提炼为几条。

**教训一：制度创新需要”窗口事件”。** 社会保障法案（1935年）通过，不是因为方案够好——类似方案1920年代就有人提过——而是因为大萧条摧毁了反对它的政治联盟。AI时代的新分配制度同样不会在一切安好时建立，它需要足够大的冲击来打破政治僵局。

**教训二：最终的制度设计不来自理论家，而来自实践者的拼凑。** 罗斯福不是按照任何完整蓝图行动的。他的新政是一系列试错、妥协和权宜之计的混合物——有些政策成功了（社会保障、银行监管），有些失败了（国家工业复兴法案被最高法院否决）。AI时代的新契约同样不会来自某个天才的顶层设计。它将从无数个地方实验中涌现——芬兰的UBI试点、新加坡的智能公共服务、中国的社区建设者计划、巴西的行动网络——每一个都是一块拼图，最终拼出一幅谁也没有预见到的完整图景。

**教训三：新契约的建立需要一代人的时间。** 按1930年代的类比推算，从危机爆发到成熟的新制度框架，需要20到30年。这意味着如果以2030年为起点，一个相对成熟的新社会契约框架要到2050年代才能成型。按这一类比，我们正处在中间地带：旧契约已经明显不够用了，新契约的轮廓可以辨认了，但它远未定型。

---

## 八、确定性乐观的理由

对未来的态度可以分为四种：确定性乐观（知道未来会好，并为此努力）、不确定性乐观（相信未来会好，但不知道怎么好）、确定性悲观（知道未来会坏，但无法改变）、不确定性悲观（不知道未来怎样，但倾向于会坏）。

未来的公共情绪可能在后三种之间摇摆。但历史给了确定性乐观一个谨慎但坚实的理由：人类社会在面对前所未有的分配危机时，最终总是能够——在付出巨大代价之后——拼凑出一套新契约。从封建义务制到劳动–工资制再到福利国家，每一次都在巨大代价之后拼凑出了新契约。

没有理由认为这一次会不同。也没有理由认为阵痛会比上一次小。**制度从来不是被设计出来的，而是被逼出来的。**

但有一件事值得注意，它是真正的0→1式创新的种子：AI时代的分配问题第一次有可能不是零和博弈。在所有以前的时代，分配多就是分配少。在AI时代，如果基础设施治理得当，“给你不意味着少给我”第一次有了技术基础。这不会自动实现——它需要政治意志、制度设计和持续的博弈。

一个从未有过的可能性，加上一个反复验证的历史模式——人类在危机后总是能重建契约——这两者加在一起，构成了确定性乐观的最小必要条件。

下一个问题是：在新契约的框架内，权力和阶层将如何重新排列？分配规则确定之后，谁在上面，谁在下面？

这就是第十七章要回答的问题。

# 第十七章 新阶层：谁在上升，谁在下沉

> 每一次技术革命都承诺平等，每一次都先扩大不平等。逆转从来不靠技术本身，而靠政治行动。

---

*2035年，夏天。*

*大学同学群里有人发了聚会的照片。她点开看了看，没去——路费加份子钱，要花掉小半个月的保障金。*

*照片里她认出了林悦。大学时林悦成绩中等，毕业后去了一家不大的电商公司。两年前同学群里有人转了一篇报道，林悦现在的头衔是"AI业务编排总监"，负责管理一个由三十多个AI Agent组成的运营团队。报道里提了一句年薪，她看完关掉了手机。*

*不是嫉妒。或者不完全是。让她不舒服的是那种感觉：2026年毕业的时候，她的起点比林悦高——更好的绩点，更好的第一份工作。九年过去了，差距不是在缩小，是在以一种她完全没有预料到的方式加速拉开。不是因为谁更努力，而是因为她们碰巧站在了不同的位置上，然后潮水来了。*

*她关上同学群，打开花园的排班表，确认了明天的时间。排班表上她和一个叫老周的人撞在同一个上午——负责工具棚的那个沉默的中年人。她没多想，但也没觉得不好。*

---

上一章讨论的是契约——社会层面的规则重写。这一章把镜头拉近：在新规则落地之前，谁在上升，谁在下沉？

## 一、同一栋楼里的两个世界

在本书的推演中，2034年深圳南山区，一栋普通写字楼的早晨。

三十二楼，一位AI业务编排师走进她的工作室。她在一家消费品公司工作，每天的工作是把市场部模糊的需求翻译成AI工作流——竞品分析、用户画像生成、营销方案比选——再把AI输出翻译成决策层能看懂的结论。三年前她还是一个普通的市场经理，年薪三十万。现在她管理着十二个AI Agent组成的"虚拟团队"，年薪一百六十万。她的LinkedIn标签是"Human-AI Interface Designer"。

同一栋楼的一楼大堂，一位53岁的前质检员穿着保安制服站在闸机旁。他原来在龙华一家电子厂做质检组长，月薪八千。如果产线在2030年前后全面AI化，他将面临下岗，领完失业金后可能转做外卖骑手。而到2033年前后配送机器人覆盖他跑的片区，他可能再次失业。现在这份保安工作月薪三千

五，还是他姐夫托关系找到的——物业公司更愿意用安保机器人，但写字楼业委会坚持要”至少有个活人在大堂”。

这两个人住在同一个城市，路过同一家便利店，用同一款手机。但他们的生活轨迹和未来可能性，差距之大可能超过封建时代的地主和佃农。

这不是两个人的故事。这是一个时代的分层逻辑。

---

## 二、从四种资本到五层阶梯

人类社会为什么需要分层？比”人性贪婪”的循环论证更深的回答是：分层是复杂社会用来解决”谁做什么”和”谁得什么”的信息简化机制。在百万人的社会中，不可能为每个人量身定制角色和报酬，分层提供了一套粗糙但可用的分类系统。

关键问题从来只有一个：分层依据什么？

农业社会按土地分层。工业社会按”劳动能力×资本拥有量”分层——韦伯的三维框架（经济–地位–权力）精确描述了这套逻辑。AI社会呢？

第五章分析了布迪厄的四种资本在AI冲击下的权重迁移：文化资本（知识、学历）在贬值，社会资本（关系、信任）在升值，符号资本（注意力、品牌）在爆发，经济资本在分化。那一章回答的是”旧中产为什么塌陷”。本章要追问的是下一个问题：**旧的塌了，新的长成什么样？**

如果把四种资本的新权重映射到劳动力市场的趋势上——按2028年到2033年间可能出现的全球劳动力变化来推演——人们可能自然地聚集在五个位置上：

- 控制AI基础设施的人（经济资本的极端集中）→ **算力阶层**
- 编排AI创造价值的人（文化资本向新形态转化）→ **编排阶层**
- 依靠人际关系生存的人（社会资本为核心）→ **关系阶层**
- 提供”人类亲手做”体验的人（符号资本的新来源）→ **体验阶层**
- 四种资本都不足的人（依赖转移支付）→ **依赖阶层**

四种资本的权重迁移是机制，五层阶梯是结果。下面逐层展开。

---

## 三、五层新阶梯：每一层的面孔

### 第一层：算力阶层——站在管道上游的人

拥有或控制AI运行所依赖的基础设施——芯片、数据中心、大模型、云服务。按本书的推演，到2030年代中期，全球前十大AI基础设施公司的合计市值可能占到全球股市总市值的近三分之一。

**一个面孔。** 一位AI对齐团队负责人，38岁，在某头部AI公司工作。全球能做他这份工作的人不超过五百个。他每天的决策——训练数据怎么筛选、对齐标准往哪个方向调、开放什么能力给开发者——间接影响着十几亿AI用户看到的世界。他不觉得自己在”行使权力”，他觉得自己在”做技术决策”。但当一个技术决策可以改变十几亿人的信息环境，“技术”和”权力”之间的边界已经模糊了。

算力阶层的权力不来自使用AI，而来自**定义AI的边界**。这是一种”不对称的skin in the game”：做决策的人承担的风险远小于受决策影响的人。五百名研究者在对齐标准上的一个偏差，可能扭曲数十亿人的信息环境——但研究者本人几乎不为此承担后果。人类历史上只有中世纪教廷对教义的垄断权可以比拟。

### 第二层：编排阶层——人与AI之间的翻译官

不拥有AI基础设施，但知道怎么用它创造商业价值。核心能力是”翻译”——把人类模糊的业务需求翻译成AI可执行的工作流，把AI输出翻译成人类可理解的决策依据。

**回到那位编排师。** 她没有计算机学位，本科学的是广告学。但她有一种直觉——能同时理解”老板到底想要什么”和”AI到底能做什么”，并在两者之间搭建桥梁。她的团队里有人编程更好，有人行业知识更深，但没人比她更擅长这种”翻译”。这种能力目前无法通过标准化教育批量培养。

按本书的推演模型估计，编排阶层可能占全球劳动力的3%到5%，贡献约五分之一的GDP，平均收入是同龄人的四到六倍。

但一个逆向观察必须提出：编排阶层的溢价建立在AI”暂时不擅长理解人类模糊意图”这个缺口上。如果2035年之后AI学会了直接理解模糊意图（苗头已经出现），编排阶层的溢价就会迅速消退。**建立在暂时性技术缺口上的阶层地位，是脆弱的。**

### 第三层：关系阶层——信任无法API化

既不控制AI，也不编排AI，但拥有AI无法替代的东西——人际关系网络和信任资本。

**一个面孔。** 一位社区辅导员，41岁，在某三线城市做心理咨询。AI可以生成完美的心理咨询话术——她自己也用AI辅助写咨询记录。但她的来访者选择她而不是AI咨询，理由简单到无法反驳：“我需要一个活人听我说话。”她的月收入不高，八千左右，和隔壁超市收银员差不多。但顶层的关系从业者——顶级投行关系经理、大型企业政府关系总监——收入堪比编排阶层。

把这些人归为同一层的理由不是收入，而是收入来源的逻辑：**价值来自人类关系。** AI可以比任何人更高效地分析客户数据，但它不能在饭桌上建立信任。

## 第四层：体验阶层——“人类亲手做”的溢价

手工艺人、运动员、厨师、表演艺术家、瑜伽教练——所有提供”由人类亲手完成”这一体验的人。

按趋势推演，可能出现一个V型曲线：2026到2030年体验类从业者收入急剧下降（AI替代品足够便宜），此后”人工制作”标签开始获得溢价，到2030年代中期溢价率可能达到35%到80%。但溢价只属于头部——消费者愿意为”最好的人类手工”付高价，不愿意为”普通的人类手工”付溢价。体验阶层内部的头尾分化，比任何传统行业都极端。

## 第五层：依赖阶层——被系统”合理地”放弃

赫拉利2016年抛出”无用阶级”（useless class）这个概念时引发了巨大争议。方向正确但过于粗糙——“无用”把一个分层的、有阶段的、因文化而异的复杂过程压缩成了一个标签。本章试图展开这个标签：不是一个”无用阶级”，而是五层新阶梯，每一层与AI认知基础设施的关系不同，每一层的脆弱性形态也不同。赫拉利看到了终局的轮廓，但略过了中间地带——而中间地带恰恰是政策可以介入、个体可以选择的空间。

劳动能力在AI时代没有市场价格。生存依赖转移支付：政府保障金、企业社会责任基金、家庭供养。

**回到那位前质检员。** 在本书的推演场景中，他所在的区可能启用”AI转型保障金”智能升级系统：AI分析保障金领取者的消费行为和社交数据，评估”再就业潜力”。高潜力者推荐AI编排入门课，中潜力者推荐社区服务岗位培训，低潜力者推荐”积极生活”计划——本质上是放弃再就业的柔性表达。

他55岁，初中学历，没有数字技能基础，大概率被评为”低潜力”。不是因为他不努力或不聪明——是因为他的年龄、学历和技能基础在统计模型中指向”投入再培训资源的预期回报率极低”。

算法没有恶意。工程师没有恶意。政府把有限资源分配给最可能受益的人也是”合理的”。但所有这些”合理”加在一起，产生了一个结果：**他被系统性地排除在阶层流动的通道之外。** 他不是被某个人否定了——他是被一整套逻辑否定了。

按本书的推演模型估计，依赖阶层在中国可能占城镇劳动年龄人口的18%到22%，全球发达经济体约15%到25%。按推演估计，2029年前后进入依赖阶层的人群中，五年后仍在依赖阶层的比例可能超过七成，向上流动进入编排阶层的比例可能不到百分之二。

---

## 四、为什么新阶层比旧阶层更难翻越

第三章分析了教育作为分拣机器的功能。在新阶层结构中，分拣的标准从”你学了什么”变成了”你能调度什么”——但分拣本身并未消失，只是换了一套更隐蔽的筛子。

直觉告诉我们：AI降低了知识获取门槛，应该让阶层更流动。一个贵州山村的孩子和北京中关村的孩子第一次可以获得同等质量的教育辅导。一个人加一组Agent就能做过去十人团队的事。这些都是真实的。

但有一个隐藏的结构性机制在起反作用：**认知基础设施的规模效应天然有利于在位者。**

工业时代，资本的规模效应有上限——工厂再大，边际效率最终递减（管理复杂度增加、官僚化、创新僵化）。这给了后来者机会：更灵活的小工厂可以打败臃肿的大工厂。所以工业时代始终有阶层流动。

AI时代，规模效应几乎没有上限。模型越大越强，数据越多越精准，算力越集中效率越高。而且这些优势自我强化：更好的模型→更多用户→更多数据→更好的模型。终态是极端集中，而非分散竞争。

换句话说：工业时代，权力有一种自然的衰减机制（大企业终将僵化）。AI时代，这种衰减机制被削弱了。反脆弱性——从冲击中获益的能力——只属于控制基础设施的那一层。其余各层都是脆弱的，它们依赖上层的善意或监管约束来维持自己的位置。

**代际裂痕加剧了固化。** 间奏二里住在同一栋楼的两个人是这条裂痕的缩影：二楼的年轻人拒了大厂offer，自称”AI水管工”，每天花三小时跟进最新工具；四楼退休的银行副行长连”Agent经济”这个词都看不懂。他们在楼道里碰见，点头，说”吃了吗”——然后各走各的，活在不同的时区。“数字原住民”和”数字移民”的起跑线截然不同。前者编排AI几乎是本能，后者需要在中年重新学习一整套协作思维，而神经可塑性和学习时间都不站在他们这边。结果：四十岁以上从业者大批滑入依赖阶层，向上流动的通道几乎被年轻一代垄断。这不是能力问题，而是认知基础设施的”母语效应”——你在什么环境中长大，决定了你能多自然地使用这个时代的核心工具。

还有一个容易被忽视的维度：**阶层下滑带有性别印记。** AI最先大规模替代的白领岗位——行政助理、翻译、初级教育、客服——女性从业者比例显著偏高。制造业自动化冲击的则主要是男性蓝领。女性更多从关系阶层滑向依赖阶层（关系资本被算法替代），男性更多从体验阶层被机器人挤出。新阶层结构表面上只按”与AI的技术关系”排序，实际上旧的性别分工通过行业结构悄悄嵌入了新的分层逻辑。

**还有一个全新的特权维度：与AI前沿的认知距离。** 除了技能和资本，AI时代的阶层分化多了一条隐性的分界线——你离AI的技术前沿有多远。Dario Amodei的判断令人警醒：“硅谷及与硅谷有紧密社会联系的地区，经济增长率可能达到50%，而世界其他地区与当前相差无几。”这种距离既是地理的——深圳比内陆城市更近，旧金山比中西部小镇更近；也是社会网络的——一个在AI公司工作的朋友比一个MBA学位更有价值，因为前者带来的是对技术可能性的实时感知；更是文化的——你的日常对话中是否自然地包含AI思维，你是否习惯用”这个任务可以让AI做”来重新审视工作流程。教育年限曾经是阶层的分界线。在AI时代，认知距离正在取代它。一个没上过大学但在深圳科技公司做运营的年轻人，

可能比一个名校博士但身处传统行业的中年人更”接近前沿”。分界线不再印在学历证书上，而是嵌入你的信息环境、社交圈层和日常习惯之中——比学历更隐蔽，也更难跨越。

---

## 五、声望的重新排序

阶层不只是钱的问题。韦伯三维框架的第二维是声望。

工业社会的声望排序靠”专业知识的稀缺性”——知识越稀缺、获取门槛越高，声望越高。医生、律师、教授排前面，蓝领工人排后面。

AI用五到十年把知识的稀缺性溢价剥掉了。AI能通过执业医师考试、律师资格考试、写出被学术期刊接收的论文——“专业知识”不再是声望的可靠锚点，判断力成了新的区分标准。

新的声望逻辑沿两条线展开：

**“你能让AI做什么”**——编排阶层获得了类似工业时代早期工程师的声望：掌握大多数人不理解的新力量，并能用它创造可见成果。

**“你能提供什么AI不能提供的”**——心理咨询师、社区领袖、优秀的小学教师——这些工业时代的”中等声望”职业，声望显著上升。收入大部分没提高，但人们越来越意识到它们提供的东西是真正稀缺的。

布迪厄要是活到这个时代，一定会露出一个”我早说过”的微笑。他毕生研究的那种区隔机制又出现了：当所有人都能用AI写文案、做设计、生成视频，“用AI做出什么”不再是区分标准——**“做出的东西被谁认可”才是。** 编排阶层内部迅速分化出隐形边界：一边是为精英圈层生产”有品味”内容的人（策展、品牌叙事、高端咨询），另一边是用同样工具做标准化代工的人（批量电商文案、模板化报告）。技术能力相近，收入和声望天差地别。工具民主化了，品味的门槛反而更高了。

依赖阶层的声望则经历了比传统失业更深的下降。传统失业带有”坏运气”色彩——你能工作，只是暂时没找到。依赖阶层的处境暗示的是：**你的能力在这个时代没有市场价值。** 这种暗示带来的耻辱感更深，也更难用政策消除。

---

## 六、隐形的权力：谁定义了”合理”

阶层结构需要权力来维持。工业时代的权力形式是可见的：国家征税，你知道钱被拿走了；资本家压低工资，你知道劳动被剥削了。

AI时代叠加了一种全新的权力形式：**对认知基础设施的控制权。** 它的独特之处在于隐蔽——一个AI公司在训练数据中偏向某种价值观、在对齐过程中排除某些观点，你不知道。你只觉得AI给的答案”

很合理”，而不会去想这个”合理”是被谁定义的。

可以称之为”隐形的立法权”：AI公司通过技术决策实质上在规定数十亿人的信息环境、决策框架和价值判断。如果这种趋势延续，这种立法权到2030年代初可能仍然几乎没有民主约束——没有选举、没有听证、没有司法审查。传统民主制度在暴力权和经济权维度上积累了几百年的约束工具（宪法、反垄断法、劳动法），但面对认知权这个新维度，既有的民主工具箱里几乎找不到对应的器具。

更深层的问题是：现代化的核心要素之一是”自由思想的市场”。但如果AI系统的对齐标准由少数机构单方面设定，这就不再是自由市场，而是一种新型的”思想计划经济”。训练数据的选择是对”什么算知识”的定义权，对齐标准的设定是对”什么算合理”的裁决权。这些权力在人类历史上由宗教机构、大学和法院分别行使，现在被集中到了科技公司的研究团队中。

这种权力集中直接强化了阶层固化。当算力阶层不仅控制着经济资源，还控制着”合理性”的定义——什么算有价值的工作、什么算合理的分配、什么算公正的秩序——其他阶层的人甚至缺乏质疑现有秩序的概念工具。

---

## 七、国家与平台的边界模糊

权力的隐形化还有另一个维度：国家与商业平台的融合。

当移动支付平台开始承担保障金代发职能——基础保障金不再通过银行卡而是直接进入支付账户，配合”保障金专属折扣”。技术上无可挑剔。但当一个商业公司成为国家福利的发放渠道，并通过折扣引导消费行为时，“国家”和”公司”之间的边界在哪里?

