

AI大压缩： 智能经济时代的生产 范式转移

The Great AI Compression:
A Production Paradigm Shift in the Age of
Intelligence

2026年5月

AI大压缩:智能经济时代的生产范式转移

The Great AI Compression: A Production Paradigm Shift in the Age of Intelligence

联合发布

香港大学人工智能、管理与组织研究中心
(HKU CAMO)

中国信息通信研究院华东分院

模驭人工智能科技(上海)有限公司

2026 年 5 月

执行摘要

纵观技术发展史,每一次重大跃迁的本质都是对价值创造链路中间环节的系统性压缩——能源、材料、信息领域概莫如此。AI 正在触发第四次大压缩,其对象从物理世界升维至认知劳动本身:“人理解意图→操作工具→输出结果”的冗长链条,正被压缩为“意图→AI 交付”。因此,今天的 AI 远不止于强大的效率工具,而是涌现全新价值生产范式的源动力。

AI 大压缩在时间、空间、价值链、组织四个维度同步展开。AI 可近乎独立地完成大量认知任务,交付周期坍缩 1-4 个数量级;专业能力与专家个体、地理位置实现解耦;承担信息转译与流程衔接的价值创造中间层正在被 AI 节点化;产能与人员规模结构性脱钩。四重压缩沿因果链传导,使得价值创造的最小可行单元突变为“少量决策者 + AI 系统”,这正是新范式的起点。

在此背景下,AI 驱动的新生产范式正在成型,其设计原则的第一性是“AI 价值最大化”。人机边界持续上移,人类职能逐步收敛于定义目标、处理例外、以及承担责任。组织资产从个人经验转向具备复利特征的系统资产;成本从按岗位预置转向随业务负载连续伸缩;定价从按投入计费迁移至按结果计费。概率生成与确定性交付的张力、系统自治与责任追溯的矛盾,成为了新范式的内生约束。

范式迁移不会均匀席卷所有行业。扩散速度取决于落地阻力与变现动力的交互。软件、营销、客服、基础财税等已处于快速渗透区;医疗、法律、金融、先进制造以切片式路径渗透,而壁垒一旦突破即转化为护城河,构成最显著的价值高地。相应价值链上的利润池向基础设施层(算力与基础模型)与意图发起层(场景与数据)持续集中。

压缩不会因任何组织拒绝采纳而暂停,未被压缩的中间层只是尚未轮到;新范式已在先行行业得到验证,过度的审慎将面对快速拉大的系统资产代差;处于传统价值链中间位置的价值创造主体,转型已成为生死存亡的问题。

压缩消除的固然多,但被释放的更多——AI 大压缩消除认知中间环节的同时,触发了远比旧世界更广阔的新价值大爆发。观望正在成为最昂贵的战略选择。

Executive Summary

Every major leap in technological history has, at its core, been a systematic compression of the intermediate links of value creation—first energy, then materials, then information. AI is now triggering the fourth compression. For the first time, the target shifts from the physical world to cognitive labor itself: the lengthy pipeline of "human interprets intent → operates tools → produces output" is collapsing into "intent → AI delivers." AI today is not merely a more powerful efficiency tool; it is the catalyst behind an emerging paradigm of value production.

The AI Compression unfolds simultaneously across four dimensions: time, space, value chain, and organization. For cognitive tasks AI can perform with little or no human intervention, delivery cycles collapse by one to four orders of magnitude. Professional capability decouples from individual experts and physical geography. The middle layers responsible for information translation and process coordination are being folded into AI nodes. Output capacity decouples structurally from headcount. As these effects propagate, the minimum viable unit of complex value creation shifts to "a small group of decision-makers paired with AI systems"—the starting point of the new paradigm.

Against this backdrop, an AI-driven production paradigm is taking shape, designed around a single first principle: maximize the value AI can create. The human-machine boundary moves steadily upward, and human roles converge on three functions—defining objectives, handling exceptions, and bearing accountability. Organizational assets shift from individual expertise to compounding system assets; cost structures shift from headcount-based provisioning to continuous scaling with workload; pricing migrates from billable inputs to outcome-based delivery. The tension between probabilistic generation and deterministic delivery, and the paradox between system autonomy and accountability tracing, become the paradigm's intrinsic constraints.