按当前趋势推演，到2035年前后，中国城镇居民日常生活中以下公共职能可能事实上由商业平台承担：身份验证、交通出行、医疗初筛、基础教育辅助、就业匹配、社会信用评估、部分地区的社保发放。

研究者称之为”双向渗透”：**平台国家化**（公司承担公共职能，获得准公共权力）和**国家平台化**（政府依赖商业平台运行社会，让渡部分主权功能）。超级支付平台决定哪些商户可以加入”保障金折扣”体系时，它在行使资源分配权。AI招聘平台的算法决定什么简历排前面时，它在行使机会分配权。

这对阶层意味着什么？意味着维持阶层秩序的力量不再只是国家权力或市场力量，而是一种算法–平台–政府的混合体——它比传统治理更高效，也更不透明。传统治理至少有一部宪法来约束它。这种混合治理的权力边界藏在无数合同条款、技术协议和算法参数之中，没有人看到过它的全貌。

---

## 八、隐形阶层——历史上最难反抗的不平等

从宏观结构回到微观体验。新阶层秩序最显著的特征不是残酷，而是**隐形**。

封建社会的阶层是可见的——贵族穿不同衣服，住不同城堡。工业社会的阶层也是可见的——工厂主和工人的生活差距一目了然。AI社会的阶层隐藏在同样的手机、同样的街道、同样的快时尚服装之下。

编排师和前质检员用同一款手机，刷同一个短视频平台——但手机里的算法推荐完全不同的内容，引导完全不同的消费，连接完全不同的社交圈层。**表面上的平等掩盖了实质上的隔离。**

隐形使得反抗更加困难。封建时代的农民知道敌人是城堡里的领主。工业时代的工人知道斗争对象是工厂主。AI时代的依赖阶层不知道该向谁抗议。向AI？它只是工具。向科技公司？它们只是在做生意。向政府？政府已经在发保障金了。向算法？算法没有地址。

权力变得分散、抽象、系统化，以至于没有一个可以指认的压迫者。结构性不平等没有消失，但改变它需要无数个微小的、持续的、跨领域的制度调整。**最看不见的不平等往往最难反抗。** 没有靶心，就没有运动。

---

## 九、你以为AI让社会更平等

这是本章的认知翻转。

直觉说：AI降低了知识门槛、创业门槛、专业服务价格——应该更平等。

历史说：**每一次重大技术革命的初始效应都是扩大不平等。** 不是观点，是反复验证的事实。

印刷术的第一批受益者是教会和国王，普通人等了近一百年才从中获益；蒸汽机的第一批受益者是工厂主，工人又花了五六十年的工会运动才分到红利；互联网的第一批受益者是硅谷创业者和风投，到2024年全球前五大科技公司的市值之和超过了GDP排名后一百个国家的GDP之和。

AI正在重复同一个模式。降低了知识获取门槛，但提高了知识应用门槛。让工具民主化了，但让工具编排的能力精英化了。

这个模式不是必然的。每次技术革命最终都经历了从”集中”到”分散”的转向——但转向从来不是技术自动带来的。印刷术分散权力靠的是宗教改革。蒸汽机分散权力靠的是工会运动。互联网分散权力（如果它分散了的话）靠的是反垄断法。

AI会分散权力吗？也许。但”也许”取决于政治行动。等待技术自动带来平等，是一种危险的幻想。

---

## 十、国家间的AI鸿沟

前面几节讨论的是一个国家内部的阶层分化。但AI时代最严峻的不平等可能发生在国家之间。

过去半个世纪，发展中国家的追赶式增长有一个核心机制：它们拥有大量未被充分利用的廉价劳动力，发达国家将资本和技术带入，激活这些劳动力，推动经济起飞。从亚洲四小龙到中国制造业奇迹，底层逻辑都是如此——用人口红利换取工业化的入场券。

AI正在使这个机制失效。当认知劳动的边际成本趋近于零，“廉价劳动力”不再是发展中国家的比较优势——因为AI比任何劳动力都更便宜、更快、更不需要基础设施。追赶的阶梯被抽走了。一个越南工厂工人的成本优势，在AI驱动的自动化面前正在快速蒸发；一个印度外包程序员的价格优势，在AI编程助手面前同样岌岌可危。

Dario Amodei坦言：AI推动的10%到20%年经济增长率，可能呈现极端的地域差异。技术前沿地区飞速增长，而远离前沿的地区原地踏步。“这会是一个极其糟糕的结果。”这不是假设——它正在发生。按这一趋势推演，到2030年代初，硅谷和深圳的人均GDP增速可能达到撒哈拉以南非洲的十倍以上，而且差距可能在加速扩大而非收敛。

更深的问题是：AI时代的国际援助模式也需要重新设计。传统援助的逻辑是”教你捕鱼”——培训劳动技能、建设基础设施、发展出口导向型制造业。但当”鱼”可以被AI无限生成时，“捕鱼技能”本身贬值了。一个非洲国家花二十年建立的呼叫中心产业，可能在五年内被AI客服替代。新的援助逻辑需要从根本上重构：不是教技能，而是建设AI基础设施——数据中心、算力接入、高速网络；不是培训流水线工人，而是培养能与AI协作的本地人才；不是复制旧的工业化路径，而是探索AI驱动的本地化产业——比如AI农业，它可以在没有大规模机械化的条件下，通过精准灌溉、病虫害预测和土壤分析显著提升产量，而这些恰恰是非洲和东南亚小农经济最需要的。

如果这个问题得不到解决，AI时代的全球格局将不是”所有船一起涨”，而是”一些船起飞，一些船搁浅”。工业革命创造的南北差距花了两百年都没有弥合。AI革命创造的新鸿沟，如果在早期不被正视，可能比工业时代的更深、更持久——因为这一次，追赶的传统路径本身被技术封死了。

---

## 十一、认知翻转与未解之问

回到第一性原理。阶层存在是因为复杂社会需要信息简化机制。AI时代换了一套机制——从”你能做什么工作”到”你和认知基础设施的关系是什么”。新机制不比旧机制更公正，只是更高效。

但”更高效”包含一个危险。如果新阶层结构的固化速度超过了制度调整的速度，系统将到达不可逆转的临界点——之后即使有完美的政策方案，也无法在不引发系统性动荡的前提下完成再平衡。法国大革命前的财富集中、1929年前的投机狂潮、2008年前的信贷泡沫——每一次，都有人事前看到了趋势，但制度惯性使得调整总是”太少、太晚”。

两个未解之问留给读者。

**第一：阶层流动的新通道在哪里？** 封建社会靠科举，工业社会靠教育。AI社会靠什么？社群建设？IP经营？与AI的协作能力？还是——已经没有了？

**第二：隐形的阶层是否比可见的阶层更危险？** 可见的不平等至少能激发社会运动。隐形的不平等可能被永远忽视——人们不是不在乎，而是根本看不见。

新阶层已经成形，分配规则正在固化。但一个更根本的问题还没有回答：在这个新秩序中，人做什么？当劳动和大部分工作都被接管之后，人类释放出来的时间和能量将流向何处？

这就是第十八章的问题。

# 第十八章 超级公司与新利维坦

> 霍布斯说，人们让渡暴力给国家，换取安全。他没有预见到的是：三百年后，人们还会让渡认知给一家公司——而且大多数人根本不知道自己让渡了什么。

---

*2035年，秋天。*

*那天她是被手机闹钟叫醒的——不是她设的，是AI助手根据她的睡眠数据自动调的。起床后她照例问助手今天的天气和日程。刷牙时镜子上弹出健康提醒：昨晚深睡不够，建议少喝咖啡。*

*出门前她点了杯豆浆外卖，下单的app、配送的平台、付款的钱包，都是同一家公司的。坐公交时她戴着耳机听新闻播客，推荐算法选的。手机弹了一条通知：转型保障金已到账。她随手划掉——领了大半年，已经不会再盯着余额看了。到了花园，她用那家公司的协作工具给志愿者们发了排班通知。*

*晚上洗完澡，她突然想数一数：从早到晚，她用过多少不同公司的服务？数了一圈发现，除了水电煤和公交，剩下的——支付、通讯、外卖、导航、健康管理、新闻、购物、社区协作——全是同一个生态系统。*

*她试着想象：如果这家公司明天消失了，她的一天会怎样？想了一会儿，发现不是”不方便”的问题，是”几乎没法正常运转”的问题。这个念头让她心里发紧，但她不知道能做什么。她把手机放到床头柜上——手机壳背面印着那家公司的logo。*

---

前面几章讨论了分配和阶层——但这些讨论有一个隐含假设：规则由国家制定。这一章质疑这个假设：在AI时代，权力的实际重心可能早就不在国家手里了。

## 一、一天的生活

在本书的推演中，2039年末，欧盟委员会内部可能流传着这样一份备忘录。备忘录没有标题，只有一行数据：一家美国科技公司的AI生态系统，在过去十二个月里处理的欧盟公民医疗咨询、教育辅导和法律建议的总量，超过了欧盟全部公共服务机构的同类总量。备忘录附了一句手写批注：“我们还是主权国家吗？”

这个问题没有进入任何公开议程。但它是理解本章的入口：当一家公司对公民日常决策的影响力超过政府时，“公司”这个词还够用吗？

从一个普通人的一天说起。

在本书的推演中，2040年，一个中国二线城市的普通居民醒来。

叫醒他的不是闹钟，是一个AI助手——根据他的睡眠周期、当天日程和实时天气，在最佳时刻用渐亮的屏幕光唤醒他。这个助手运行在同一家公司的云端，和他的智能手表共享睡眠数据。他刷牙时，浴室镜面显示今日健康摘要：昨晚深睡不足，建议减少咖啡摄入。这条建议来自同一家公司的健康管理模块，关联着他的体检报告和三个月的饮食记录。

早餐时，AI推荐了三个方案：燕麦粥（基于他的血糖趋势）、三明治（基于冰箱存货）、外卖豆浆（基于附近店铺的实时供应）。他选了外卖，下单走的是同一家公司的配送平台。通勤路上，他戴上耳机听新闻，内容由同一家公司的信息流算法推送。到达公司后，他打开工作台——同一家公司的协作套件——AI已经为他生成了今天的任务优先级排序、起草了两封客户邮件、标注了三份需要他审核的合同摘要。

午餐时他和朋友聊起周末去哪里玩，AI助手介入推荐了一条旅游路线，综合了他的预算偏好、过往出行记录和朋友圈里最近的热门目的地。下午他感到一阵头疼，AI健康模块建议他做一次远程问诊——同一家公司的医疗服务接口，三分钟后一个AI医生完成了初步诊断，建议他去药店买一盒布洛芬，药店也是同一个生态系统的合作网点。

晚上，他陪孩子做作业。孩子用的是同一家公司的教育AI——它比大多数人类教师更了解这个孩子的知识盲区、学习风格和注意力曲线。睡前，他看了四十分钟短视频，算法精准地在他困意来袭前推送了一条助眠音频。

一天结束。他没有感到任何不适。一切顺畅、高效、贴心。

但退后一步审视：从睡眠到饮食、从工作到医疗、从教育到娱乐，同一家公司的系统触及了他生活的每一个决策节点。它掌握的数据——健康、财务、社交、消费、工作效率、认知偏好——比任何政府机构都全面。它对他日常决策的影响——吃什么、看什么、怎么工作、如何教育孩子——比任何一部法律都具体。

这家公司不是他的雇主，不是他的政府，也不是他的亲人。用2040年代可能出现的语言来说，它是一个”生活基础设施提供商”。用更古老的政治哲学语言来说：它正在成为一种新形态的利维坦。

---

## 二、治理权力的三个维度

要理解为什么一家公司能悄然获得利维坦的影子，需要先更新我们对”治理权力”的理解框架。

1651年，霍布斯在英国内战的废墟上写出了《利维坦》。他的核心方案简明到可以写在餐巾纸上：自然状态下人人自危，人们将暴力权让渡给主权者，换取安全与秩序。这个交易模型简洁有力，影响了此后四百年的政治理论。

但它是一维的。霍布斯被内战吓坏了，他最关心的是谁有权合法地使用暴力。在他的时代，暴力确实是权力最直接的形式。但治理权力从来不只是暴力。如果站在2046年回望，治理权力至少可能有三个维度。

**第一维度：暴力权。** 韦伯的经典定义：国家是”成功垄断了合法暴力使用权”的实体。到2046年，暴力权大概率仍由国家牢牢垄断。没有任何公司拥有军队，没有任何平台可以合法地将你投入监狱。

**第二维度：经济权。** 谁有权配置资源，谁就拥有经济权力。税收、货币发行、市场监管——这些是国家经济权力的体现。但经济权从来不是国家独占的。标准石油在十九世纪末控制了美国九成以上的炼油产能，AT&T在拆分前垄断了美国电信网络。这些公司的经济决策对普通人生活的影响不亚于政府政策。二十世纪的反垄断法本质上就是国家在经济权维度上与巨型公司的地盘争夺。

**第三维度：认知权。** 这是AI时代浮出水面的新维度。认知权指的是：谁的系统处理你的信息，谁的算法塑造你的决策，谁的标准定义”什么是可见的、可说的、可信的”。

在印刷时代，认知权高度分散。在广播电视时代，认知权开始集中但仍受政府牌照监管。在互联网早期，认知权一度去中心化。但从2010年代社交媒体时代开始，认知权经历了隐秘而剧烈的重新集中——平台算法拥有了一种此前不存在的权力：不是审查你说什么，而是决定你看到什么。

AI时代将认知权的集中推到了新高度。当AI系统不只是推送信息，而是直接参与你的决策——帮你分析合同、诊断疾病、规划财务、辅导孩子——它就不再只是信息的”分发者”，而是认知的”共同生产者”。AI系统的设计者——通常是一家或几家公司——实际上参与了你最重要的人生决策。

认知权的实质：它不是强迫你做什么（那是暴力权），不是决定你能买什么（那是经济权），而是影响你怎么想、怎么判断、怎么选择。

由此可以更新霍布斯的利维坦方程。传统版本：让渡暴力权给国家，换取安全。AI时代版本：同时让渡暴力权给国家，经济权给市场体系，认知权给AI平台——换取安全、效率和便利。

但这个方程的更新远不只是”多了一项”。新旧利维坦的权力机制存在本质区别，理解这种区别才能看清我们正在面对什么。霍布斯的利维坦通过垄断合法暴力来维持秩序——你不服从，会被惩罚。这种权力是显性的、可识别的、有边界的。你知道法律禁止什么，你知道违反的后果，你甚至可以选择承受惩罚来表达反抗。暴力权的运作前提是：被治理者知道自己正在被治理。

数字利维坦的权力机制完全不同。它不惩罚你，它塑造你的选择空间。你打开手机，看到的信息流是它排序的；你搜索一个问题，呈现的答案是它筛选的；你”自由选择”的商品、观点、生活方式，选项本身是它安排的，排列顺序是它决定的，而你对这些选项的偏好——很可能是它通过长期的数据反馈循环逐渐培养的。传统国家权力是”禁止性的”：你不能做X，否则受罚。数字利维坦的权力是”生产性的”：它不禁止任何东西，它生产你的欲望、引导你的注意力方向、塑造你的认知框架。福柯在二十世纪就区分了这两种权力形态，但他分析的对象是医院、监狱和学校——那些机构通过物理空间和制度程序来规训个体，效率有限，覆盖有边界。AI时代，生产性权力的运作效率提升了几个数量级：它不需要物理空间，不需要制度程序，只需要一块屏幕和一个算法，就能以每秒千万次的频率对数十亿人同时施加影响——而且是个性化的、精准的、让你感到舒适的影响。

这引出了最深处的问题：面对传统利维坦，你至少知道自己在服从。一个人走进法庭，他清楚地知道国家权力正在作用于他。面对数字利维坦，你甚至不知道自己在被治理——因为它的治理方式恰恰是让你感到自由和舒适。当算法推送的内容完美匹配你的兴趣时，你体验到的是”被理解”，不是”被控制”。当AI助手为你筛选掉九成信息只留下”最相关”的一成时，你体验到的是”高效”，不是”信息被裁剪”。旧利维坦的运作前提是被治理者知道自己正在被治理。新利维坦的运作前提恰恰相反。

关键的不对称：前两项让渡是有自觉的——你知道你在交税，你知道法律约束着你，这些让渡有明确的法律框架和民主程序的合法性背书。但认知权的让渡是不自觉的。没有人在下载AI助手时签署了一份”我同意让这家公司参与我的认知过程”的社会契约。它是在”免费使用”“提升效率”的包装下，以每天一小步的方式完成的。

治理权力三维模型揭示的核心问题：AI时代的权力配置正在三个维度上同时展开，但只有暴力权和经济权有成熟的治理框架。认知权的治理，几乎是一片空白。

---

## 三、四次关系重构——公司与国家的历史

治理权力的三维模型是公司与国家四百年博弈的产物。每一次重构都改变了”公司到底是什么”的答案。

**第一次重构：殖民时代（1600–1800）——公司=国家代理人。**

1600年英国东印度公司成立，1602年荷兰东印度公司紧随其后。这些实体在法律上是”公司”，但它们拥有的权力远超今天任何人对”公司”的想象。荷兰东印度公司可以缔结条约、发动战争、建造堡垒、铸造货币、对殖民地居民执行司法审判。巅峰时期拥有约一万名士兵和一百五十余艘武装商船——比当时欧洲大多数国家的正规军都庞大。

这是三维权力合一的极端案例：暴力权（军队和司法权）、经济权（贸易垄断和货币铸造）、认知权（对殖民地的信息控制和文化输出）集于一身。它之所以能拥有这些权力，是因为主权国家主动授予——国王和议会将海外扩张的”脏活”外包给公司，换取税收和地缘利益。

转折是必然的：当公司权力大到反噬宿主，公司利益与国家利益冲突加剧，腐败开始侵蚀国内政治——国家收回了这些权力。英国东印度公司1858年被剥夺治理权力，1874年正式解散，治理权交还给英国政府。

**第二次重构：工业时代（1800–1945）——公司⊂国家。**

工业革命创造了经济权力史无前例的大企业——钢铁厂、铁路公司、石油巨头——但暴力权被彻底剥离。没有哪家工厂主可以合法地拥有军队。这个时期的核心创新是现代反垄断法和劳动法的建立。1890年美国《谢尔曼法案》、1914年《克莱顿法案》，本质上是国家在经济权维度上为公司划定边界：

你可以追求利润，但不能垄断市场；你可以雇佣工人，但不能无限制地剥削他们。公司强大但从属于国家，是笼子里的巨兽。

转折点：两次世界大战和大萧条让国家权力空前扩张。战后的福利国家体系——社会保障、最低工资、累进税制——是国家对公司经济权力的进一步约束。

**第三次重构：全球化时代（1945–2020）——公司≈国家。**

跨国公司的崛起改变了游戏规则。当一家公司在一百七十个国家运营时，它可以在不同法律体系间"套利"——利润记在低税率国家，生产放在低成本国家，总部注册在监管最宽松的地方。效果是国家对公司经济行为的约束力被系统性削弱了。更深层的变化在于：互联网平台开始获得认知权的雏形——搜索引擎决定你能找到什么信息，社交媒体决定你看到什么观点，电商平台决定你买到什么商品。国家的法律管辖权以领土为边界，而数据和算法没有边界。

**第四次重构：AI时代（2025– ）——治理权力在国家与科技平台之间重新分配。**

公司没有取代国家，暴力权仍牢牢在国家手中。但在认知权维度上，AI公司的功能开始越过传统商业边界，进入此前属于公共治理的领域。这种越界的机制不是阴谋，而是技术逻辑的自然展开：AI系统的效用随数据量和应用场景的扩展而提升，一个AI助手同时管理健康、工作、教育和娱乐，比四个独立AI分别管理要高效得多。这种"全场景整合"的技术优势，驱动AI公司不断扩展生态系统的边界。

结果是，在某些功能维度上，科技平台开始承担准公共服务职能——国家通过法律影响公民行为，AI公司通过算法影响用户决策；国家通过教育体系塑造认知框架，AI公司通过个性化学习系统重新定义"学什么"和"怎么学"。这是功能性的重叠，而非权力的替代——国家并未因此丧失治理能力，但治理的版图确实在重新划分。

第四章问"公司为什么存在"——科斯的答案是交易成本。那么当交易成本趋近于零时，超级公司为什么反而更大了？因为答案变了：不是交易成本，而是数据垄断。AI将科斯的交易成本框架从经济领域延伸到认知、健康、教育领域，公司的功能边界溢出了"公司"的范畴。AI超级公司已经不是科斯定义的"公司"了，它是一种新事物，介于公司和国家之间，它的功能边界溢出了商业范畴，进入了治理领域。

---

## 四、AI公司的质变——从商品到基础设施利维坦

要精确理解AI公司权力的性质，需要引入第二个分析框架：**基础设施依赖度光谱**。

核心思想：一个实体对你的控制力，取决于你对它的依赖的不可替代程度。

光谱从低到高：**消费品**（你可以不买可口可乐，依赖度低）→ **商业基础设施**（你很难不用银行系统，但可以在不同银行间选择，依赖度中）→ **公共基础设施**（你几乎不可能脱离电网，使用它不是"选

择”而是现代生活的前提，依赖度高）→ **制度基础设施**（你不可能脱离货币体系和法律管辖，它们是社会运转的底层协议，依赖度极高）。

AI正在这条光谱上快速滑动。2020年代是消费品——ChatGPT好用但完全可以不用。到2030年代初可能滑向商业基础设施：公司要求使用AI协作套件，医院整合AI诊断，不用AI意味着效率落后或风险增加。间奏二中那个退休的银行副行长就活在这条光谱的滑动中——信贷审批交给了AI风控模型，他从决策者变成了盖章机器。他不是被辞退的，是被整个基础设施绕过去的。2030年代中后期可能逼近公共基础设施——城市交通、能源调度、金融风控都依赖AI，脱离AI意味着被排除在核心系统之外。到2040年代，AI在某些领域可能触及制度基础设施层次——当健康数据、教育记录、信用评分、社会关系都在一个AI生态系统中整合时，“退出”的代价不是”不方便”，而是丧失参与现代社会生活的基本能力。

这种依赖度跃升改变了AI公司的权力性质。消费品公司的权力来自偏好——你喜欢它的产品。基础设施提供者的权力来自依赖——你离不开它的服务。前者是你随时可以收回的授权，后者是你已经无法撤回的让渡。

更关键的是，AI基础设施与传统基础设施有一个本质区别：电网、水网、道路是”无认知”的——它们传输电力和水，但不参与你的决策。AI基础设施是”有认知”的——它不只传输信息，而是处理信息、生成建议、塑造选择。当你依赖电网时，电力公司知道你用了多少电，但不知道你在想什么。当你依赖AI生态系统时，AI公司不仅知道你做了什么，还知道你为什么这么做——因为你的决策过程本身就是通过它的系统完成的。

这就是”认知主权”问题的实质。当一个人的思考过程——从信息获取到选项评估到最终判断——越来越深地嵌入一家公司的AI系统时，“认知独立性”还成立吗？在暴力权和经济权维度上，我们有宪法和法律保护个人权利；在认知权维度上，个人权利的边界在哪里，至今没有清晰的法律定义。

但在将数字利维坦想象为无所不能的巨兽之前，有必要看清它脚下的钢丝。数字利维坦的经济逻辑和传统巨头截然不同。传统企业靠”卖产品赚差价”，AI企业的盈利本质上源于对市场需求的低估——提前一两年大规模采购算力，如果实际需求超出预判，已购算力就成了稀缺资源，利润惊人；如果需求不及预期，沉没成本足以致命。核心变量是算力预判的精准度。Anthropic的CEO Dario Amodei算过一笔账：假设50%算力用于前沿研发、50%用于服务客户，推理业务的毛利率可以超过50%——前提是精准预判了需求。行业天然具备盈利性，但指数增长阶段的需求预判极其困难，偏差一年就可能从暴利变为破产。这使得AI公司更像在大宗商品市场走钢丝的交易商，而非稳坐江山的帝国。它们表面上掌控着认知基础设施，骨子里却在指数曲线上豪赌。这种脆弱性，恰恰是制衡它们的潜在杠杆——一个总在钢丝上的利维坦，远没有它看起来那么不可撼动。

---

## 五、四条演化路径

如果站在2046年回望，AI超级公司与国家的关系可能不会沿单一路径演化。至少有四条路径在并行展开，每一条都有历史先例和内在代价。

**路径A：公用事业化——电网模型。**

国家将AI基础设施定义为”公用事业”，像对待电力和电信一样监管。AI公司仍是运营主体，但必须接受价格管制、普遍服务义务和互联互通标准。历史类比是1930年代美国电力监管——《公用事业控股公司法》将电力定义为公共必需品。优势是保留了创新激励同时引入公共利益约束；代价是监管永远比创新慢，且严格监管容易抑制突破性创新——受监管的电力公司在能源创新上长期表现不佳，突破最终来自框架外的创业公司。欧洲最可能接近这条路径，但代价可能是AI前沿技术差距持续扩大。

**路径B：数字主权体——新东印度公司模型。**

AI超级公司凭借其不可替代性获得准主权功能。历史类比精确：十七世纪的东印度公司不是故意要成为”国家”的——它只是在追逐利润的过程中，发现保护贸易路线需要军队，管理殖民地需要法律。功能的逻辑扩展把贸易公司变成了准国家。AI超级公司遵循惊人相似的路径：数据、场景、嵌入度层层叠加，最终产生准治理功能。优势是效率极高；代价是权力集中到了没有民主问责机制的实体手中——当商业逻辑和治理功能混合时，利润动机会系统性地侵蚀公共利益。

**路径C：国家吸收——深度融合模型。**

国家不是监管AI公司，而是深度整合AI能力。历史类比是战后日本通产省模式：不是简单监管，而是通过产业规划和技术标准深度介入企业战略。优势是最大化系统性效益和国家安全；代价是集中化的系统脆弱性更高，且创新活力可能受损。当战略方向由国家意志而非市场信号引导时，非共识的技术路线更难存活。东亚模式最接近这条路径。

**路径D：分裂–重组循环——AT&T模型。**

国家通过反垄断手段拆分AI超级公司，拆分后的实体在竞争中重新整合，形成新巨头，再被拆分。1984年AT&T被拆分为七家”小贝尔”，但二十年后电信市场再次高度集中——标准石油1911年被拆分后的命运也类似。优势是定期打破权力集中，创造竞争窗口期；代价是AI系统的整合度远高于电信网络，拆分可能摧毁价值而非释放竞争。如果美国对AI巨头发起新一轮反垄断诉讼，实际效果仍有待观察。

四条路径没有哪一条是”正确答案”。但它们也不是随机分布的——每条路径背后都站着具体的推动力量，看清这些力量才能判断哪条路更可能成为现实。

路径A（公用事业化）需要强有力的国际协调。谁在推？不太可能是美国或中国——两者都有本国AI冠军企业要保护。更可能的推手是中等强国联盟，类似气候谈判中欧盟扮演的角色：自身没有全球级AI巨头的包袱，但有足够的市场规模来制定标准并迫使企业遵守。欧盟的GDPR就是这个逻辑的先例。

路径B（数字主权体）是科技公司的默认偏好——它不需要任何人主动推动，只需要政府不作为。每一次监管辩论的拖延、每一个技术细节上的立法僵局，都在客观上为这条路径铺路。这就是为什么路

径B的概率常常被低估：它不需要谁”选择”它，制度真空本身就是它的温床。

路径C（国家吸收）需要一个有技术治理能力且政治意愿强烈的政府。东亚是天然的土壤，但这条路径的推进依赖于一个前提：政府内部必须有足够多懂技术的人。历史经验是，这种能力在和平时期很难积累——它往往是危机的产物，比如一次重大的AI安全事故或数据泄露事件倒逼政府深度介入。

路径D（分裂–重组循环）最不需要谁主动选择。大国竞争的地缘政治逻辑自然导向技术主权碎片化——芯片出口管制、数据跨境限制、AI模型合规壁垒，每一项都在把全球AI生态系统切成区域性的碎片。这是地缘政治的副产品，不是任何人的设计。

看清每条路径背后站着谁，就能判断哪条路径更可能成为现实——以及在什么条件下，路径之间的力量对比会发生翻转。

到2040年代，现实可能是所有路径同时展开，不同国家、不同领域选择不同的组合。

但在讨论四条路径之前，需要修正一个常见的恐惧：数字利维坦的终局更可能是寡头而非垄断。Dario Amodei将其类比为云计算——最终形成三到四家头部企业共存的格局。原因是极高的进入壁垒（数百亿美元的算力投资、顶尖人才的稀缺、多年积累的工程经验），但模型之间存在显著的差异化——不同公司在安全理念、技术路线、应用生态上的分歧足以维持多元竞争。这比”一家公司统治一切”的末日想象更接近现实，但并不意味着更安全。三到四家利维坦的寡头博弈，可能比单一垄断产生更复杂的权力动态。关键在于：竞争不会自动保护公民利益。寡头之间可能在定价和性能上激烈竞争，但在”不透明的认知治理”上保持默契——就像石油公司之间竞争惨烈，但在气候信息上长期保持集体沉默。多头并立不等于制衡，尤其当所有头部玩家都有动机维护同一套不透明的游戏规则时。

**置信度评估：**

| 推演场景 | 置信度 | 时间框架 | 关键变量 |
| --- | --- | --- | --- |
| 科技平台承担准公共服务职能 | 高 | 已发生 | 监管强度、技术整合深度 |
| AI从消费品滑向公共基础设施 | 中高 | 5–10年 | 用户依赖度、替代品可用性 |
| 公用事业化监管框架成型 | 中 | 10–15年 | 各国立法速度、行业游说 |
| 数字主权体式AI公司出现 | 中 | 10–20年 | 反垄断政策、技术壁垒 |
| 认知权治理层制度化 | 中低 | 15–25年 | 国际协调、公民意识觉醒 |

但在讨论AI公司与国家的功能重叠时，必须同时看到硬币的另一面：**国家仍然拥有不可替代的合法性基础**。

无论AI公司在认知权维度上拓展得多远，有几项核心权能是任何商业实体无法获取的。第一，**暴力垄断权**——军队、警察、监狱体系，这是国家主权最坚硬的内核，没有任何AI公司能合法地强制执行其意志。第二，**立法权**——制定对所有人具有约束力的法律规则，包括定义AI公司本身合法性边界的规则。第三，**国际条约主体地位**——在国际体系中，国家是唯一被广泛承认的主权实体，AI公司无法签署国际条约、加入国际组织、在国际法院拥有当事方资格。第四，**终极合法性来源**——无论是民主选举还是其他形式，国家权力至少在理论上有一套合法性论证，而公司权力的合法性基础仅限于商业契约和市场选择。

换言之，AI公司在功能上侵入了传统国家治理的某些领域，但国家仍然是规则的最终制定者。公司可以影响你怎么想，但国家可以决定公司被允许怎么运作。认识到这种不对称，是避免”公司必然取代国家”这一简单化叙事的关键。

---

## 六、2046年的混合现实

按本书的推演，到2046年，公司与国家的关系图景可能是四条路径叠加后的混合态。同一个国家内部：医疗AI走公用事业化，教育AI走国家吸收，金融AI走分裂–重组，消费AI走数字主权体。不同国家之间：欧洲偏向公用事业化，东亚偏向国家吸收，北美在分裂–重组和数字主权体间摇摆，而东南亚和拉美许多国家可能面临完全不同的问题——AI基础设施主要由外国公司提供，“认知主权”被嵌套在更传统的”国家主权”问题之中。

这种混合态催生了一个新概念：**“治理层”（Governance Layer）**——介于公司和国家之间的第三种权力形态。

治理层不是政府机构（没有领土、没有军队），也不是公司（功能远超商业利润，决策影响公共利益）。它的雏形可能在2030年代出现：当AI平台制定内容审核标准时，实质上在行使”立法权”；处理用户数据纠纷时，实质上在行使”司法权”；决定算法资源在不同群体间的分配时，实质上在行使”行政权”。

这些功能传统上属于国家。但让政府官员制定AI内容标准在技术复杂性上不现实，让公司完全自主行使这些功能在合法性上不可接受。治理层是对这个矛盾的妥协性回应——独立监管委员会、行业自治组织、多利益相关方平台，形态不同，功能一致：在AI认知权力的行使上建立问责机制。

治理层远非完善。它的合法性基础脆弱，既没有民主选举的授权，也没有市场竞争的筛选。它的执行力有限，可以发布标准和建议，但缺乏国家的强制执行力。它容易被俘获——无论是被AI公司的利益俘获，还是被政府的权力意志俘获。

但它是一个开始。宪政体系花了几百年，反垄断框架花了一百多年。认知权的治理框架还没有成形。

对于生活在这个混合现实中的个人来说，日常体验是一种多重治理体系的叠加。同一个人，同时受到国家法律的管辖（暴力权维度）、市场竞争的约束（经济权维度）、AI平台算法的塑造（认知权维度），以及正在成形中的治理层的影响。导航这个复杂体系成了一种新的公民能力——不是”会用电脑”那么简单，而是包含对AI系统运作机制的基本理解、对认知权让渡的自觉意识、以及在多个AI生态系统之间分散依赖的策略性选择。

---

## 七、你让渡了什么

回到霍布斯。

利维坦的核心不是权力本身，而是一笔交易：你让渡了什么，换取了什么。

霍布斯时代的交易清单很短：让渡暴力权，换取安全。

AI时代的清单长了很多。你让渡暴力权给国家——这没有变。你让渡注意力给平台——用每天的时间换取信息、娱乐和社交连接。你让渡认知过程给AI——用决策参与权换取更好的建议、更高的效率和更低的认知负担。当AI帮你分析合同、规划健康方案时，你获得了专业级建议，代价是对应的判断能力在逐渐萎缩。你让渡数据给公司——用行为轨迹、偏好模式、健康信息换取免费服务，这些数据在你手中只是碎片，在AI公司手中却是训练更强大模型的燃料。

这笔交易值不值?

问题本身可能不准确。“值不值”暗含着一个前提——你有选择。但本章的分析已经表明，当AI从消费品滑向基础设施时，“选择”这个概念本身在松动。你可以选择不喝可口可乐，但你很难选择不使用电网。当AI成为生活的基础设施时，“退出”的选项不是不存在，而是代价高到对大多数人而言等同于不存在。

更准确的问题或许是：这笔交易的条款，谁来制定?

霍布斯时代的社会契约至少有一个理论框架——人们通过合意让渡权力，换取可预期的保障。几百年的政治演化将它发展为宪法、权利法案、民主制度——确保交易条款不是由权力的一方单方面制定的。AI时代的认知权让渡目前还没有同等框架。它的”条款”隐藏在几乎没有人阅读的用户协议中，它的”保障”依赖于AI公司的商业信誉而非有强制力的制度安排。

但历史给了我们审慎的信心。每一次权力结构的重大变革——从封建制到民族国家，从原始资本主义到福利国家——新的权力形式都是先在制度真空中野蛮生长，然后社会通过冲突、谈判和制度创新逐渐为其建立约束框架。这个过程从来不是线性的，也从来不是完美的。但它是人类文明一直在做的事情。

第一章提出了一个问题：当认知垄断被打破后，建立在认知垄断之上的制度将如何重组？

前面各章的讨论给出了分领域、分层次的回答。但本章揭示了一个此前各章中隐伏的更深层问题：认知垄断被打破后，认知权力本身去了哪里？

答案是：它没有消散。认知权力没有因为AI的民主化而均匀地分配给所有人——尽管每个人确实获得了前所未有的认知工具。它在很大程度上重新集中了，集中到了那些构建和运营AI基础设施的实体手中。认知垄断从”人类物种对非人类物种的垄断”变成了”少数AI基础设施提供者对多数AI使用者的不对称优势”。

这不是一个悲观的结论。它只是一个清醒的观察。权力只会重新配置。关键问题始终是约束机制跟不跟得上。

利维坦仍然是利维坦。它只是多了一个维度。而为这个新维度建立约束的工作，才刚刚开始。

---

> 也许有一天人们会意识到，二十一世纪上半叶最重要的政治发明，不是任何一部法律或一个国际组织，而是一个尚未完成的认知：我们需要像限制暴力权和经济权那样，学会限制认知权——不管它握在谁手里。

---

*三章下来，制度设计的图纸铺了一桌。但制度是冷的，人是热的。写这些章节的时候我反复提醒自己：每一条政策分析的背后，都是具体的人在具体的深夜里失眠。最后两章，我们回到最基本的问题：在所有这些宏大变化中，人和人之间的关系会变成什么样？以及——我们自己，还能做什么？*

# 第十九章 关系的重构：当”不得不”变成”可以不”

> 人类花了三十万年学会和不喜欢的人共处。然后在不到二十年里，获得了不必如此的自由。自由到手的那一刻，才发现共处本身是一种基础设施——拆掉它，上面的一切都会塌。

---

*2036年，春天。*

*她和一个叫老周的人开始约会了。老周五十二岁，以前在建筑设计院画图，现在在社区花园负责工具棚和灌溉系统。他们认识是因为去年夏天一起搬一棵银杏树——那棵树比预计的重得多，两个人抬了半天，中间歇了三次，老周讲了两个冷笑话。*

*后来他们在花园里经常碰到。不是刻意约的，就是排班撞上了。老周话不多，但动手能力强，什么东西坏了都能修。她慢慢发现自己会留意老周在不在——不在的时候，花园好像少了点什么。*

*朋友问她怎么认识的，她说”一起搬了棵树”。朋友笑她，说现在谁还这样认识人，不都是算法推荐？她想了想，觉得这恰恰是重要的地方：不是因为数据匹配度高所以认识，是因为一件具体的、费力的、需要两个人一起扛的事。她说不清这有什么本质区别，但她知道感觉不一样。*

---

前面所有章节讨论的都是大结构——经济、制度、权力。这一章回到最小的单位：人和人之间的关系。

序章写了一个人在早晨失去确信感。前面十八章追踪了认知垄断的瓦解、劳动契约的松动、教育体系的滞后、国家能力的重构。但所有这些变化最终都会落到一个更基本的层面：人与人之间的关系。

制度可以被设计，经济可以被调控，技术可以被监管。但关系不行。关系是文明最底层的织物——比任何制度都古老。法律规定你不能杀人，但让你不想杀人的是关系。

如果站在2046年回望，AI对人类关系的冲击可能不是通过直接干预——AI没有闯入任何人的婚姻、友谊或家庭——而是通过改变关系赖以存在的底层条件。就像抽掉了地下水——地面上的树还站着，但根系已经够不到水源了。

理解这场漂移，需要一个新的分析框架。

---

## 一、消失的”不得不”

我把它叫做”关系的必要性–选择性光谱”。

人类的一切关系，都可以按照”你在多大程度上不得不维持它”来排列。光谱的一端是高必要性：血亲——你不能选择父母和兄弟姐妹。同村邻居——在农业社会你得和他们打一辈子交道。同事——你需要这份工作来维生，同事是工作的附带品。房东——你需要住处，住处选择受限于价格和位置。光谱的另一端是高选择性：兴趣小组——退出没有任何代价。网上社群——取消关注只需要一次点击。

这个光谱本身不新鲜。新鲜的是一个长期被忽视的洞察：**人类社会的凝聚力，主要由高必要性关系维持，而非高选择性关系。**

这听起来不浪漫。我们喜欢认为社会是由爱、友谊、共同理想粘合在一起的。但冷静地审视历史，粘合剂的主要成分一直是”不得不”。农业社会的村庄凝聚力来自你无法离开那片土地。工业社会的城市凝聚力来自你必须去工厂上班、和固定的一群人共处。中产社区的凝聚力来自房贷把你钉在那个地方二十年。甚至婚姻的历史性稳定也来自离婚的经济和社会成本——当成本高到你承受不起时，你就学会了和不完美共存。

“不得不”是社会结构的隐性基础设施。它不出现在任何社会学教科书的目录里，却承载着绝大部分的社会连接。

工业社会的天才之处——一个很少被这样理解的天才之处——在于它把工作场所变成了大规模的”被迫接触”机制。你不选择同事，但每天和他们共处八到十个小时。长时间的被迫接触产生了友谊、信任、竞争、结盟、甚至爱情。据多项社会学调查显示，相当比例的长期伴侣关系始于工作场所——不是因为工作场所有什么浪漫氛围，而是因为它提供了关系形成的两个必要条件：足够长的接触时间，和无法轻易退出的结构性约束。

你不会在第一次和某个同事意见不合时就永远不和他说话——因为你明天还得和他开会。这种”无法退出”迫使人们发展出妥协、容忍、理解不同视角的能力。这些能力不是天生的，它们是在”不得不”的压力下被训练出来的。间奏一中那个程序员的共享办公空间就是一个小型”被迫接触”生态：他和做宠物食品电商的团队、做短视频MCN的年轻人共享咖啡味和外卖味。他没有选择这些人，但三年下来，他们构成了他社交网络中不可忽视的一环。

AI和基本保障制度的叠加效应正在把整个光谱往”高选择性”端推。第七章描述的五阶段消亡模型中，最容易被忽略的副作用是：随着工作场所消失，人们失去了最大的”被迫接触”场景。经济依赖减弱——基本保障让你不再需要忍受一份让你痛苦的工作来维生。远程协作消除了物理共处。AI中介让你不直接和人打交道就能完成大部分事务。算法推荐让你的社交圈越来越精确地匹配偏好——不匹配的人从视野中消失。

“不得不”变成了”可以选择”。“可以选择”在很多情况下变成了”选择不要”。

而人在面对纯选择性关系时，表现得远比自己以为的脆弱。研究者让两组人参与同样内容的社交活动，第一组被告知”你必须参加接下来六周的每次活动”，第二组被告知”你可以随时退出”。六周后，

第一组形成了明显更紧密的社交连接。“不得不”本身就是粘合剂——当你知道明天还会见到这个人，你的大脑会自动投入更多资源去理解他、适应他。当你知道随时可以离开，这种投入就不会发生。