The shift will not sweep through all industries uniformly. Its pace is governed by the interplay of adoption barriers and commercial drivers. Software, marketing, customer service, and basic finance already sit in the rapid-penetration zone. Healthcare, legal services, financial services, and advanced manufacturing diffuse along a slicing path; once the barriers are broken, they convert into moats and define the most significant value reservoirs. Profit pools concentrate persistently at both ends of the value chain—the infrastructure layer (compute and foundation models) and the intent-origination layer (scenarios and data).

Compression will not pause because any organization declines to participate; the middle layers not yet compressed are simply waiting their turn. The new paradigm has already been validated by early movers; excessive caution will face a rapidly widening gap in system assets. For value-creating entities positioned in the traditional middle of the value chain, transformation has become a question of survival.

What compression removes is much, but what it releases is more—as AI compresses the intermediate links of cognition, it triggers an explosion of new value far broader than the world it replaces. Waiting is becoming the most expensive strategic choice.

目录

前言	01
历史规律与AI大压缩的本质	02
四维压缩:从物理效应到商业结构	06
新生产范式:围绕AI能力最大化的重构	16
行业扩散:新范式的不均匀渗透	27
大爆发:被AI大压缩释放的新价值与创新机会	35
结语:压缩即创造	39
附录	42

CONTENTS

Preface	01
The Essence of AI Compression	02
Four Dimensions of Compression	06
New Production Paradigm	16
The Industrial Diffusion of New Paradigm	27
The Great Explosion: New Values and Opportunities	35
Conclusion: Compression Is Creation	39
Appendix	42

前言 Preface

纵观人类技术发展史，每一次伟大的跃迁，本质都是对原有价值创造链条上中间环节的系统性压缩——能源、材料、信息领域概莫如此。今天，我们正站在第四次科技大跃迁的临界点，其本质正是AI对认知链条中间环节的系统性压缩。

将AI仅仅视作“更强大的效率工具”是对本轮技术跃迁的巨大低估，因为AI引发的并非渐进式的工具升级，而是技术工具首次对“认知劳动链路”的降维折叠。当“人理解意图 → 操作工具 → 输出结果”的冗长链条被瞬间压缩为“意图 → AI端到端直接交付”时，依附于旧链路之上的价值构成、组织形态、商业生态，必将发生彻底的重构，即“AI大压缩”。

目标读者与阅读建议：

本文提出“我们正处于何种历史时刻”的趋势性判断。我们相信，理解这一判断的人，将能够更清醒地做出关于组织、战略与投入的决策。目标受众是企业决策者、组织管理者、以及一切正在思考“AI 究竟对我的行业意味着什么”的人。本文不是 AI 技术科普，不是自动化实施手册，更不是裁员方案。它是一份关于“正在发生什么、为什么发生、以及需要做什么准备”的客观判断与理性分析。

01

历史规律 与 AI 大压缩的本质

1.1 技术史的Déjà Vu¹:从物理坍塌到认知折叠

技术发展史揭示的一般规律是：原本冗长的价值创造中间环节，不断被压缩进新的技术节点，价值创造链路变得越来越简洁优雅。这种“压缩”规律在能源、材料、信息与智能领域依次展开，呈现出高度一致的结构特征。前三次压缩，消除的都是人类认知活动的外部约束——获取能量、塑造材料、传递信息的链路被逐次缩短：

能源的压缩：从钻木取火、薪柴燃烧，到煤炭开采、蒸汽机转化，再到电力系统的建立与电机的即时调用，人类获取和使用能源的链路不断缩短、损耗不断降低。技术进步持续消除了能量获取、传输与转换过程中的物理摩擦。

材料的压缩：现代化工与增材制造（3D打印）消除了克服材料硬度的物理摩擦，将“采矿→粗炼→锻造→切削”的减材制造链条，压缩为分子级合成或数据模型直接成型，大幅压缩了加工损耗与中间处理层。



图 1-1 信息生成、复制、分发的中间过程逐步被压缩至 Agent

图1-1中的倒三角结构刻画了信息生成、复制、分发、传播链路的进化和压缩过程：传统印刷范式（多环节、强分工）、机械化与早期数字化阶段（部分流程整合）、以及数字化与智能体驱动阶段（端到端传播）。随着技术进步，信息传播链路中的中间环节数量与复杂度持续被压缩进新的技术工具中，最终由Agent完成信息的生成、复制、分发、消费的全流程。