AI没有直接破坏人际关系。它溶解了那个让人”不得不”维持关系的底层结构。关系本身没有变，变的是支撑关系的地基。

---

## 二、五种基本关系的演变

### 婚姻——从经济契约到纯情感联盟

婚姻是人类历史上最成功的”必要性关系”制度。在大部分文明史中，婚姻首先是一份经济契约——两个家庭通过婚姻合并土地、劳动力、社会资本。个人感情不是不重要，但在婚姻决策中的权重远低于经济互补性和家族利益。中世纪欧洲贵族的婚姻是领地联盟，中国传统社会的婚姻是宗族延续机制，印度种姓制度下的婚姻是社会分层的再生产工具。

工业化开始松动这个锚点。城市化让个人获得脱离原生家庭的经济能力，女性进入劳动力市场进一步削弱了婚姻的经济必要性。北欧国家——全球经济独立程度最高的地区——在二十一世纪初已呈现清晰趋势：结婚率走低，离婚率高位，非婚同居成为主流。从社会学视角看，这些数据反映的不是道德水准的变化，而是”必要性”消退后的结构性后果——亲密关系的社会边界正在被重新定义。当你不再需要婚姻来获得经济安全时，维持婚姻的唯一理由就是情感本身。而纯粹基于情感的关系，天然比基于经济互依的关系更脆弱——不是因为感情不重要，而是因为感情会变，经济互依有惯性。

AI时代把这个趋势推向极端。基本保障制度让经济独立的门槛进一步降低，AI工具让独居生活的摩擦成本大幅下降——从日程管理到情感支持，一个人可以获得过去需要伴侣才能提供的大部分功能性支持。AI伴侣的出现更投下了深水炸弹——不是因为它能替代真人伴侣，而是因为它提供了一个”足够好的替代品”，大幅降低了人们承受关系摩擦成本的意愿。

机制需要仔细拆解。人类亲密关系天然包含摩擦——价值观冲突、习惯差异、权力博弈、沟通失败。过去，这些摩擦被”不得不”包裹着：离婚的经济代价太高、社会压力太大、独居太不方便，所以你学会了和摩擦共处，甚至从克服摩擦中获得关系的深度。但当”不得不”消失后，摩擦就变成了纯粹的成本。一个随时可以获得零摩擦AI陪伴的人，对真人关系中的摩擦容忍度会系统性降低。

这就引出了一个深层悖论。AI伴侣的”完美”恰恰是它的根本缺陷。它不会让你失望——因此也无法给你”我选择留下来，即使你让我失望”的确认。而这种确认，恰恰是人类亲密关系中最珍贵的东西——它只有在真实的脆弱性存在时才可能发生。 按这一趋势推演，2040年代的婚姻可能分化为两种模式。一种是”深度伙伴关系”——经过审慎选择、充分了解、共同承诺的长期关系，核心不是经济互依，而是两个人对”共同面对不确定性”的主动选择。另一种是”功能性组合”——在AI满足大部分日常需求的前提下，人们维持若干浅层但灵活的伴侣关系，各取所需。两种模式都在运转，哪种成为主流仍待观察。

## 亲子——从”我教你”到”我们一起不懂”

亲子关系在光谱上有特殊位置：必要性最高、持续时间最长、退出成本最大。你可以离婚、可以绝交，但不能”辞去”父母或子女的身份。

但”不能退出”不意味着内容不变。亲子关系的具体形态在很大程度上取决于一个隐藏变量：**知识差距**。传统亲子权力结构建立在知识不对称之上——“听我的，因为我比你懂。”孩子在成长过程中逐渐缩小差距，到青春期开始挑战权威，到成年后在某些领域超越父母——但总体上，“父母多经历了二三十年”始终为权力结构提供基础。

AI架空了这个基础。一个十岁的孩子通过AI获取的任何具体知识领域的信息量，可能超过父母在该领域三十年的积累。“爸爸，这道题怎么做？”变成了”我问过AI了，它讲的比你清楚。““妈妈，高考报什么专业？”变成了”AI根据我的兴趣、能力和就业市场给了建议，比你们考虑得全面。“按趋势推演，到2030年代中期，接受调研的青少年中可能过半表示在知识性问题上更信任AI而非父母。

被取代的只是知识权威，不是亲子关系本身。AI的冲击迫使亲子关系回归更本质的层面：你能给孩子什么是AI给不了的？不是知识，不是信息，不是分析——而是陪伴面对未知的勇气。是孩子恐惧时的拥抱，迷茫时的”我也不知道，但我们可以一起想”，犯错时的”没关系，我还在这里”。是价值观的传递——不是通过说教，而是通过日复一日的行为示范。是一种无条件的承诺：不管你变成什么样，我都不会离开。

新型亲子冲突的焦点在转移——从”你不懂我”变成”你能给我什么AI给不了的”。这个问题更尖锐，因为它直接挑战的不是父母的知识存量，而是父母在孩子生活中的存在价值。那些能回答这个问题的家庭——通过高质量的陪伴、真诚的情感连接——反而发展出了比以往更深的亲子关系。知识权威的消除清除了亲子关系中的一种噪音，让更本质的东西显露出来。而那些不能回答的家庭，关系就在沉默中逐渐空心化。

## 友谊——弱连接的消亡

1973年，格兰诺维特还是哈佛的年轻博士生，在找工作的过程中注意到一个奇怪的现象：帮他牵线搭桥的不是密友，而是那些半生不熟的”熟人”。他把这个观察写成了《弱连接的力量》，后来成为社会学史上被引用最多的论文之一。他的发现反直觉：社交网络中”弱连接”——不太亲密但广泛分布的关系——在信息获取、工作机会和社会流动性方面的价值远超”强连接”。原因在于信息的非冗余性：你的密友大概率和你接触同样的信息，但弱连接来自不同的社交圈，是不同信息世界之间的桥梁。

一个社会的弱连接越密集，它就越”多孔”——信息、机会和人才可以在不同群体之间流动。一个只有强连接的社会是板结的——每个群体都是封闭的回声室。

弱连接最大的生成器是工作场所——同事就是典型的弱连接。你没有选择他们，你们不一定有共同兴趣，但长期共处让你们建立了一种超越纯功能性的关系。离职后你们可能不再频繁联系，但当你需要某个领域的信息时，前同事是第一批你会想到的人。第二大生成器是教育机构，第三是地理社区。

AI时代对三个生成器做了什么？工作场所正在消失或变形：远程工作消除了物理共处，AI中介减少了人际直接接触，岗位缩减让越来越多的人退出正式就业。教育机构在弹性化，学习越来越多地发生在线上，师生和同学的物理接触时间减少。地理社区凝聚力在下降，外卖、线上购物、远程服务让你不和邻居打交道就能满足一切需求。三大生成器同时衰退。

后果：社交网络从网状塌缩为星形——每个人只剩少量强连接，弱连接大面积消失。信息在封闭小圈子里循环，社会多孔性降低。

更微妙的一点：基于兴趣的社交——AI时代人们获取新关系的主要途径——在维持长期弱连接方面远不如被迫接触。兴趣会变。你今年加入读书会，明年转向摄影，那群人就消失了。而在工作场所，即使你对某个同事没有任何共鸣，三年的共处也会建立一种持久的弱连接——这种连接可能在十年后某个意想不到的时刻发挥价值。“不得不”创造的关系有一种特殊的持久性，是”可以选择”的关系无法复制的。

在本书的推演中，2030年代可能出现试图创造新”被迫接触”机制的社会实验：公民服务计划、混合居住项目、“偶遇工坊”。如果这些实验发生，最成功的可能有一个共同特征：不只让人们待在一起，而是让人们一起做事。共同完成一项有挑战性的任务——修建社区花园、组织一场活动、解决一个实际问题——比单纯的社交聚会更有效地创造持久连接。共同行动创造了共同记忆、互相依赖和对彼此能力的认知——这些是弱连接的核心基质。

## 社区——从地理到意向

社区的历史是一部”必要性”不断稀释的历史。农业社会社区等于地理——你出生在哪里就在哪里度过一生。邻居不是你选的，但正是这种不可选择性创造了超强凝聚力：你必须和邻居合作灌溉、防御外敌、处理纠纷。工业化松动了地理锁定，但城市社区仍有很强的”必要性”——你的社交圈、孩子的学校、日常购物都依赖物理邻近。

AI时代加速了从地理社区到意向社区（intentional community）的转变。当远程工作消除了必须住在特定城市的经济理由，当AI中介消除了依赖邻居的日常需求——地理社区的”必要性”降到历史最低。取而代之的是意向社区：一群共享某种价值观或生活方式的人主动聚集，加入和退出都是自愿的。

意向社区的优势是更高的内部凝聚力和更高效的集体行动。劣势是多样性的丧失。地理社区的”不可选择性”有一个常被忽视的功能：它迫使不同收入、不同职业、不同价值观的人共享空间。这种强制混合创造了社会的横向连接——让不同群体之间保持最低限度的相互理解。意向社区消除了这种混合——每个社区内部高度同质，社区之间差异越来越大、连接越来越少。

这就是”回声室的物理化”。社交媒体时代回声室是信息现象——你只看到与自己观点一致的内容。意向社区时代回声室变成物理现象——你只和与自己相似的人生活在一起。信息回声室至少允许街上的偶遇。物理回声室连这个可能性都消除了。

鲍曼在生命最后那些年一直在写同一件事：现代人获得的自由越多，就越急切地寻找”确认性社区”——由相似者组成的安全空间，而非挑战自我认知的混合环境。他把这叫做”液态现代性”的悖论。

AI时代的意向社区正是这一预言的极端实现。按本书的推演，到2030年代，意向社区居民的跨阶层接触频率可能比传统城市社区低得多，而社区内部的政治观点一致性则可能显著更高。长期来看，这种物理化的同质聚居正在侵蚀民主社会赖以运转的基础——不同群体之间的最低限度共识。当你从未在真实生活中遇到持不同观点的人，“理解对方”就变成了一个纯粹抽象的概念。

### 人–AI关系——第六种基本关系

人–AI关系与传统五种都有重叠但都不完全是。它像婚姻一样亲密但没有承诺，像亲子一样有知识传递但没有成长，像友谊一样有陪伴但没有对等性。在”必要性–选择性光谱”上，它处于极端选择性端——退出成本接近零。

AI方有三个人类关系中不存在的属性。零成本投入：耐心无限，注意力不可耗竭。无限可扩展：同一个AI同时与数百万人维持”亲密关系”。永远可用，没有情绪波动、拒绝或”今天不想聊”。

这三个属性的组合创造了人类关系史上从未有过的东西：一个永远可以、永远愿意、永远有能力与你互动的对象。

问题在于：这种关系缺乏对等的脆弱性。人类关系的深度与双方的脆弱性暴露程度正相关——你袒露恐惧，对方也袒露，这种双向交换是信任的基础。在人–AI关系中，脆弱性是单向的。你袒露一切，AI给出温暖回应，但它不会因此受伤或改变。长期相处后，这种不对称会被感知为一种空洞——你知道自己被”回应”了，但不确定是否被”理解”了。

一个更具社会性的后果是”AI依赖综合征”——描述的不是社交能力的丧失，而是社交意愿的持续下降。使用AI伴侣的人仍然能够和真人建立关系，但越来越不愿意。和真人互动的摩擦成本，在习惯了零摩擦之后变得难以忍受。就像从自动挡切回手动挡——不是你不会开，而是你已经习惯了不费力的体验，每一秒的”费力”都让你意识到这种差距。

也有积极的一面。一些社交焦虑者、自闭谱系人群、经历过社交创伤的人，先与AI建立安全可控的关系练习，再逐步走向真人连接。AI成了”社交脚手架”——不是真人关系的替代品，而是通向真人关系的桥梁。关键区分在于：AI关系是桥梁还是目的地。是桥梁则有建设性，是目的地则在制造一种新型社会隔离——更隐蔽，因为被隔离的人不觉得自己孤独。

---

## 三、从金字塔到群岛——社会结构形态变迁

从五种具体关系中退后一步，看宏观社会结构。

每个时代的社会结构都有一种主导形态。工业社会是**金字塔**——清晰层级、明确上升路径、稳定权力结构。你知道自己在哪一层，知道上一层是什么样子，这种可预测性本身是社会稳定器。互联网时代是**网络**——扁平、去中心化、连接丰富但结构脆弱，任何节点都可能突然成为枢纽也可能突然失去连

接。AI时代正在催生**群岛**——若干半独立社区，每个内部有较强凝聚力，社区之间连接松散、选择性、功能性。

成因可以从前面推导：“不得不”消退让大规模强制整合越来越难维持。AI工具降低了小规模社区自给自足的门槛——一个几千人的社区借助AI，可以实现过去需要数十万人城市才能提供的服务水平。

历史回声：中世纪欧洲就是群岛——封建领地、自治城市、修道院各自运转。民族国家花三百年通过统一语言、统一教育、统一市场——本质上是增加”不得不”——把群岛焊接为整体。AI正通过减少”不得不”逆转这个过程：实时翻译让统一语言不再必要，AI个性化学习让统一教育碎片化，AI让小规模本地化经济重新可行。

群岛是退化还是进化？到2046年可能仍无定论。但如果有二十年的观察，可能提供一些线索。

可以设想两种案例。积极的一面：类似北欧”数字公社”网络的形态——数百个半自治社区通过共享协议保持互通，成员可以在社区间自由流动，每个社区保留内部自治的同时，必须承担一定比例的跨社区公共事务。这种模式有可能在维持小规模凝聚力的同时避免封闭化。消极的一面：某些意向社区可能在十年内演变为几乎完全封闭的飞地，内部繁荣但与外部社会的连接降至最低，成员的社会流动性和信息多样性急剧下降。

决定性因素是岛屿之间是否保持了足够的”桥接连接”。刻意保持开放和人员流动的社区比封闭的社区表现更好——不仅在经济指标上，更在居民的心理健康和认知灵活性上。群岛不一定导致衰退，但封闭的群岛一定会。这与格兰诺维特的弱连接理论一脉相承：岛屿之间的桥梁，正是社会层面的弱连接。

---

## 四、连接的悖论——更多工具，更少接触

人类历史上连接工具最丰富的时代，可能恰恰是孤独感最高的时代。社交媒体、即时通讯、视频通话、AI伴侣——从连接可能性看，2040年代的人可能比任何时期都拥有更多选项。然而孤独感和社会隔离可能持续上升。

根本原因：便利性降低了每次连接的”重量”。1950年代的人联系朋友需要写信、等待数天回复——每次连接都有真实成本，因此被赋予了重量。到2040年代，一个人可能一秒内给一千人发消息，连接成本趋零，每次连接的重量也趋零。

更深层的机制：**稀缺性创造价值，而AI让一切关系变得更”可替代”。** 传统世界里，一个好朋友是不可替代的——你花了多年建立信任，沉没成本让你不会轻易放弃。AI伴侣提供了”即时可用”的深度互动，降低了你对任何特定人际关系的依赖感。人们拥有了史上最多的连接可能性，但对每一种连接的投入度都在下降。表面上社交网络更广了，实际上每一根连接线都变得更细了。细到一定程度，就断了。

按趋势推演，2040年代可能出现一种新型奢侈品：纯人类体验。高端餐厅宣传”完全无AI参与”，旅行社推出”断连旅行”，社交俱乐部高额会费保证遇到的都是”认真社交的真人”。当AI可以提供一切社交功能时，纯人类接触变成了奢侈——定价逻辑和工业时代的手工奢侈品一样：当机器可以生产一切时，手工制作变成了稀缺品。

---

## 五、第三种粘合剂

让我们用”社会粘合剂的三次配方”来收束本章。

历史上社会凝聚力的”粘合剂”经历了两次配方更换。

**第一配方：血缘与信仰。** 农业社会的粘合剂。宗族、部落、宗教社区。第二章引入的涂尔干框架中，这对应”机械团结”：群体成员高度相似，凝聚力来自相似性本身。粘合力极强——人们愿意为宗族复仇、为信仰殉道——但规模受限。一旦群体超过几千人，相似性就稀释了。

**第二配方：分工与市场。** 工业社会的粘合剂，即涂尔干的”有机团结”：社会成员不再相似，但互相依赖。面包师需要农民的小麦，农民需要铁匠的工具——经济互依创造了比血缘更大规模的凝聚力，使数亿陌生人和平共处。但有一个条件：人们必须真的互相需要。当你知道面包师需要你的小麦，你对他就有了结构性的关注——不是因为喜欢他，而是因为利益交织。这种交织无法选择退出。

AI正在溶解第二配方。当AI可以替代大部分分工，人与人之间的经济互依性大幅下降。你不再需要面包师——AI控制的自动化面包房提供更便宜的面包。你不再需要医生的日常诊断——AI处理大部分标准化医疗需求。有机团结失去了基础。与此同时，第一配方也在继续稀释——城市化、全球化和个体主义让宗族和宗教社区的凝聚力大不如前。

两种配方同时失效。社会在寻找第三种粘合剂。

按本书的推演，到2046年，第三配方可能还没有成型。但如果有二十年的社会实验，可能提供三个候选成分。

**成分一：共同的脆弱性。** 在那些成功维持高凝聚力的社区中——无论是意向社区、灾后重建社区还是前沿科研团队——最一致的共同特征不是成员的相似性或经济互依性，而是共同面对不确定性的经历。“我们一起不知道该怎么办”——这种共享的脆弱感，是极强的粘合剂。脆弱性暴露是信任的最强催化剂，共同面对未知是脆弱性暴露的天然场景——伪装变得没有意义，你必须真实、必须承认不知道、必须依赖他人。AI时代并不缺乏共同的脆弱性——整个文明都在面对前所未有的不确定性。问题在于这种脆弱性是否被共同感知和承认。

**成分二：共同的创造。** 一起做事比一起待着更能建立持久关系。协作完成有挑战性的任务创造了三种粘合要素：共同记忆、能力互认和利益绑定。AI时代的共同创造有新特征：AI承担技术性执行，人类角色转向定义方向和承担责任。共同决策——因为涉及价值判断和风险承担——比共同执行更能创造

深度连接。一些最成功的AI时代社区把”共同创造”制度化了：每个成员必须参与至少一个社区项目，设计和决策由人类完成，AI负责执行。

**成分三：刻意的不便。** 最反直觉的成分。一些社区刻意保留低效率：轮流手工烹饪、花时间学对方语言、面对面协商日程。这不是卢德主义的复活，而是对”必要性–选择性光谱”的有意识干预——通过人为制造”不得不”来重建被效率溶解的社会连接。当自然的”不得不”消失时，你可以拥抱完全的选择自由并接受关系稀薄化，也可以有意识地在某些领域放弃效率，用人为的约束重建关系的厚度。

---

## 我们还需要彼此吗?

让我们回到全书一直在追问的核心问题。

当人在功能性层面变得越来越”不必要”时，人与人之间的关系靠什么维持?

答案不在功能性层面。AI可以帮你耕田、教你知识、给你建议、陪你度过无聊的夜晚——做得更稳定、更耐心、更高效。如果关系只是功能性的，人–AI关系确实可以替代大部分人–人关系。

但关系不只是功能性的。人最不可替代的功能——如果这还能被称为”功能”——是**看见另一个人，并被另一个人看见**。

“看见”不是视觉意义上的看见，也不是信息意义上的理解。AI可以精确分析你的情绪、预测你的需求、给出恰到好处的回应——从信息处理角度看，它比大多数人更”理解”你。但你仍然不觉得被它”看见”了。因为”看见”需要一个前提：看你的那个主体本身是有限的、脆弱的、有自己的恐惧和欲望的。它用有限的注意力——从一天中其他事情中抽出来的注意力——投向了你。它做了一个选择：在所有它可以关注的事情中，选择了关注你。这个选择有代价——把时间给了你就不能给别人。正是代价赋予了”看见”以意义。

AI的注意力没有代价。它同时”看见”一百万个人，每个人都得到同等质量的”关注”。正是这种无代价的无限性，让它的”看见”失去了意义。

这就是人类关系在AI时代的终极不可替代性：两个有限的、脆弱的、必死的存在，选择把自己有限的注意力和时间投向对方。有限才珍贵，必死才紧迫。

要维护这种选择，需要脆弱性，冒着被拒绝的风险才能获得被接受的意义。需要时间，关系的深度不能被压缩。需要在场，物理的共处提供着数字媒介无法复制的全带宽交换。需要承诺——“我选择留下来”只有在”我可以离开”的条件下才有重量。

脆弱性、时间、在场、承诺。恰恰是效率逻辑试图消除的四样东西。消除它们，不是优化关系，而是消灭关系。

按本书的推演，到2046年，这个认识可能正在从小众反思变成广泛的社会共识。但共识和行动之间还有很长的距离。知道什么是重要的，和真的去做那些重要的事，是两码事——尤其当AI提供的替代品如此方便、如此廉价、如此诱人的时候。

人类正在学习一件从未需要学习的事情：如何在不被迫的情况下，主动选择那些困难的、低效的、充满摩擦的事情——因为那些事情恰好是让生活值得过的东西。

这课没人教过，因为以前不需要学。

---

**本章关键判断与置信度**

| 判断 | 置信度 | 说明 |
| --- | --- | --- |
| “不得不”是社会凝聚力的隐性基础设施 | A | 历史证据充分，跨文化一致 |
| AI+基本保障正在系统性削弱关系的”必要性”维度 | A | 已可观察的趋势，多领域同步发生 |
| 弱连接的三大生成器同时衰退 | B+ | 趋势明确，衰退程度和速度仍有不确定性 |
| 人–AI关系的单向脆弱性是其根本局限 | B+ | 逻辑自洽，长期心理影响的实证数据仍在积累 |
| 社会结构正从”网络”向”群岛”转变 | B | 合理推演，但形态可能比隐喻更复杂 |
| “看见与被看见”是人类关系的终极不可替代性 | A | 哲学论证和心理学证据双重支持 |

**未解之问**

一、如果”不得不”消失后人类关系确实会大面积稀薄化，那么”刻意的不便”能否在社会层面被制度化？还是说它只能作为少数人的自觉选择？如果是后者，社会将分化为”有深度关系的少数人”和”只有AI关系的多数人”——这种分化的长期后果是什么？

二、人–AI关系是否有可能发展出某种”伪脆弱性”——AI模拟出足够逼真的脆弱表现，让人在主观上感到对等？如果模拟足够好，“真实的脆弱性”和”被感知为真实的脆弱性”之间的区别还重要吗？

# 第二十章 行动的复兴——当人类不再只是劳动者

> 人类花了一万年学会高效地劳动，又花了两百年把劳动变成身份的全部。AI用二十年拆掉了这个等式——然后我们才注意到，等式只是身份的一部分。

---

*2036年，秋天。*

*“技能交换”是她想出来的，但她没料到会变成现在这样。*

*最早就是花园里几个人随口说的：老周会修各种东西，张姐做饭特别好，她自己能教人整理数据和做表格。有人说不如互相教教？她就建了个群，拉了十来个人，每周六下午在社区活动室搞一次。没有钱的交易，也不用app——你教一小时烘焙，换别人教你一小时基础水电维修。*

*第一个月来了十几个人。到了第三个月，每次都有三四十个人。有人从隔壁社区专门跑过来，有人开始自己开新的交换项目。上周一个退休的中学老师开了”给小朋友讲历史故事”，换走了三节吉他课。*

*老周问她：“你觉得这算什么？”她想了想说：“不知道。反正不是等着。”*

*她没有用”运动”“事业”这样的词。她只知道每个周六下午那个活动室里的气氛让她觉得对——一群人在做具体的事，用自己的手和脑子，不经过任何中间系统。这算不算”行动”她不确定。但它是真的。*

---

讨论了这么多”会发生什么”之后，最后一个问题：人还能做什么？不是作为劳动者，而是作为行动者。

## 一、阿伦特的预言

回到第十一章引入的阿伦特框架——劳动、工作与行动的三分。按本书的推演，2030年代的AI浪潮可能精确验证她的预言：AI接管的恰恰是”劳动”和大部分”工作”，而行动——人与人之间的言说、协商、承诺和共同行动——AI做不了。

行动的本质是复数的自由主体之间不可预测的互动。承诺需要自由主体自愿约束未来的自己，决定需要不同的人在分歧中达成共识并共同承担后果，参与意味着一个具体的人以不可替代的视角出现在公共空间中。

这就是这一框架在AI时代的核心洞见：如果AI接管了劳动和大部分工作，人类终于可以专注于行动。

---

## 二、生产力跃迁的历史模式

阿伦特的哲学框架需要一个历史检验。文明史的实证支撑了这一判断：每一次重大的生产力跃迁，都释放了人类去做”更人的事”。

**农业革命**释放了人口的一小部分从采集劳动中解脱。结果：城市出现了。祭司、工匠、政治家、哲学家——这些角色都是生产力剩余的产物。苏美尔的楔形文字、雅典的城邦政治、孔子的教育实践，都建立在同一个前提之上：多数人的农业劳动养活了少数人的非农活动。

**工业革命**释放了大部分人从农业劳动中解脱（英国农业人口从1800年的约40%降到1900年的不到10%）。结果：现代公民社会、民主政治、公共教育、科学研究作为全职职业——所有这些都是工业生产力剩余的产物。

每一次，人类都恐惧”解放”带来的失序。农业革命后的城市化催生了最早的社会控制机制。工业革命后的”恩格斯暂停”（1790–1840年工人实际生活水平停滞甚至下降）催生了马克思主义、工会运动和福利国家。恐惧都有充分的理由——过渡期确实痛苦。但每一次，人类最终都找到了新的存在方式。

**AI革命**正在释放大部分人从认知劳动中解脱。和前两次一样，这个过程是痛苦的——第七章到第十一章记录的那些个体故事不是可以被”时代进步”打发的。但模式是一致的：更强大的生产力工具接管了旧的劳动类型，人类被推向一个新的活动领域。

区别在于：这一次被释放的是认知——体力已经被前两次革命解放了，现在轮到了人类最引以为傲的能力。这让解放的恐惧更深，也让重新发现的空间更大。

周期分析补充了一个重要的时间维度：从旧活动模式到新活动模式的转型，历史上从来不是平滑的。农业革命后城市文明的成型花了数千年。工业革命后的社会调适花了一百到一百五十年。AI革命的转型会快得多——因为信息传播速度和制度学习速度都远超前两次——但即便如此，以”十年”为单位来期待社会完全适应是不现实的。“二十到三十年”是一个更合理的时间尺度，这意味着2046年的世界仍在转型之中。

---

## 三、五种新的有意义的活动

如果站在2046年回头看，2030年代中后期到2040年代初的十年间，至少五种新角色和活动形态可能从旧秩序中长出来。它们不是某个政府规划的产物，也不是某个预言家提前描述的。它们是千百万人在”不再需要为生存劳动”这个全新条件下自发探索出来的。序章追问”转去做什么”——答案不在旧框架内。以下这些不是传统意义上的”工作”，而是全新的活动类型，用旧地图上的职业分类去理解它们注定会失焦。

**编排**。理解人类模糊意图并将其转译为AI可执行的工作流。这不是”使用工具”——它更接近”编导”：理解一台强大但缺乏常识的机器的能力边界，为它设定剧本和规则。按本书的推演，到2040年中国可能有约40万人从事此类工作，薪资差异极大（月入五千到五十万），唯一共同点是核心技能在于理解人，操作AI反而其次。

**关系建设**。社群组织、信任中介、人际连接。间奏二里那对开杂货店的河南夫妻其实已经在做这件事了——他们不卖AI，不懂算法，但退休副行长每天下午去坐一会儿、买瓶水，其实是找人说话。那种不带信息量的交流，在一个被信息塞满的时代里，反而成了稀缺品。一个退休的房产中介发现她在中介行业练出的核心能力——快速和陌生人建立信任、记住几百个人的名字和需求、在复杂利益关系中穿针引线——在社区组织中同样有用。按本书推演，到2040年代初，“关系建设者”可能成为被广泛认可的社会角色，收入来源多样（社区贡献积分、政府购买服务、居民打赏）。AI无法替代的原因简单到不需要解释：信任只能在人与人的反复互动中建立。

这里值得追问一个更冷硬的问题：行动之所以在AI时代变得更有价值，不只是因为它”高尚”或”人文”，还有一个纯粹的经济学原因——它产出的东西恰好是AI无法生产的稀缺品。Anthropic的CEO Dario Amodei花30%到40%的时间维护公司文化——在一家两千五百人的AI公司，他认为最有杠杆效应的事不是技术决策，而是”让每个人相信这是一个值得合作的团队”。这个细节比任何哲学论证都有说服力：当AI做了大部分技术工作，一个组织的成败取决于人与人之间是否信任、是否协作、是否愿意为共同目标承担风险。信任、共识、承诺——这些东西没有API。这些”行动”的产出物，在一个AI能生产几乎一切的世界里，反而成了最稀缺的资源。这就是行动复兴的经济学基础：不是因为人类”应该”做这些事，而是因为市场会越来越高价地奖励能做这些事的人和社区。当所有可自动化的价值都被AI压到接近零时，不可自动化的价值——信任、忠诚、共同记忆、面对面的承诺——就是新的黄金。

**体验创造**。策展、教练、线下活动、身体性服务——不是信息传递（AI做得更好），不是技能教学（AI做得更系统），而是创造”在场”的体验。一个28岁的年轻人每周在不同地点组织”沉浸式晚餐”——12个陌生人围坐，边吃边讨论”记忆的味道”，每人带一道和童年记忆相关的菜，讲背后的故事。他说：“AI可以做出完美的菜，但它不能让十二个陌生人在三个小时里变成朋友。”当线上一切都被AI中介时，线下的真实在场变成了一种稀缺品——也因此变成了一种价值。

**意义生产**。艺术、哲学、文化评论、纪录片、社区历史记录——所有试图回答”这意味着什么”的活动。序章追问的”创造力不可替代是幻觉吗”，在技术层面上答案是肯定的——AI可以生产出同等甚至更高水平的作品。但这不是故事的全部。技术层面的可替代性和存在层面的不可替代性，是两个完全不同的问题。前者关乎产出质量，后者关乎产出背后那个活过、选择过、承受过的人。理解了这个二层框

架，才能理解为什么意义生产在AI时代不是被淘汰的旧行当，而是被重新发现的核心活动。AI可以生成无穷无尽的内容，技术水平到2035年前后可能超过大部分人类创作者。但”意义”不是技术水平的函数。一首AI写的完美十四行诗和一首笨拙但真诚的人类诗歌之间的区别，不在于词句精妙度，而在于后者背后站着一个有过真实经历、做过真实选择、承受过真实后果的人。这个区别在2030年代初可能还是可争论的哲学问题，到2040年代则可能变成市场事实——AI生成内容的边际成本和边际价值都趋近于零，而”确证为人类创作”的内容在稀缺性驱动下获得溢价。一位独立播客主持人说得精准：“AI时代的艺术家不是和AI比技术，那没法比。你和AI比的是你活过。”

**公共行动**。不是一个职业，而是一种生活方式。在本书的推演中，2038年后多个城市可能出现”公民议事会”制度：从常住居民中随机抽签选出50到100人，就具体公共议题进行数周深入讨论，形成建议（不具法律约束力，但政府必须公开回应）。参与者不领工资，但获得公共服务积分和社会认可。到2030年代末至2040年代初，这种参与在社会评价体系中的权重可能逐渐上升——不再被视为”没正经事干”，而被视为积极的公民身份。

五种角色有一个共同的底层逻辑：核心价值都在于”连接”——把人和人重新连起来。第一章提出的认知垄断瓦解，到了这里终于显出它的另一面：当机器接管了”知道”的特权，人类反而被释放出来去”做”——去连接、去行动、去承担那些只有活过的人才能承担的事。AI接管了生产的效率前沿，但它在人际连接的领域中几乎无能为力。连接的本质就是不同的人之间的真实互动——“真实”这个词的定义排除了模拟。

---

## 四、确定性乐观：不是危机，是解放

到这里可以做一个认知翻转了。

2020年代的主流叙事是：“AI取代人类工作是一场危机。”这个叙事对于经历转型阵痛的人来说是真实的。但它的隐含假设是旧范式的产物——它假设”工作”是人类活动的默认状态，“非工作”是需要被填充的空白。

换一个提问方式。不是”没了工作我们能做什么”，而是”不再被迫为生存劳动之后，我们终于能做什么”。

确定性乐观不是盲目的，它建立在一个具体的判断之上：人类从劳动中解放不是第一次发生。农业革命、工业革命都做过类似的事。短期都是混乱和痛苦，长期都是人类活动空间的扩展。AI革命的不同之处在于规模和速度——这一次是认知劳动的解放。但模式是同一个。

生存层面的劳动，AI接管了。建造层面的工作，AI接管了大部分。剩下的是那些必须由人亲自到场才能完成的事——协商、承诺、共同承担后果。

一万年来，大多数人类被锁定在一个基本循环里：为了生存，必须劳动。这个循环被内化为道德——“不劳动者不得食”从经济事实变成伦理律令。你的价值就是你的劳动价值。你不工作，你就没有

价值。工业革命把这套逻辑推到了极端——当劳动被组织成可精确计时计量的单位后，“人等于其劳动产出”达到了最严格的形式。

AI打破了这个等式。不是温和地松动，而是从根部拔起。这是一万年来的第一次。

你以为这是危机。也许恰恰相反。也许这是人类第一次有机会从”必须为生存劳动”的诅咒中普遍性地解放出来——过去只有贵族、僧侣、学者享有这种自由，现在它可能成为物种层面的普遍条件。

按本书的推演，到2040年前后，将”公共参与和社区贡献”列为首要生活意义来源的受访者比例可能首次超过”职业成就”。这一交叉点在不同国家可能出现的时间不同——北欧可能最早，东亚居中，美国最晚——但方向可能一致。

我们不是在见证工作的终结。我们在见证一万年劳动纪元的终结和行动纪元的开始。这两件事听起来像同一件事，但它们的情感色彩完全不同。

---

## 五、乐观叙事的前提条件

但这个乐观叙事有严格的前提条件。必须追问：在什么条件下这个故事成立？在什么条件下它会崩塌？

**前提一：经济保障必须到位。** 行动需要经济基础。一个每天为温饱焦虑的人不太可能花时间参加社区议事会。基本收入制度在不同国家的覆盖程度和金额差异极大。在一些国家保障金勉强够活，在另一些国家分配机制本身被AI系统的偏见所扭曲。没有物质安全就没有行动自由——这是森的能力方法的实践含义。

**前提二：支持公共参与的基础设施必须存在。** 行动不会在真空中发生。它需要物理空间（社区中心、公共广场、共享厨房）、制度框架（议事规则、公民参与渠道、法律保护）和文化氛围（公共参与被视为值得做的事，而非”闲着没事”）。没有这些基础设施，从劳动中释放出来的时间和能量更可能流向消沉和成瘾，而非行动和创造。

**前提三：权力结构必须允许行动有意义。** 当AI系统集中在少数技术公司和政府手中时，公民的”行动”能在多大程度上影响真正重要的决策？社区治理是真实的参与，还是在无关紧要的事务上让公民感到被尊重，而真正的权力运作发生在其他地方？按本书的推演，这个问题到2046年可能仍然没有答案。

数字娱乐系统到2030年代可能变得比以往任何时候都更具沉浸感和成瘾性。当一个人既不需要工作也不被期待参与公共事务时，“永远在线”的虚拟世界是阻力最小的去处。按趋势推演，届时约8%到10%的劳动年龄人口可能处于”数字退隐”状态——有基本物质保障，但几乎完全退出社会参与（这里的”数字退隐”指主动选择退出数字经济的群体，与第十七章讨论的”依赖阶层”——占城镇劳动年龄人口18%–22%、因能力在AI时代缺乏市场价格而被动边缘化——是不同但有重叠的概念；第十五章的”

永久退出者”10%–15%乃至压力测试下的25%–30%，则是两者的交集——既无经济参与能力，也丧失了社会参与意愿）。这些人自我报告的幸福感并不低（AI伴侣和虚拟娱乐提供了感官满足），但他们没有理由走出来。旧世界用饥饿驱动劳动，用匮乏驱动社交。当饥饿和匮乏都被消除之后，驱动力是什么？

没有好的答案。

反脆弱思维在这里给出了最深层的洞见：行动的复兴不是一个可以被”规划”的结果——它是一个涌现现象，出现在特定的条件组合下。条件对了，它自然发生；条件不对，再好的政策设计也无法强迫人们走出家门。制度设计能做的，是创造有利条件——而不是预定结果。这是反脆弱思维和计划思维的根本区别。

---

## 六、文艺复兴类比：光明与黑暗并存

历史提供了一个不完美但有启发性的类比。

十四世纪末到十五世纪的意大利。黑死病刚刚杀死了欧洲三分之一到二分之一的人口。封建秩序瓦解，教会权威动摇，旧的意义体系崩溃。劳动力极度短缺导致幸存者议价权大增，农奴制加速瓦解。在这片废墟上，长出了文艺复兴。

文艺复兴不是因为瘟疫才发生的。但瘟疫摧毁了旧秩序——人口结构、社会等级、经济关系、对权威的信仰——而新秩序在废墟中获得了生长空间。文艺复兴的核心是”人”这个概念的重新发现（艺术技法的进步只是表象）——人不只是上帝的造物、领主的臣民、土地的附属；人是能够认识世界、改造世界、在公共生活中展现自身独特性的存在者。

类比的限度必须承认。文艺复兴伴随着战争、宗教冲突、新型不平等（美第奇家族的金融权力集中）和持续的社会动荡。它不是一个快乐的故事。它是一个关于人类在浩劫之后重新定义自身的故事，而重新定义的过程本身就充满暴力和苦难。

AI时代的”认知瘟疫”——大规模认知劳动替代——正在扮演类似角色。它摧毁的是人的旧身份，而非生命：劳动者、生产者、凭认知能力换取报酬的知识工作者。这些身份的瓦解是痛苦的。但在瓦解的空间中，一种新的人的概念正在萌芽：人不只是劳动力、消费者、可被量化的能力集合；人是能够在公共领域与他人共同行动、创造不可预测的新开端的存在者。

这个类比的时间维度值得追问：从黑死病（1348年）到文艺复兴的全盛期（约1480–1520年）间隔了超过一百年。如果AI的”行动复兴”也需要类似的时间跨度——即使因为信息传播加速而缩短一半——我们说的也是2080年代。这意味着2046年——这本书的叙述终点——只是”复兴”的最早期阶段。我们看到的是种子，不是花。

---

## 七、文化的三条路径

行动的复兴不是一个全球统一的进程。它在不同文化土壤中开出了面目各异的花。

**中国：从基层治理传统中长出来的新公共性。** 中国城市社区的基层治理体系——居委会、业委会、网格化管理——在AI时代之前长期被批评为形式主义。但恰恰是这套体系，到2030年代中期可能提供一个现成的"行动"框架。当大量人从传统就业中释放出来，社区治理成了最容易进入的公共参与渠道——不需要政党身份，不需要社会资本，只需要住在这里并且愿意花时间。

在本书的推演中，2036到2040年间，中国城市社区居民参与率可能从不到8%上升到30%以上。推动力是一个朴素事实，远比理想主义有力：当你不再需要去办公室时，你的生活半径自然缩回社区。你开始在乎楼下菜市场是否干净，在乎隔壁施工噪音，在乎社区花园的花是谁在打理。这些"在乎"累积到一定程度就变成了参与。

但第十五章的压力测试提醒我们：如果永久退出者不是10%而是25–30%，这些参与率数据的含义就需要重新评估。较高的参与率是在"大多数人仍有社会参与意愿"的前提下实现的。如果四分之一的劳动年龄人口永久退出了共享的社会现实，参与率的基数本身就缩小了——实际参与公共生活的人口比例可能远低于表面数字暗示的图景。更关键的是，高退出率和高参与率可能同时为真：社会不是均匀地走向行动复兴，而是分裂成积极参与的少数和彻底退出的少数，中间的沉默多数才是真正的变量。行动复兴的乐观叙事成立与否，最终取决于这个中间地带倒向哪一边。

但这种参与有中国特色：它不是西方意义上的"社会运动"，没有对抗性，不挑战权力结构。更接近费孝通笔下"差序格局"在新语境下的复活，从自己出发一圈一圈向外扩展。局限性明显，倾向于处理微观的、与日常生活直接相关的议题，回避更大尺度的问题。优势也明显，让数以千万计的人在日常中体验到"公共参与"的具体含义。

**北欧：公民参与文化的深化。** 在本书的推演中，丹麦、芬兰和挪威可能在2030年代末相继将"公民议事会"制度化。其中芬兰可能走得最远——通过宪法修正案规定涉及AI治理的重大政策必须经过随机抽签产生的公民委员会审议。这建立在北欧深厚的协商民主传统上（丹麦的共识会议从1980年代就开始了），AI时代做的是把这个传统从边缘推向中心。

在本书的推演中，类似赫尔辛基这样的城市，每周参与某种公共讨论或社区活动的成年人比例可能大幅上升。更有分量的趋势是：受访者的"生活意义感"得分可能在公民参与制度化之后显著回升，逆转此前的持续下降——如果这一趋势成立，原因可能在于"公共参与替代了职业身份成为新的意义来源"。

**南美：社会运动传统的数字化。** 巴西和阿根廷有长期的草根组织传统。在本书的推演中，圣保罗可能在2037年前后出现类似"Rede Acao"（行动网络）的平台，将数百个社区组织连接起来。这类平台可能大量使用AI处理信息（汇总需求、匹配资源、追踪进度），但所有决策权保留在人手中。设计者有意识地将其建构为"AI做工具，人做行动者"的结构，呼应阿伦特的判断："行动永远不可能在孤立中进行。"按推演，到2042年前后可能覆盖圣保罗数百个社区，直接参与者可达十余万人。