1 Déjà Vu来自法语，原意为“已经见过，或似曾相识”。

信息的压缩:互联网消除了信息载体的物理质量,将“创作→排版→印刷→分发”的重资产链条,压缩为“创作→数字传播”,压缩了信息复制、运输与分发的几乎全部中间环节。

智能的压缩:正在发生的第四次大压缩是迄今最深刻的一次,压缩对象由认知活动的外部约束条件转向了认知活动本身。

1.2 AI大压缩的本质:认知价值链条的节点化

“AI大压缩”这一概念由香港大学李晋教授于2025年首次提出²,用以描述AI对认知劳动价值链条的系统性压缩效应。传统知识工作本质上是一条漫长的数据转译链。在旧有的价值创造体系中,其标准结构为:“人理解业务意图 → 人将意图转译为系统指令(操作工具) → 工具输出中间结果 → 人进行校验与拼装”。在这个链条中,人类实际上扮演了极其昂贵且低效的“API(应用程序接口)”,负责在不同的系统与流程之间搬运和加工信息。

AI Agent(智能体)的出现,直接将这一中间层整体抹除。价值创造的链路被粗暴且优雅地压缩为:“发起意图 → AI端到端直接交付”。被压缩掉的并不是人类的创造力,而是那些不产生核心业务洞察、仅负责信息转译与标准加工的认知冗余节点。

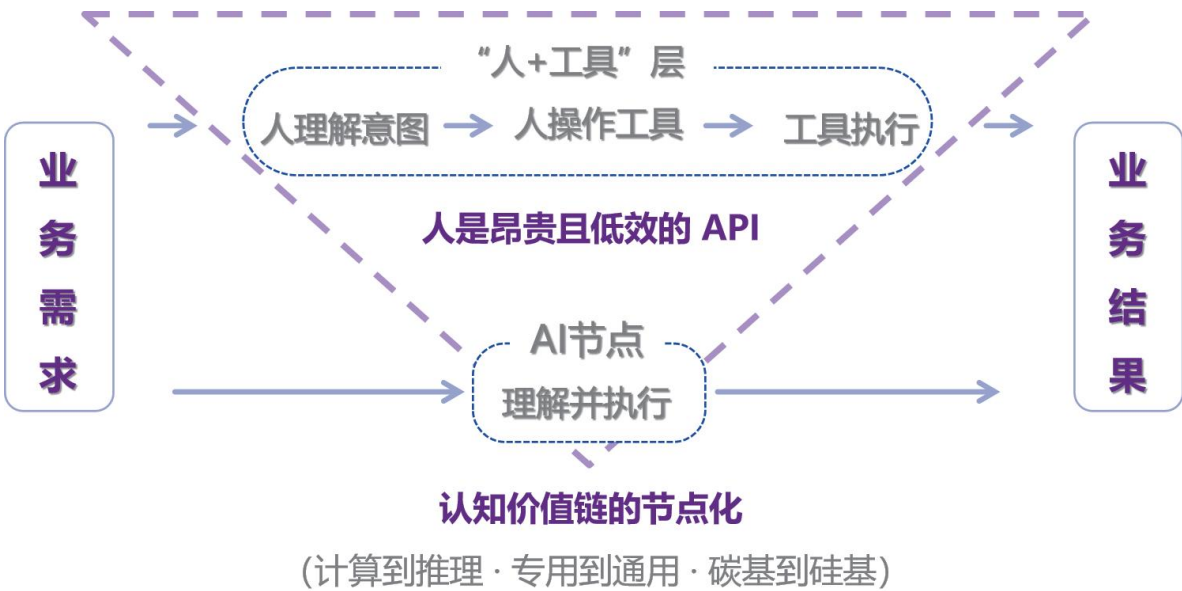


图 1-2 认知价值链条压缩为 AI 节点

传统“人+工具”范式下,业务需求经由“人理解意图→人操作工具→工具执行”的串联链路完成,人类充当连接需求与系统的“通用API”,成本高、效率低且不可规模化。AI节点同时具备理解与执行能力,直接连接业务需求与业务结果,将中间人工作业链路整体压缩。其本质是认知价值链从以人为中心向以AI节点为中心的迁移——认知能力被模块化、节点化并沉淀为可调用的智能基础设施。

2 “AI大压缩”概念由香港大学人工智能与组织管理中心主任李晋教授于2025年首次提出,参见:罗汉堂 | 李晋:AI 时代企业组织的“大压缩效应”,微信公众号“罗汉堂观点”,2025年7月22日。 <https://mp.weixin.qq.com/s/l-UU1Jg4fDwFpdccIC612A>

与历次技术革命相比, AI对认知劳动的压缩在深度、广度与持续性上, 均展现出前所未有的强度:

深度的突破——从计算到推理: 过去被认为只有人类大脑才能处理的非结构化推理、跨模态解析与模糊意图对齐, 首次被工程化、可计算化。系统不再需要人类将任务拆解为精确的代码, 而是直接承接自然语言表达的模糊意图。

广度的拓展——从专用到通用: 无论是编写代码、审查合同、生成营销内容, 还是辅助医疗初筛, 其底层都在被统一为同一种模型能力的不同调用方式。正因如此, AI的压缩效应不再局限于某一行业或某一岗位, 而是可以跨越专业边界, 渗透几乎所有高信息密度的知识工作场景。

持久性的跃迁——从碳基上限到硅基扩展: 认知能力的供给, 第一次在相当程度上摆脱了人类脑容量、训练周期与经验积累的生物学约束。随着模型、算力与数据基础设施的持续迭代, 这种压缩并非一次性的效率跃升, 而是一个会不断深化、并持续降低边际成本的动态过程。