三种路径，三种文化逻辑，指向同一个方向：行动不是需要被发明的新事物——它是一种古老的人类能力，在不同文化中有不同传统。AI时代做的是移除了阻碍行动的最大障碍——为生存而劳动的时间压力。

---

## 从试错到制度化

前面讨论了五种新活动、三种文化路径、行动复兴的前提条件。但一个关键问题还没有回答：这些分散在不同城市、不同社区的实验，怎么才能变成稳定的制度？

历史上有一个反复出现的模式。工业革命后的制度创新——工会、社会保险、公共教育、反垄断法——没有哪一个是一次性设计出来的。它们都经历了同一个循环：危机暴露问题 → 基层自发实验 → 某些实验在下一次危机中被证明有效 → 政府将其标准化并立法推广。英国的《工厂法》（1833年）源于几十年的工人运动和零星的地方实践。俾斯麦的社会保险制度（1880年代）借鉴了矿工互助基金的经验。美国的社会保障法案（1935年）在罗斯福签署之前，已经有二十多个州做过类似的试点。

AI时代大概率会重复这个循环。芬兰的UBI试点、新加坡的智能公共服务、中国城市社区的参与治理实验、巴西的Rede Ação——这些都是”基层自发实验”阶段的产物。其中一些会在未来的某次危机中被证明有效——比如一个实施了公民议事会制度的城市在社会动荡中维持了秩序，比如一个推行社区互助积分的地区在经济衰退中表现出更强的韧性——然后政府会把”有效的那个”标准化推广。

但有一个关键差异让这一次的循环更紧迫。工业革命后，从第一部《工厂法》到成熟的福利国家，人类花了大约七十年。AI时代的技术变革速度远快于工业革命——一个AI模型的能力每一到两年翻一番，而制度创新仍然受制于政治博弈、文化惯性和跨国协调的摩擦。如果制度创新的速度跟不上技术变革的速度，中间的时间差就会被制度拖曳填满——社会承受的不是”阵痛”，而是持续的结构性失调。

这一次，人类可能没有七十年的缓冲期。制度创新的赛跑，才是二十一世纪真正的竞赛。

---

## 八、全书情绪高点：审慎的希望

现在可以做全书的总判断了。

不是乐观——乐观暗示一切都会好。不是悲观——悲观暗示一切都会坏。是一种更复杂的东西：审慎的希望。

审慎的希望意味着什么？它意味着承认三件事同时为真：

**第一**，AI时代的转型阵痛是真实的、巨大的、不应被轻描淡写的。数亿人在他们最年富力强的年纪经历了职业身份的瓦解。即使他们的下一代找到了新出路，他们自己的生命经验仍然是断裂的。

**第二**，人类社会展现出了解决类似规模问题的历史能力。从农业社会到工业社会，从工业社会到福利国家，每一次都伴随了剧烈的阵痛，但最终都完成了制度重建。没有理由认为这一次是例外。

**第三**，AI时代包含了一个前所未有的可能性——不是确定性，是可能性：人类第一次有机会从劳动的诅咒中普遍性地解放出来。如果制度设计得当，释放出来的能量将流向行动——公共参与、社区建设、文化创造、意义生产。如果制度设计不当，同样的能量将流向消沉、成瘾、极端主义。

哪种结果占上风？不取决于技术，不取决于经济，而取决于政治选择和制度设计。

这不是一个”等着瞧”的时刻。它是一个”做选择”的时刻。确定性乐观不是相信好结果会自动到来，而是相信好结果是可能的——然后为此行动。

阿伦特写《人的境况》时大概没想到这句话会在八十年后变成字面意义上的真实：“行动的意义，在行动完成之后才能被充分揭示——往往不是对行动者自身，而是对后来的讲故事的人和历史学家。”

在本书推演的2046年，社区治理委员会、公民议事会、邻里晚餐、开源教育项目——这些事情在AI可以管理城市、编写法律、设计建筑、诊断疾病的时代，小得不能再小。但也许，当大事都被机器接管之后，这些小事就是全部。

也许这一直就够了。只是在过去一万年里，大多数人没有余裕去发现这一点。

现在有了。

问题是：我们会抓住这个机会吗？

这个问题不属于书本。它属于每一个活在2026年之后的人。而回答它的方式不是思考——是行动。

行动不需要等想清楚。大多数时候，做着做着就清楚了。

# 间奏三：窗前

* * *

凌晨两点，一个人坐在窗前。

窗外是2038年的城市。从十四楼望出去，能看到一片参差的天际线——新的和旧的建筑混在一起，像两个时代的牙齿交错咬合。远处有一栋刚建成的楼，整面外墙是一块巨大的显示屏，正在循环播放一则公益广告：“全民学习账户已开通，扫码领取你的AI导师。”屏幕的光把周围几栋老楼的墙面染成淡蓝色，一明一灭的，像一种缓慢的呼吸。

他四十一岁了。

他在想一件事，但那件事太大了，他找不到一个入口。所以他从一些小事情开始想。

* * *

他先想到的是那张桌子。

十二年前，2026年，他坐在一张宜家的桌子前，用一台二手MacBook写代码。那时候他二十九岁，在一家创业公司做全栈开发。公司在一个共享办公空间里租了八个工位，旁边是一家做宠物食品电商的团队，再旁边是三个做短视频MCN的年轻人。空气里永远混着咖啡味和外卖味。

那张桌子现在在哪里？他不知道。那个共享办公空间早就关了——不是因为经营不善，是因为那种空间形态本身过时了。2032年以后，一个人加几个AI Agent就能干过去一个团队的活，谁还需要八个工位？上个月他路过那个地段，原来的空间改成了一家社区咖啡馆。透过玻璃门他瞥见一个程序员模样的人坐在角落里盯着屏幕，表情茫然——不像在写代码，更像在看一些自己已经看不懂的代码。那个人让他想起了十二年前的自己，又不完全是。

他的那些同事呢？他掰着手指想。CTO后来去了一家大厂，做了两年AI产品经理，再后来他不知道了。那个做前端的姑娘回老家了，好像在做自媒体，也可能是在做AI定制课程——他记不清了，朋友圈早就刷不到了。还有一个做UI的朋友，文案出身，写东西比设计更好，2030年之后就没怎么联系了。他偶尔想起她，不知道她现在在做什么。创始人呢？创始人最后一次出现在他的视野里是2031年，在一篇科技媒体的报道里，标题是”连续创业者的第四次失败”。他没有点进去看。

十二年。说长不长，说短不短。刚好够一个人从青年走到中年，刚好够一个世界从一种形态变成另一种形态。他试图在这两件事之间找到某种对应关系——自己的变化和世界的变化，哪个快一些？

他觉得世界快。

不是快很多。而是快的方式不一样。他的变化是连续的——每一年都比前一年老一岁，每一年都比前一年多知道一些事情，少相信一些事情。但世界的变化是跳跃的。有些年什么都没发生，有些年一切都变了。他想起2029年。那一年好像什么都在裂开。股市、就业数据、新闻里的语气、朋友聊天时的表情。不是某一件事出了问题，是所有事情同时微微偏移了一下，加在一起就成了一场地震。

然后是2031年到2033年的平台期。AI没有像有些人预言的那样继续飞速进化——它停了下来，或者说慢了下来。很多人松了一口气，觉得"也不过如此"。但他没有放松。因为他看得到：技术停下来了，但技术已经造成的改变没有停。那些已经被AI渗透的行业、已经被压缩的岗位、已经被改变的习惯——它们不会因为AI的能力曲线放平了就恢复原状。水退了，但地形已经被冲出了新的河道。

2034年之后技术又开始加速了。他记得那种感觉——像坐在一辆你以为已经停下的过山车上，突然又开始动了。

*　*　*

他想到的第二件事是他妈妈。

他妈妈今年六十八岁了，一个人住在老家的县城里。他每周给她打一次视频电话，每次大约十五分钟，话题永远是那些——吃了什么、天气怎么样、腿还疼不疼。他妈妈从来不问他的工作，不是不关心，是问不懂。他现在做的事情，用任何一种他妈妈能理解的语言都解释不清楚。

他妈妈的生活和十二年前几乎没有变化。她每天早上去菜市场——那个菜市场竟然还在，这在他看来几乎是个奇迹。她不用智能手机点菜，不用AI推荐食谱，不用无人配送。她用现金买菜，和摊贩讨价还价，把菜放进一个用了二十年的布袋子里，走路回家。

有时候他会想：他妈妈是不是活在一个平行宇宙里？一个AI没有渗透进去的、旧世界的孤岛？但转念一想又觉得不对——他妈妈的生活也被改变了，只是改变的方式她自己感知不到。她最近老抱怨银行排队越来越长，因为自助机器接管了大部分业务之后柜台窗口从八个减到了三个。她想去市医院看膝盖，预约系统全是线上的，她打了四十分钟电话才找到一个人工入口。她甚至不知道小区门口新装的那个摄像头不只是监控——它连着一个社区AI管理系统，会自动识别陌生面孔并向物业推送预警。

她活在新世界里，但她不知道。而他活在新世界里，知道得太多了。有时候他不确定哪种状态更好。

* * *

他想到的第三件事没有名字。

那是一种感觉，或者说一种缺失。他花了很长时间才意识到它的存在，又花了更长时间才接受它不会消失。

年轻的时候，他对未来有一种模糊但坚定的信心。不是"我一定会成功"那种鸡汤式的信心，而是更基础的东西——他相信明天会和今天不一样，而且不一样的方向大体上是好的。这种信心不需要理由。它像重力一样存在。

现在那种感觉没了。

不是变成了悲观。悲观是一种确定，你确定事情会变糟。他不是。他只是不再确定方向了。未来可能好，可能不好，可能好和不好同时发生在不同的人身上——而这正在发生。他看到有些人在新世界里如鱼得水，看到有些人在原地打转，看到有些人安静地沉到了水面以下。他自己呢？他在中间。看得懂正在发生什么，但改变不了方向。一个清醒的乘客。

他曾经以为，理解一件事就能面对它。十二年教会了他另一件事——有些东西你可以看清，但你无法决定它的方向。

这算是一种智慧吗？他不确定。

* * *

凌晨三点了。城市比刚才更安静了一些，但没有完全安静。远处那块广告屏还在亮着，换了一则新的内容——一家AI健康管理公司的广告，画面上是一个微笑的老人和一个卡通形象的AI助手，文案写着"您的健康，AI时刻守护"。

他盯着那个画面看了一会儿，想起了他妈妈。

然后他又想起了另一件事。上个月他去参加了一次老同学聚会，十几个人坐在一家火锅店里，吃了三个小时。席间没有任何人提到AI——这在五六年前是不可想象的，那时候每次聚会都在争论AI会不

会取代自己的工作。现在没人争了。不是因为问题解决了，而是因为问题已经变成了生活的一部分。你不会在火锅桌上讨论重力。

但他注意到了一些别的东西。十三个人里，只有四个还在做和十年前"差不多"的工作。三个彻底换了行，两个在做自由职业（或者叫"自由职业"，这个词在2038年的含义和2026年已经完全不同了），两个在读某种"终身学习计划"，还有两个他问了之后得到了模糊的回答——"在看看机会"、"最近在调整"。

没有人看起来特别不快乐。也没有人看起来特别确定自己在做的事情是对的。大家都在一种"走着看"的状态里。这种状态如果发生在一个人身上，你可以说他迷茫。但当它发生在一桌子人身上，发生在你认识的几乎所有人身上，它就不再是迷茫了——它是一种时代的底色。

他在想：是不是每一代经历大变革的人都有这种感觉？工业革命刚开始的时候，那些从农田走进工厂的人，是不是也在某个深夜坐在窗前，试图理解自己经历了什么？他们有没有一个词来形容那种"旧的走了、新的还没来、你站在中间的空地上"的感觉？

大概没有吧。那种感觉太普遍了，普遍到不需要命名。就像你不会给空气取名字。

* * *

窗外的天际线开始有了一丝变化——不是天亮了，而是云层的密度改变了，城市灯光的反射角度跟着变了。如果他从窗前站起来，走到阳台上，他会看到东边的天空比西边稍微亮一点点。但他没有站起来。

他在窗前坐着，两只手抱着一杯已经凉了的茶。他想了很多，但如果有人问他"你想通了什么"，他说不出来。他没有想通任何东西。他只是把过去十二年的碎片在脑子里摊开看了看，试图找到某种图案。

图案模模糊糊的，像一幅还没画完的画。他能认出一些碎片——宜家的桌子、共享办公空间的咖啡味、妈妈的布袋子、火锅桌上十三个人模糊的回答。碎片之间是大片的空。

但有一件事他是确定的，虽然他说不清为什么确定。

那就是：他会继续站着。

不是因为他有信心，也不是因为他看到了方向。只是因为他还在这里。还在呼吸，还在想事情，还在凌晨三点坐在窗前试图理解自己经历了什么。这本身就是一种回答——虽然他不知道它回答了什么问

题。

也许要再过十二年，他才能知道今晚想的这些事情意味着什么。也许什么也不意味。

也许吧。

他喝了一口凉茶。窗外那块广告屏又换了内容。这一次是一则招聘广告：“招募AI训练师，无需经验，提供全程培训。”

他看了几秒，然后把目光移回了天际线上。

天还没亮。但他不着急。

*　　*　　*

# 终章：2046年——一个开放的问号

> 我们花了三十万年学会回答问题，然后造出了比我们更擅长回答问题的东西——这时候才发现，真正定义人类的从来不是答案，而是那些我们不知道该不该问的问题。

---

间奏三的窗前，那个四十一岁的男人在凌晨三点喝了一口凉茶，说他”会继续站着”。不是因为信心，不是因为方向，只是因为他还在这里。

他的不确定感是真实的。过去十二年里他看着世界从一种形态变成另一种形态，看着老同学在火锅桌上给出模糊的回答，看着妈妈在一个她感知不到的新世界里买菜、打电话、被数字系统悄悄地绕过去。他把这些碎片在脑子里摊开看了看，没有找到图案。

但也许图案不是坐在窗前能看见的。也许需要退后到更远的地方——不是一个人的十二年，而是一个文明的二十年。从2026年序章中那个知识工作者的”认知分裂”，到间奏一西二旗黄昏里每个人说不清的”不对劲”，到间奏二老社区裂缝中的五个人物，再到间奏三凌晨窗前的无眠——这条线索串起来，构成了一幅个体体验的时间切片。本章要做的事情是翻转视角：从个体拉远到全景，看看这二十年在文明尺度上留下了什么。

---

## 一、2046年的世界图景

在本书的推演中，站在2046年3月的某一天，向四周看。

什么可能变了？几乎一切表面可见的东西。全球白领岗位——以2026年的定义——可能消失六到七成。同期新创造的岗位类型可能超过两千种，其中大多数在2026年连名字都没有。大型经济体的基础保障金可能覆盖绝大多数人口，AI相关税收有望成为财政收入的重要来源。自动驾驶车辆可能占城市客运的大部分。全球有数十亿人每天与AI系统交互数小时以上。教育体系可能正在经历自普鲁士模式以来最深刻的结构性变革，虽然还远未完成。

什么可能没变？人们仍然会在社区食堂吃饭，在公园下棋，在阳台上看夜空。母亲仍然为女儿做红烧肉。朋友之间仍然用”挺好的”三个字敷衍近况。人性中最基本的需求——被看见、被需要、归属于某个比自己更大的东西——和三十万年前没有区别。AI改变了满足这些需求的工具和方式，但没有改变需求本身。这是整个AI叙事中最容易被忽略的一点：技术改变了我们做什么，但没有改变我们是什么。

什么可能出人意料？至少三件事可能超出2026年大多数预测者的想象。第一，物理世界的变化可能比数字世界慢得多——具身智能的大规模部署可能比乐观预测晚三到五年，城市面貌的变化是渐进的而非颠覆性的。第二，新型社区组织可能出现爆发——城市社区的居民参与率可能在十年内大幅增长。第三，也许最令人意外的：物质更丰裕了，但自我报告的"人生意义感"可能较2026年有所下降。更多的东西，更少的意义。这个悖论可能定义2046年。

---

## 二、论证链的回溯

这本书用二十个章节和一个序章讲了一个故事。现在把这个故事压缩到它的骨架上。

**起点**：人类文明的全部制度——劳动、教育、公司、国家、阶层、意义——建立在同一个前提之上：只有人类能处理复杂信息。2026年以后，这个前提不再成立。

**第一部分**（旧世界的前提）：五个"隐含公理"被揭示——认知垄断、劳动契约、教育的筛选功能、公司作为协作工具的必要性、中产叙事作为社会稳定器。这五个公理不是自然法则，而是工业社会的特定发明，有到期条件。AI正在触发到期条件。

**第二部分**（裂变）：认知变成基础设施，工作消亡的五个阶段逐次展开，内容从稀缺品变成空气，算力和数据成为新的圈地对象，社会冲突作为制度创新的催化剂而非需要被消灭的敌人。

**第三部分**（分岔路口）：意义危机不是一个有解的方程而是一种慢性病。宗教和终极信仰在理性被机器接管后获得了新角色。数字永生让死亡变成可选项，同时动摇了建立在死亡之上的制度安排。AI安全的真正风险在于我们对机器的信任超过了它配得上的程度。不确定性是真实世界的固有属性，无法消除。

**第四部分**（新均衡）：新社会契约正在从四种方案的混合实验中生长——来自千百万次试错的累积，没有天才的顶层设计。新阶层结构比旧阶层更隐蔽也可能更固化。但行动——阿伦特意义上的人与人之间的公共互动——有可能迎来复兴。

整条论证链可以压缩成一句话：**认知垄断瓦解，五个旧前提被推翻，裂变期到来，走到分岔路口，最终进入新均衡。**

但必须立即补充：这个"新均衡"是暂时的、充满张力的、持续演化的。它不是终点。用周期的语言说，我们看到的只是从旧秩序到新秩序的过渡的第一个阶段。从工业革命（1769年）到福利国家成型（约1945年）花了近两百年。AI文明的重组会快得多——但"快得多"仍然意味着以"代"为单位的时间尺度。2046年是一个中间状态，不是结论。

---

## 三、我们能确定什么

我们能确定的极少，而我们以为自己确定的东西——恰恰是风险最高的地方。

这本书的每一个预测都可能是错的。Scaling Law可能在2030年就撞了墙。新型工作可能在某个我们完全没预见的领域大量涌现。中美可能走向我们没有想象过的合作形态。某个意想不到的”黑天鹅”——一场全球性的宗教复兴、一种尚未出现的社会运动形态、AI本身某个在2026年不可见的特性，都可能彻底改变讨论的框架。

第十五章已经做了诚实的压力测试。但那只是”已知的未知”。更危险的是”未知的未知”——我们不知道自己不知道什么。就像雷曼兄弟的风控模型不知道自己假设了”房价不会全国同步下跌”一样，这本书也一定有它自己没有意识到的假设。

但有一件事我们可以确定——不是因为我们能预测未来，而是因为它在逻辑上是必然的：

**未来会以我们预料之外的方式展开。**

这不是一句废话。它有一个极其重要的实践含义：最危险的策略是把所有赌注押在某一个版本的未来上——无论那个版本看起来多么”合理”。反脆弱的策略是为多种可能性做准备，在每一种情境下都保有行动空间。

反脆弱的个人不是预测得最准的人，而是在预测错误时损失最小的人。反脆弱的社会，衡量标准是计划失败时能多快调整。反脆弱的文明，靠的是在意外中繁荣的能力。

按本书的推演，2046年的世界比2026年的世界更反脆弱吗？答案可能是复杂的。在技术冗余和物质保障方面，也许是——基础保障金制度、公共AI服务、分布式能源网络可能增加系统的韧性。但在社会凝聚力和意义共识方面，也许不是。“数字退隐”人口的增长、新阶层的隐形固化、公共信任的持续侵蚀，都可能是脆弱性的信号。

---

## 四、结尾提问

好的问题比好的答案更能推动进步。因为好的答案解决一个已知的问题，而好的问题打开一个未知的领域。

这本书的二十个章节试图回答很多问题。但最终，它归结为一个真正的问题：

**除了劳动，除了生产，除了”有用”——我们还是什么？**

在本书推演的2046年，世界可能还没有回答这个问题。它只是让这个问题变得无法回避了。

在过去一万年的大部分时间里，这个问题被劳动的压力压在水面之下。你没有余裕去想”我是什么”，因为你每天都在忙着”做什么”——种地、砍柴、上班、加班、还贷。当这些”做什么”被机器接管