02

四维压缩： 从物理效应到商业结构

Overview

压缩如何在商业世界中具象化？本章将其结构化四个维度：时间、空间、价值链、组织——前两者是AI的直接物理效应，后两者是压缩效应沿因果链向商业结构的传导。

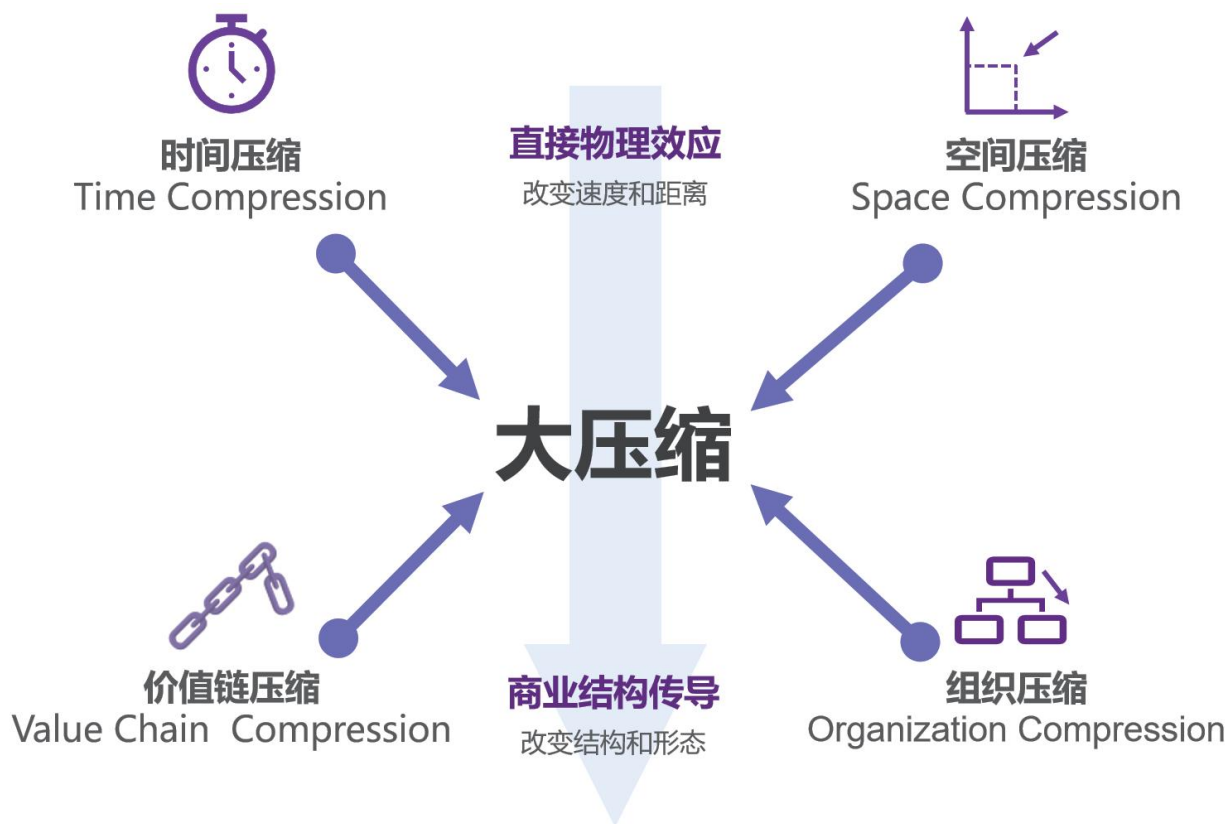


图 2-1 四维压缩：时间、空间、价值链、组织

AI大压缩在商业世界中的四个作用维度。时间压缩——交付周期坍缩1-4个数量级；空间压缩——专业能力与专家个体及地理位置解耦；价值链压缩——信息转译与流程衔接的中间层整体塌陷；组织压缩——产能与人员规模结构性脱钩。前两者是AI的直接物理效应，后两者是效应沿因果链向商业结构的传导结果。

2.1 时间压缩:交付周期的数量级坍缩

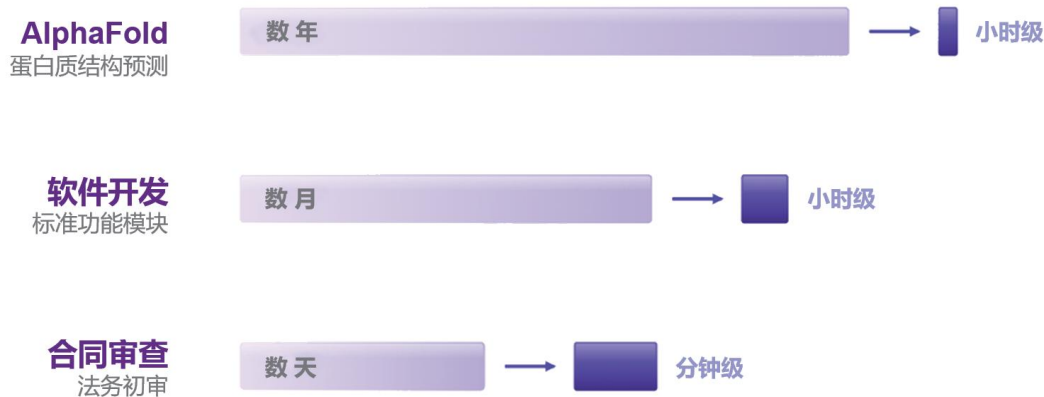
在AI能够独立或近乎独立完成的认知任务中，交付周期的压缩幅度可达2-4个数量级——AlphaFold将蛋白质结构预测从数年压缩至小时级³，AI辅助编程将标准功能模块的开发周期从数月压缩至小时级⁴，AI合同审查将法务初审从数天缩短至分钟级。这种坍缩的直接商业后果，远不止“做得更快”：

3 Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature, 596, 583–589 (2021).

4 Peng, S. et al. The Impact of AI on Developer Productivity: Evidence from GitHub Copilot. arXiv:2302.06590 (2023).

第一,交付基准线被重置。当AI将合同初审从两周压缩到十分钟,“两周交付”从正常节奏变为不可接受的低效。任何一个环节率先达到AI交付速度,都会倒逼上下游同步提速,否则就成为整条价值链的瓶颈。行业竞争的入场门槛,正在从“能不能做”陡然变成“能否在新基准时间内做完”。

第二,任务周期的决定因素发生质变。旧流程中大量看似正常的环节——排期等待、交接间隙、审批流转——本质上是人类生理极限与组织协调摩擦的副产物。当执行主体切换为可并发、可连续运行的AI系统,任务周期便从“按组织节奏交付”转向“按系统吞吐交付”。产能的约束变量从“人手数”变为“算力供给”。