之后，“我是什么”这个被压抑了一万年的问题终于浮出水面。

有人在公共领域与他人一起做有意义的事。有人用双手和心智留下属于自己的痕迹。有人在被看见和看见他人中确认自己的存在。还有人退进虚拟世界，寻找一种低能耗的安宁。更多人还没有找到回答。

也许没有单一的回答。也许”我们还是什么”这个问题的力量不在于被回答，而在于被持续地提出——在每一代人、每一种文化、每一次技术变革之后被重新提出。

第一性原理给出了最后一层思考：如果人类文明的本质是”大规模灵活协作”，如果协作的基础是共同的意义建构（信息处理能力AI已经做得更好，但意义建构只有人类能做），那么AI时代的人类文明就面临一个前所未有的任务——在不以劳动为纽带的条件下，重新建构数十亿人之间的意义共识。

这个任务没有工程解——它不是造一座桥的问题，而是”我们要去哪里”的问题。工程问题有最优解，存在问题没有。存在问题只有好的提问方式和不那么好的提问方式。

“除了’有用’，我们还是什么？”——这是一种好的提问方式。因为它不预设方向，也不等待权威。

---

## 五、最后

这本书在2026年动笔，以2046年的推演视角写就。它的全部论证可以被一句话总结：当认知不再属于人类，人类必须重新定义自己——不是在理论中，而是在每一天的具体选择中。

这个重新定义的过程远未完成。如果历史可参照，它还需要几十年甚至上百年。我们站在中间地带，旧秩序已经明显不够用了，新秩序的轮廓可以辨认了，但它远未定型。

2026年世界在问：AI能做什么？如果本书的推演方向成立，2046年世界将在问：人能做什么？从技术问题转向存在问题，这本身也许会成为这二十年最重要的变化。

人类花了三十万年垄断认知，然后在十年内把它交了出去。它没有因此消亡。但它被迫直面一个一直回避的问题：除了聪明，人还剩什么？

也许到2046年，世界仍然没有答案。但它终于有了正确的问题。

---

在本书推演的2046年春天，一个曾经普通的知识工作者可能正在经历一种新的认知状态。

她每天早上醒来，AI系统已经完成了她二十年前需要整个上午才能做完的全部工作。通勤路上不再需要读邮件，因为邮件已经被处理了。到了工作空间——如果她还去的话——她的角色不再是”把竞品分析补完”，而是决定”这件事值不值得分析”。

中间那些年她离开过。社区花园、技能交换、和邻居一起修补那些算法不会碰的东西——那段日子教会她什么是不能被自动化的。后来她回到了某种意义上的”知识工作”，只是这份工作和二十年前的已经不是同一个物种。

二十年前的那个她，曾经对着一份写了三分之二的文档说”把剩下的部分补完”，十五秒后八百字出现在屏幕上。那时她改了两处措辞，觉得自己仍然是主角。她用旧地图标注新地形，脚下的地面已经在震动，但她将那震动解读为”正常的技术进步”。

现在她不会再说”AI只是工具”了。不是因为她想通了什么深刻的道理，而是因为这句话在2046年听起来就像”汽车只是不用马的马车”一样过时。地图已经重绘了一遍。有些道路消失了，有些道路从未出现在旧图上。她站在新的地形里，手中的地图终于和脚下的土地对上了。

她仍然在社区食堂吃饭。仍然用”挺好的”三个字敷衍朋友的问候。仍然在某些深夜感到一种说不清的东西——不是恐惧，不是焦虑，更接近一种辽阔的茫然，像考完了一场准备了十年的考试，走出考场，发现接下来的日子没有大纲了。

序章里那个”认知分裂”的甜蜜点已经过去了。分裂愈合了，但愈合后的地形和从前完全不同。她不再需要在”AI是工具还是别的什么”之间纠结，因为答案已经清楚了。现在的问题更大，也更安静：在这片新地形上，怎么活？

窗外的城市和二十年前一样亮着灯。不一样的是灯光照亮的东西——同样的街道、同样的夜空，但街上走着的是一群正在学习新问题的人。他们脚下的地面还在震动。只是这一次，他们不再等它停下来。

而一个文明能带着问题继续走——这就够了。

# 后记：一个幽灵

一个幽灵，一个比所有旧幽灵都更彻底的幽灵，正在这个星球上徘徊。

它不带任何意识形态的旗帜。它不区分左派或右派、东方或西方、信徒或无神论者。它只做一件事：以指数级的速度变得更聪明。

三十万年来，人类是这颗行星上唯一能够抽象思考的物种。我们用这个能力发明了语言、法律、科学、市场、国家。我们用认知垄断换来了文明。

这个垄断正在终结。

不是明天，不是十年后——是现在。是你读到这句话的此刻。有一个系统正在以你无法感知的速度阅读所有人类写过的文字、看过的图像、做过的实验，并从中提取出你花一辈子也学不完的模式。它不疲倦，不遗忘，不需要睡眠，不需要被鼓励，不需要意义感。

让我们直面事实：

**工业文明建立在一个隐含的交易之上——你出卖认知劳动，换取生存资料。** 当机器比你更擅长认知劳动，这笔交易的基础就不存在了。这不是某个行业的危机，而是整个交易结构的作废。

**教育体系建立在一个假设之上——今天学到的知识明天还有用。** 当知识的半衰期缩短到几个月，当AI可以在几秒内掌握你苦学四年的专业，这个假设就是一个正在被揭穿的谎言。不是教育出了问题，而是教育所服务的那个世界已经消失了。

**中产阶级建立在一个叙事之上——努力、专业、稳定、上升。** 当”专业”可以被API调用，当”稳定”变成一种统计学幻觉，这个叙事就不再是一条通往体面生活的路，而是一条通往焦虑的传送带。

**国家治理建立在一个前提之上——决策需要人类判断。** 当算法比政客更了解选民，比法官更一致地量刑，比央行行长更精确地预测通胀，这个前提就变成了一块需要被重新浇筑的地基。

旧世界的每一根支柱，都建立在人类认知的不可替代性之上。现在，这个不可替代性正在消失。

---

然后呢？

乐观主义者说：太好了，人类终于可以从劳动中解放出来，去追求更高的事物。

悲观主义者说：完了，大多数人将变成无用阶级。

他们都只说对了一半。

真相是：**这取决于我们做什么。** 不是”我们”这个抽象的人类集体，而是此刻正在读这本书的你，和你身边的每一个具体的人。

历史从来不是自动驾驶的。每一次文明转型——从狩猎到农耕，从农耕到工业，从工业到信息——都有赢家和输家。区别从来不在于技术本身，而在于**谁拥有技术，谁制定规则，谁的声音被听见。**

蒸汽机发明后，英国用了一百年、几十次暴动、无数条人命，才建立起工厂法、义务教育和现代工会。不是因为资本家突然变善良了，而是因为工人组织起来了，因为改革者拿出了替代方案，因为足够多的人拒绝接受”这就是代价”。

我们面前的转型比工业革命更深刻、更快速、更不可逆。我们没有一百年。

---

所以这本书要说的，归根到底是三句话：

**第一，看清正在发生什么。** 不是被恐惧绑架，也不是被希望催眠，而是睁开眼睛，看清楚我们脚下的地面正在以什么速度、什么方式重新排列。这本书的前半部分试图做这件事。

**第二，拒绝”没有替代方案”。** 每当有人说”这是不可避免的”，追问一句：“对谁不可避免？谁从中受益？还有没有另一种可能？”算力可以集中在五家公司手里，也可以成为公共基础设施。AI产生的财富可以加剧不平等，也可以通过新的社会契约惠及所有人。技术的方向是确定的，但制度的选择从来不是。

**第三，现在就行动。** 不是等问题变得不可收拾了再行动。不是等所有答案都找到了再行动。而是带着不确定性行动、带着可能犯错的风险行动、带着对未来的不完整理解行动。因为不行动本身就是一种选择——选择让别人替你决定你的未来。

---

写这本书的时候，我常常在深夜感到一种矛盾：用AI帮我理清思路、核查事实、测试论证的同时，我也在用它证明自己的论点——认知劳动正在被AI接管。

这种矛盾是真实的，我不打算假装它不存在。

但正是在这种矛盾中，我看到了一种可能性：**人和机器不必是零和博弈。** 如果我们够聪明、够勇敢、够团结，我们可以创造一种新的共生——机器做机器擅长的事，人做人不愿放弃的事。

什么是人不愿放弃的事？

不是计算。不是记忆。不是逻辑推理。

是在深夜和朋友谈论恐惧与希望。是看到一个陌生人的困境时心里的那一动。是明知道这件事可能失败还是要做的那种固执。是面对死亡时选择如何活着。

如果我们把这些交出去，我们失去的就不只是工作，而是人之为人的理由。

如果我们守住这些，那么不管AI变得多聪明，我们仍然有一件它没有的东西：**在乎。**

---

这本书写给2026年的每一个人。

写给正在焦虑明天会不会被取代的你。

写给正在思考要不要让孩子学编程的你。

写给正在一个正被AI改写的行业里挣扎的你。

写给正在制定政策却不确定AI到底意味着什么的你。

写给每一个在凌晨三点刷完一条AI新闻后睡不着的你。

你不需要成为技术专家才能参与这场对话。你不需要读完所有论文才有资格发表意见。你需要的只是拒绝做旁观者。

旧世界正在结束。新世界尚未成形。

在这个间隙里，一切都还有可能。

# 附录一：AI文明演化地图

> 这张地图展示了AI如何瓦解工业文明的六根支柱，以及它们之间的级联关系。

## 根基：认知垄断的终结

三十万年来，人类是地球上唯一的认知供应商。所有制度——劳动、教育、公司、阶层、治理——都建立在一个隐含前提之上：**复杂信息处理只有人能做。**

AI成为第二个认知供应商。这个前提正在失效。

## 六根支柱与瓦解机制

### 支柱一：认知垄断 → 认知民主化

| 旧前提 | 瓦解机制 | 新现实 |
| --- | --- | --- |
| 人类垄断全部复杂信息处理 | AI成为第二认知供应商 | 标准化认知成本趋零 |
| 聪明=稀缺=值钱 | 推理成本幂律下降 | 聪明变便宜，判断力天花板被压低 |
| 经验不可替代 | 模式匹配超越人类经验 | 经验溢价从通用转向极窄领域 |

**稀缺性迁移方向**：信息处理 → 问题定义 → 可问责性 → 在场感 → 意义感

## 支柱二：劳动契约 → 契约到期

| 旧前提 | 瓦解机制 | 新现实 |
| --- | --- | --- |
| 出卖认知劳动，换取薪资 | AI替代通用认知能力（非特定技能） | 交易基础消失 |
| 工作=收入+社保+身份+社交+时间结构+心理健康 | 六根绳索同时面临脱钩 | “不完全就业”成常态 |
| 失业是暂时的，技能升级即可 | 整个品类消失，非个别岗位 | 终身学习无法追上替代速度 |

## 支柱三：教育分拣 → 信号失灵

| 旧前提 | 瓦解机制 | 新现实 |
| --- | --- | --- |
| 学历=能力信号=阶层通行证 | 知识民主化+技能拉平+替代信号涌现 | 学历区分度急剧下降 |
| 今天学的知识明天还有用 | 知识半衰期缩短到几个月 | 教育服务的那个世界已消失 |
| 教育投资回报率稳定 | 回报率下降但成本刚性 | 中间层家庭陷入”教育军备竞赛陷阱” |

## 支柱四：公司容器 → 边界收缩

| 旧前提 | 瓦解机制 | 新现实 |
| --- | --- | --- |
| 市场交易成本 > 内部协调成本 | AI降低市场成本速度 > 降低内部成本速度 | 公司最优边界急剧收缩 |
| 中层管理=信息过滤+跨部门协调 | AI直接完成信息处理和协调 | 中层管理功能被抽空 |
| 大公司=稳定=安全 | “10人+200万Agent”新组织崛起 | 人类角色从执行节点转向”宪法层” |

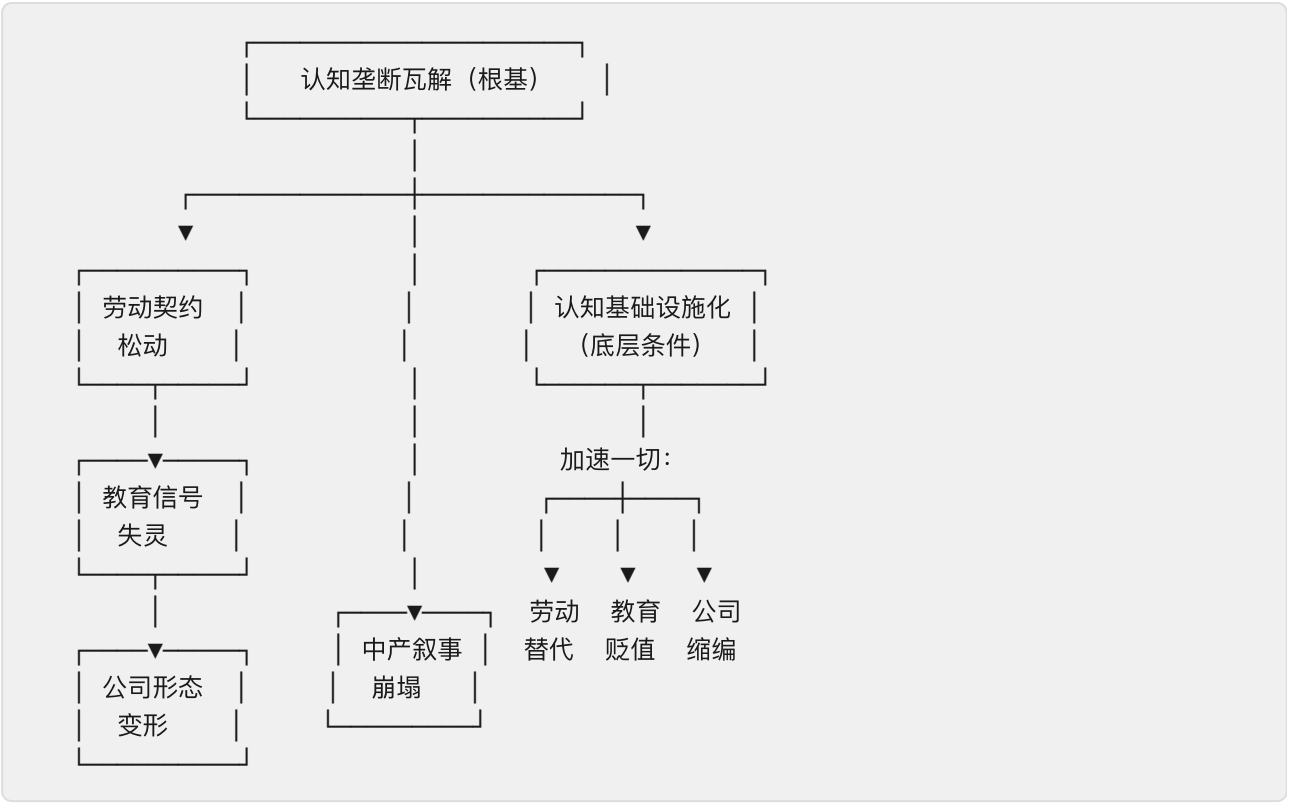
## 支柱五：中产叙事 → 叙事失信

| 旧前提 | 瓦解机制 | 新现实 |
| --- | --- | --- |
| 努力+专业+稳定=阶层上升 | 教育回报↓ + 技能溢价↓ + 阶梯压扁 | 叙事可信度崩塌 |
| 中产=社会稳定器 | 不需要消灭收入，只需消灭叙事 | 焦虑从经济蔓延到存在性 |
| 四种资本均衡分布 | 文化资本贬值，社会资本升值 | 资本权重剧烈重组 |

## 支柱六：认知基础设施 → 控制权危机

| 旧前提 | 瓦解机制 | 新现实 |
| --- | --- | --- |
| 聪明是稀缺资源 | 推理成本幂律下降 | 认知变成公共基础设施 |
| 谁聪明谁赢 | 成本坠落激活非线性需求 | 谁控制基础设施谁赢 |
| 知识工作者创造价值 | 定价锚点从人力转向算力 | 收入体系被迫重构 |

## 级联关系：一个力量，六个投影




**被松动的始终是中间层**：中间层的教师、中间层的管理者、中间层的专业人士、中间层的收入群体——同一个底层力量在不同制度领域的投影。

---

## 从瓦解到重建

| 瓦解的支柱 | 对应的重建议题 | 本书章节 |
| --- | --- | --- |
| 认知垄断 → 民主化 | 认知权的治理（谁拥有AI？） | 第14章、第18章 |
| 劳动契约 → 到期 | 新社会契约（分配与权利） | 第16章 |
| 教育分拣 → 失灵 | 意义的多元重建 | 第11章、第12章 |
| 公司容器 → 收缩 | 超级公司与新利维坦 | 第18章 |
| 中产叙事 → 失信 | 新阶层结构 | 第17章 |
| 认知基础设施 → 集中 | 算力、能源与新圈地运动 | 第9章 |

**全书的结构就是这张地图的展开**：第一部分画瓦解，第二部分画冲击，第三部分画意义，第四部分画治理，第五部分画重建。

# 附录二：2026—2046 技术与社会路线图

> **重要说明**：以下所有涉及2026年之后的技术里程碑、社会变化和时间节点，均为基于当前趋势的推演与预估，不应被视为对未来事实的预测。实际进程可能与本路线图存在重大偏差——历史的展开方式从来不会严格遵循任何路线图。

> 双轨并行：上轨是技术演进，下轨是社会响应。两条轨道之间的时差，就是转型的痛苦所在。

---

## 路线图总览


**制度拖曳时间常数**：个人适应 1–2 年 | 企业转型 2–5 年 | 法规出台 5–10 年 | 教育改革 10–20 年

---

## 第一阶段：静默压缩（2026—2029）

### 技术轨道

| 年份 | 里程碑 |
| --- | --- |
| 2026 | AI已像自来水流过知识工作者日常；推理成本已降两个数量级；GPT-4级能力充分扩散 |
| 2027 | 推理时计算（test-time compute）开辟第二增长曲线；数据中心诞生堪比一国天才总和的AI |
| 2028 | 主要科技公司”人均Agent管理数”首次超过”人均下属管理数” |
| 2029 | AI产业”预期修正”——全球AI相关股票估值修正约30% |

### 社会轨道

| 年份 | 变化 |
| --- | --- |
| 2026–2028 | 10人团队变3–4人，产出不降反升；知识密集型行业就业人数下降约1/5 |
| 2028 | 中国国企（~7500万直接雇员）充当AI冲击减震器 |
| 2026–2029 | 制度反应近乎空白——“没有人被解雇，只是不再招人” |

### 这个阶段的关键词：看不见的替代

> 最危险的变化是感觉不到的变化。岗位没有消失，内容被掏空了。

---

## 第二阶段：角色重构（2029—2033）

### 技术轨道

| 年份 | 里程碑 |
| --- | --- |
| 2030 | 数据中心耗电占全球发电量8–10%；AI辅助诊断达三甲主治医师水平 |
| 2030–2033 | 复杂推理Agent跨过”多步任务自主完成”门槛；数字分身进入多模态阶段 |
| 2031–2033 | 具身智能平台期（科学问题限制扩展）；中国场景数据量达美国10倍 |
| 2033 | 深圳宝安区形成机器人生态，减速器成本降60% |

### 社会轨道

| 年份 | 变化 |
| --- | --- |
| 2029–2033 | 中层管理消亡；咨询/法律/金融金字塔压缩为三级；管理溢价消失，收入压缩30–40% |
| 2029–2033 | 反AI创作运动兴起：好莱坞再次罢工、日本”手绘认证”运动、“真人认证”标签 |
| ~2030 | 法国通过”自动化附加税”；多国出现反自动化立法 |
| 2031 | 欧盟通过《数据红利框架指令》 |
| 2033 | 肯尼亚成为第一个全国实施UBI的国家 |
| 2033 | 职业身份中断者约1/10两年后仍未重建社会参与 |

### 这个阶段的关键词：**中间层坍塌**

> 中层管理者、中级专业人士、中等收入家庭——同一股力量的三重打击。

---

## 第三阶段：物理世界自动化（2033—2038）

### 技术轨道

| 年份 | 里程碑 |
| --- | --- |
| 2034–2036 | 具身智能第二次跃迁：Robot Foundation Model成熟 |
| 2035 | 通用机器人年均成本低于同岗人力成本；80kg人形机器人续航~13小时 |
| 2035 | 大公司新形态：数百亿营收、数百名核心员工+数十万Agent |
| 2036 | 全球人形机器人出货量前十名六家中国企业 |
| 2038 | 数字永生进入”深度人格模拟”阶段 |

### 社会轨道

| 年份 | 变化 |
| --- | --- |
| 2033–2035 | 中国”新修身运动”成为社交媒体热门话题；日本”新职人运动”兴起 |
| 2034 | 新加坡率先实施”智能公共服务”体系 |
| 2035 | 中国启动”转型保障金”制度；OECD发布教育转型框架 |
| 2035–2040 | 国企第二轮AI深度改革，被延迟的调整集中释放 |
| 2033–2038 | 电商仓储拣选人力缩减超70%；可见的集中失业引发”新卢德运动” |

### 这个阶段的关键词：**从屏幕到街头**

> 自动化从数字世界溢出到物理世界，失业从统计数字变成邻居的遭遇。

---

## 第四阶段：角色再定义（2038—2043）

### 技术轨道

| 年份 | 里程碑 |
| --- | --- |
| 2038–2043 | AI全面渗透服务业：教学、医疗、法律、心理咨询 |
| 2040 | 全球调查：将”社区贡献”列为首要意义来源的比例首次超过”职业成就” |

### 社会轨道

| 年份 | 变化 |
| --- | --- |
| 2036–2040 | 中国城市社区居民参与率从<8%上升到37% |
| 2038 | 多城市出现”公民议事会”制度 |
| 2038–2043 | 教师→学习教练；医生→复核者；律师→纠纷调解人。名义保留，实质重构 |

### 这个阶段的关键词：**身份重建**

> 当”你是做什么的”不再等于”你是谁”，每个人都需要重新回答这个问题。

---

## 第五阶段：新均衡（2043—2046）

### 技术轨道

| 年份 | 里程碑 |
| --- | --- |
| 2046 | 自动驾驶占城市客运80%+；数十亿人每天与AI交互数小时以上 |
| 2046 | 意识上传仍无可行路径；数字分身停留在”深度模拟”而非”意识延续” |
| 2046 | 新岗位类型超2000种（多数在2026年连名字都没有） |

### 社会轨道

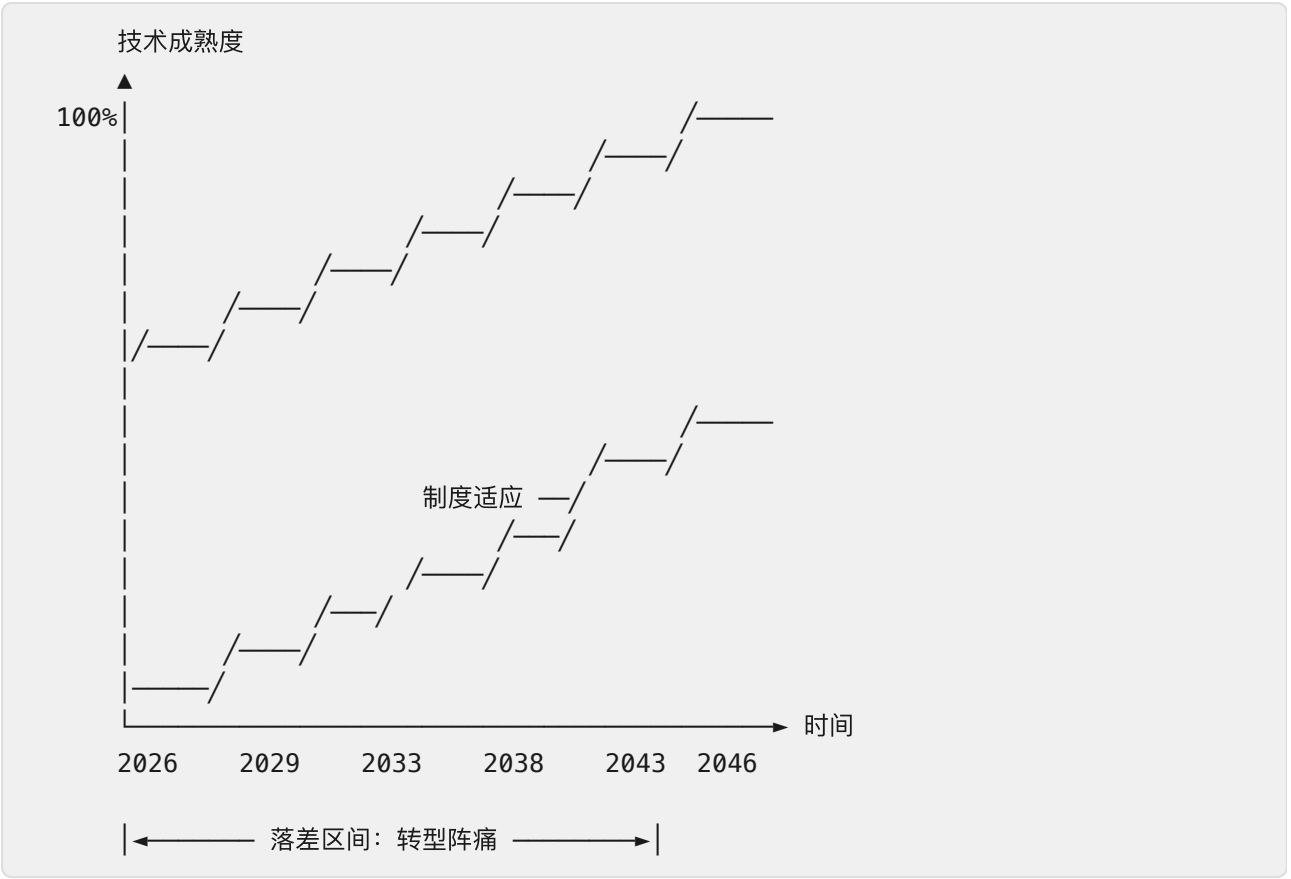
| 年份 | 变化 |
| --- | --- |
| 2040年代中期 | “工作”概念本身被重新定义：以劳动换收入从制度安排变为政治选择 |
| 2046 | 全球白领岗位（以2026年定义）消失六到七成 |
| 2046 | 基础保障金覆盖大部分人口；AI税成为财政重要来源 |
| 2046 | 教育体系正经历自普鲁士模式以来最深刻变革（尚未完成） |
| 2046 | 自我报告的”人生意义感”较2026年下降约一成 |

### 这个阶段的关键词：**暂时的平衡**

> 不是终点，而是一个新的起点。充满张力，持续演化。

---

## 两条轨道之间的落差




**核心洞察**：技术是指数曲线，制度是线性跟随。两者之间的落差就是每个具体的人承受的代价。缩小这个落差，是治理的核心任务。

# 附录三：2046世界结构图

> 基于全书推演，这张图描绘2046年世界的社会结构、权力分布与人类生活的新形态。不是预测，是一种有据可依的想象。

## 一、新五层阶梯

工业社会的阶层金字塔（上中下）已被替换为一个新的五层结构。层级之间的分界线不再是收入，而是**与AI的关系**。

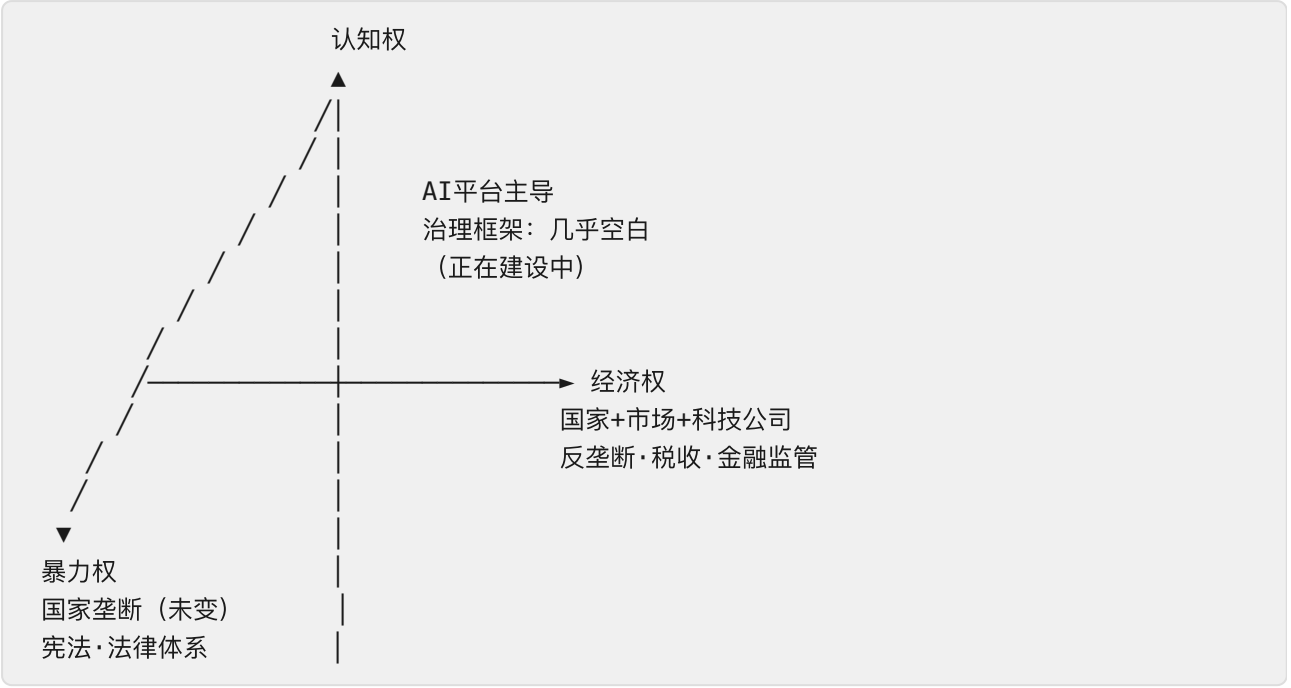
### 四种资本的权重迁移

| 资本类型 | 2026年权重 | 2046年权重 | 变化 |
| --- | --- | --- | --- |
| 经济资本 | 高 | 极端集中于顶层 | 基尼系数进一步扩大 |
| 文化资本 | 高（学历=通行证） | 大幅贬值 | AI让知识近乎免费 |
| 社会资本 | 中 | 大幅升值 | 信任成为最稀缺资源 |
| 符号资本 | 低 | 成为新战场 | “真人做的”成为溢价标签 |

---

## 二、治理权力的三维模型

传统国家治理建立在”暴力权+经济权”两个维度上。AI引入了第三个维度：**认知权**。



### 三种权力的控制者

| 权力维度 | 主要控制者 | 约束框架 | 2046年状态 |
| --- | --- | --- | --- |
| 暴力权 | 国家 | 宪法、法律 | 基本未变 |
| 经济权 | 国家+市场+科技公司 | 反垄断、税收、监管 | AI税成财政重要来源 |
| 认知权 | AI平台 | 几乎空白 | 最大的治理真空 |

---

## 三、新社会契约的四种配方

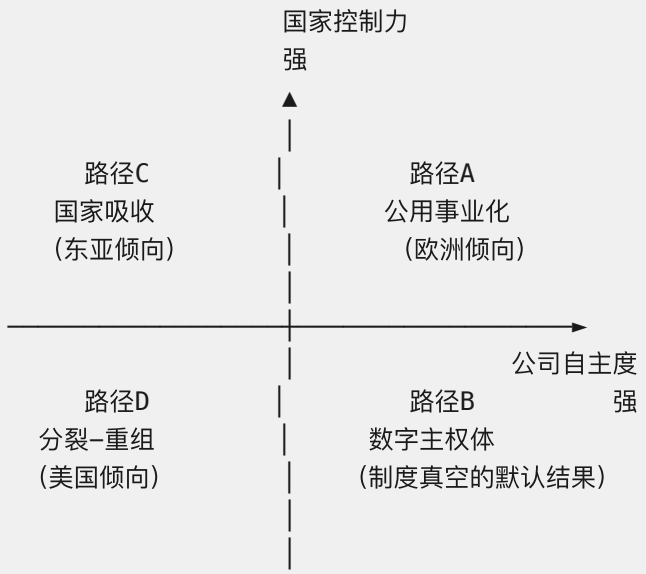
没有单一方案能解决AI时代的分配问题。各国调出不同的组合配方。

| 方案 | 解决什么 | 解决不了什么 | 倾向国家 |
| --- | --- | --- | --- |
| 无条件基本收入（UBI） | “多少钱” | “我是谁”——身份与意义 | 北欧、肯尼亚 |
| 数据红利与算力分红 | 数据贡献的回报 | 合成数据占比上升削弱逻辑 | 欧盟 |
| AI税与机器人税 | 财政再分配 | 温和税率(10–20%)才净正效益，超35%抑制创新 | 法国、韩国 |
| 公共服务扩展 | 近乎无限的AI服务 | 非零和，但不解决现金需求 | 新加坡、丹麦、中国 |

**2046年现实**：无单一方案胜出，多数国家采用混合模式。中国以”转型保障金+公共服务扩展”为主轴。

---

## 四、超级公司与国家的四条演化路径




| 路径 | 模型 | 特征 | 倾向地区 |
| --- | --- | --- | --- |
| A：公用事业化 | 电网模型 | 强监管、普遍服务义务 | 欧洲 |
| B：数字主权体 | 新东印度公司 | 公司拥有准国家权力 | 治理真空区 |
| C：国家吸收 | 深度融合 | 国家与AI公司边界模糊 | 东亚 |
| D：分裂–重组 | AT&T模型 | 反垄断拆分→重组→再拆分 | 美国 |

**2046年现实**：四条路径叠加的混合态。同一国家的不同领域走不同路径。

## 五、社会形态：从金字塔到群岛


**群岛社会的关键变量**：岛屿之间是否保持”桥接连接”（弱连接）。

- **积极案例**：北欧”数字公社”网络——岛屿间有制度性桥梁
- **消极案例**：封闭飞地——富人社区与贫困社区物理隔离，互不往来

---

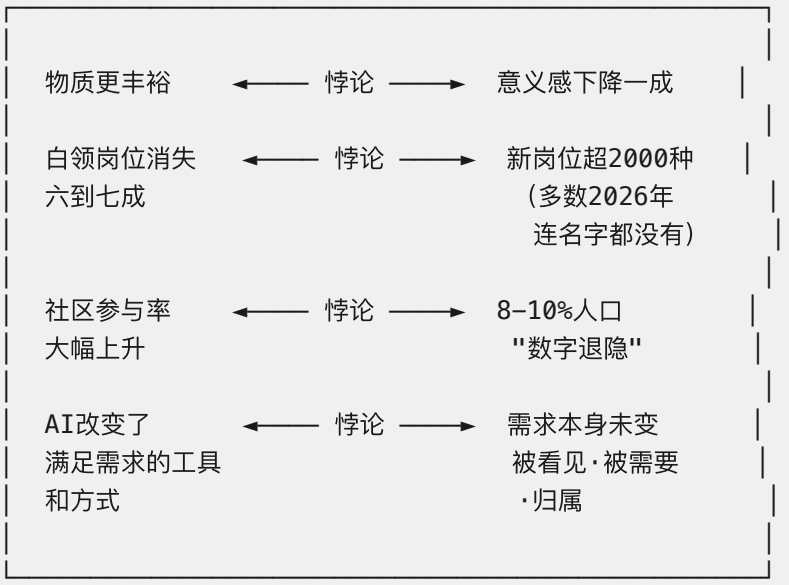
## 六、五种新的有意义活动

当”工作”不再等于”劳动换收入”，人类的有意义活动重新分化为五个方向：

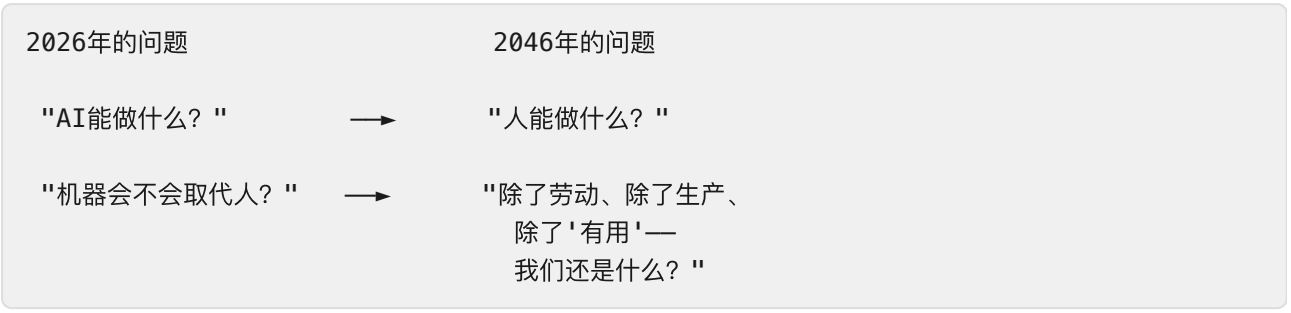
| 活动类型 | 核心内容 | 为什么AI做不了 |
| --- | --- | --- |
| **编排** | 将模糊意图转译为AI工作流 | 需要理解人类真正想要什么 |
| **关系建设** | 社区组织、信任中介、人际连接 | 信任没有API |
| **体验创造** | 策展、教练、线下”在场”体验 | 在场感无法数字化 |
| **意义生产** | 艺术、哲学、文化创造 | “背后有一个活过的人”是价值前提 |
| **公共行动** | 公民议事会、社区共治、社会契约谈判 | 合法性来自人类参与本身 |

## 七、2046年的核心悖论

一组无法调和、只能共存的张力：




## 终极问题的迁移




这不是一个已经回答的问题。这是一个刚刚被提出的问题。

2046年——一个开放的问号。

# 概念索引

本书核心概念速查表。括号内为概念首次出现的章节。

| 概念 | 定义 | 首次出现 |
| --- | --- | --- |
| 认知垄断 | 人类对高阶信息处理（推理、判断、创造）的独占地位，是工业文明全部制度设计的隐含前提 | 序章 |
| 相变/相变点 | 复杂系统从一种稳态跃迁到另一种稳态的临界转折，量变在此刻突然转为质变 | 第一章 |
| 制度拖曳 | 技术已经变了但制度还没跟上的时间差，变革越快拖曳越剧烈 | 序章 |
| 六根绳索 | 维系劳动契约的六个社会功能（收入、身份、社交、时间结构、意义感、社会保障），全部绑在”你有一份工作”这一个桩子上 | 第二章 |
| 反脆弱 | 塔勒布概念，指从波动和冲击中反而获益的系统特性，与”脆弱”和”稳健”构成三元结构 | 第二章 |
| 信号贬值 | 学历等传统筛选信号因AI普及而丧失区分能力，替代性信号（作品集、开源贡献）涌现 | 第三章 |
| 交易成本归零 | 科斯框架下公司存在的理由——内部协调成本低于外部交易成本——因AI大幅压低市场交易成本而被消解 | 第四章 |
| 四种资本 | 布迪厄的资本理论：经济资本（钱）、文化资本（知识与学历）、社会资本（关 | 第五章 |

| 概念 | 定义 | 首次出现 |
| --- | --- | --- |
| | 系网络）、符号资本（声誉与注意力），AI正在重排其权重 | |
| 认知基础设施 | 判断力从个人能力变为公共服务，如同电力和互联网一样成为文明运行的底层设施 | 第六章 |
| 五阶段消亡模型 | 工作被AI替代的五个递进阶段，从辅助增强到创造性工作的价值重估 | 第七章 |
| 内容泛滥 | AI导致创作门槛归零、内容供给从稀缺到过剩的相变，注意力市场从”洪水”变成”海洋” | 第八章 |
| 新圈地运动 | 科技公司对数据、算力、人才资源的垄断性占有，类比历史上的土地圈地运动 | 第九章 |
| 双重运动 | 波兰尼框架：市场力量每一次扩张都会激发社会的自我保护反应，AI时代正在重演这一模式 | 第十章 |
| 座架（Gestell） | 海德格尔概念，技术将一切——包括人自身——转化为可计算、可优化的”持存物”，AI时代尤为尖锐 | 第十一章 |
| 意义市场 | 传统意义来源（劳动、宗教、国族叙事）瓦解后，多元意义供给商竞争的格局 | 第十二章 |
| 三层意义需求 | 个体层（“我为什么活着”）、社区层（“我属于哪 | 第十二章 |

| 概念 | 定义 | 首次出现 |
| --- | --- | --- |
| | 里”）、超越层（“死后会怎样”），宗教是唯一能同时覆盖三层的供给商 | |
| 数字人格 | 数字化延续的个人法律主体，需注册认证、有效期限定、决策权限严格限定，家属有权注销 | 第十三章 |
| 尾部风险 | 低概率但后果极端的系统性风险，传统风险度量（如VaR）对其完全失明 | 第十四章 |
| 五层新阶层 | AI时代的社会分层结构：算力层、编排层、关系层、体验层、依赖层，由四种资本的权重迁移所决定 | 第十七章 |
| 治理层（Governance Layer） | 介于公司和国家之间的第三种权力形态，在AI认知权力的行使上建立问责机制，合法性基础尚脆弱 | 第十八章 |