- **交付基准线被重置**——“两周交付”从正常变为不可接受
- **产能约束变量质变**——从「人头数」变为「算力供给」

图 2-2 时间压缩：任务周期基准线重置，产能约束从人头变算力

AI已经在蛋白质晶体结构预测、软件开发、文本写作与审核等众多领域实现了工作时间2-4个数量级的压缩

2.2 空间压缩:专业能力的去地理化

前互联网时代:能力被专家与物理空间锁死。专业判断的输出受限于专家个体的物理在场——患者必须亲赴三甲医院,企业必须到一线城市律所面谈,偏远学生无法获得优质师资。供给公式是“专家人数 × 单人时间 × 空间距离”——专家不可分身,能力不可远程投送。

互联网时代:信息可达,但能力仍受限专家时间。远程问诊、在线教育、法律咨询App在相当程度上消解了物理空间的阻隔,供给公式进化为“专家人数 × 单人时间”。但核心瓶颈并未消除:每一次专业判断仍需要真人专家实时投入,一位医生同时只能面对一个视频窗口。互联网拓宽了投送半径,却没有改变产能的生物学上限。

AI大压缩时代:能力算法化,供给理论上可无限并发。大规模训练与微调将影像医生的模式识别、律师的条款敏感度等隐性知识算法化,转化为可并发调用的计算服务。供给公式跃迁为“模型能力 × 算力网络”,实现了信息与能力均可达。

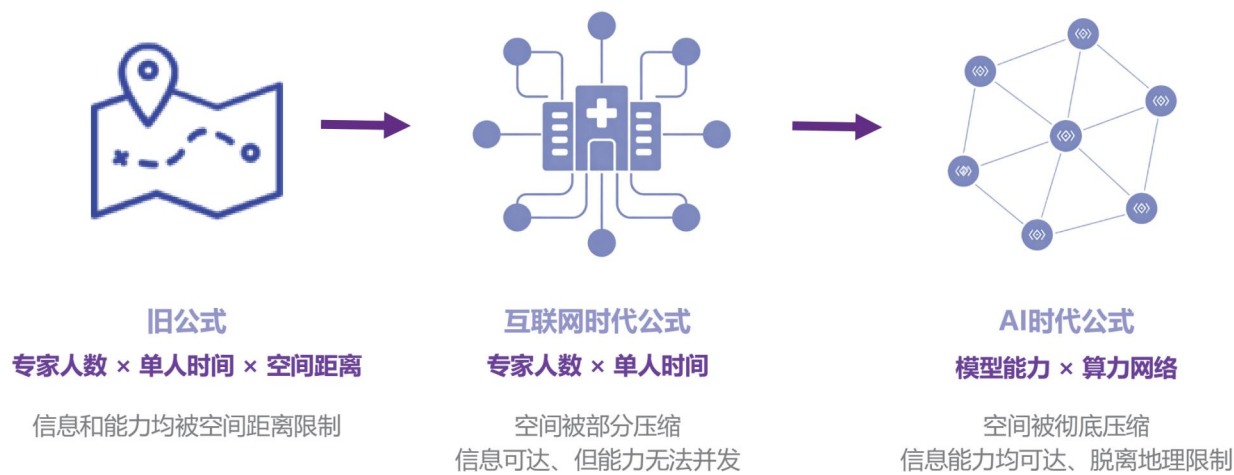


图 2-3 空间压缩：专业能力与专家、空间解耦

由此产生的商业形变是:专业能力的获取条件,从“是否身处正确的地理位置”转向“是否接入了AI服务”。在医疗影像初筛、长尾教育答疑、中小企业法务支持等场景中,外围节点首次拥有了相对直接的生产手段。“地理聚集”这一持续了整个工业时代的竞争护城河,正在被AI显著地削弱。

案例2-1：基层医院 · 时间压缩 + 空间压缩

3分钟诊断，零边际成本扩散

一家西北二甲医院的能力跃迁——基层开始常规开展三甲级复杂筛查

5-7_x

诊断时间压缩

15-20min → 3min

42%

患者等待时间缩短

全科室体验跃升

≈0

能力扩散边际成本

三甲经验 → 基层可得

旧范式 · 经验驱动逐层阅片

- 01 医生逐层调阅影像序列
- 02 肉眼寻找病灶 + 手动测量
- 03 撰写文字诊断报告
- 04 复杂病例转诊三甲等待复核

→
AI
压缩

新范式 · AI初筛 + 人复核

AI秒级全序列初筛 → 医生复核确认

深度学习多模态影像分析，数秒完成初筛并自动标记可疑病灶、量化指标、生成结构化报告。医生从“逐层找病灶”转为“复核+确认诊断”。

— 背景

西北地区日均300+例的二甲综合医院。放射科双重困境：常规阅片负荷饱和，复杂病例排至下一班次；更根本的是能力天花板——早期乳腺癌微小钙化灶、肋骨隐匿性骨折等高难检出项，基层医生“看不到影像，看不准病灶”，患者被迫转诊三甲。2024年引入AI辅助影像诊断系统，覆盖肺结节、骨折、乳腺癌等高频场景。

— 发现

最重要的变化不是“让专家更快”，而是“让基层具备专家级能力”。AI几秒内标记出人眼易遗漏的微小病灶，过去在基层漏检率较高的病例，初筛能力直接内嵌于基层流程。当三甲数十年阅片经验被压缩为可部署的模型，优质诊断以趋近于零的边际成本向基层扩散——一次典型的能力平权。

2.3 价值链压缩:中间层的整体塌陷

传统价值链之所以冗长，根本原因在于：复杂任务必须被拆解为“人类可处理、可交接”的小步骤，而每一次交接都伴随信息损耗与协调成本。传统软件开发的典型流程分工“产品经理→架构师→前端→后端→测试→运维”的六层分工，并非最优解，而是人类认知带宽有限这一约束条件下的妥协方案。

AI之所以能将中间层整体压缩，正是因为它同时突破了这一约束的三个子条件：多模态能力打破了专业壁垒（文本、图像、代码不再需要不同工种接力）；Agent 工具调用打通了环节衔接（“查信息—处理—执行—触发下一步”被串联到同一回路）；长上下文维持了跨步骤连贯性（传统组织用会议和文档维持的一致性，被记忆机制部分替代）。



图 2-4 价值链压缩：AI 整合全流程、消除交接损耗，让传统分工中间层整体压缩

由此产生的对传统价值链的不可逆压缩已在真实场景中显现。以影视创作为例：传统流程需要“创意构思→剧本创作/分镜设计→选角·布景·服·化·道/拍摄→素材剪辑→后期特效→成片输出”，AI正将其收敛为“人定义创意需求—AI交付视频结果”的闭环。类似的压缩同样发生在软件开发、内容生产、法律服务等诸多领域。被压缩的，正是那些因人类处理能力有限而不得不切分出的垂直细分操作或标准化重复步骤——AI正在将这种妥协变为历史。

案例2-2: AI原生CRM开发 · 价值链压缩

6个角色, 1个人干完了

软件开发的水平分层如何被AI垂直整合——中间层整体塌陷的极限验证



— 背景

一套基于AI Agent框架的**客户管理与销售自动化系统**，需处理销售管线、Agent自主编排等复杂业务逻辑。传统模式下6个角色各守职能边界，通过PRD和会议逐层接力。真正拖慢项目的不是编码，而是**角色之间的翻译成本**——需求经6人传递到部署，信息在每次交接中衰减。50人月估算中，多少是价值创造，多少是在为“人与人之间的翻译”付费？

— 发现

33倍压缩最大的效率来源不是“AI写代码更快”，而是“人与人之间的**翻译成本被消除**”。交付链从6节点收缩到1节点，信息衰减和交接等待同时归零——这才是压缩比远超预期的真正原因。与传统国企8倍压缩案例对照：**差距不在人的能力，而在组织起点。历史包袱越重，压缩比越接近下界。**

2.4 组织压缩: 产能与人员规模和解耦

价值链的塌陷进一步传导至组织结构:当多层分工被压缩为端到端节点(详见2.3节), 支撑这些分工层级的组织架构就失去了存在的必要性。传统组织扩张的底层假设是“产能与人数线性相关”——业务翻倍, 团队翻倍。AI正在打破这一假设:产能的扩缩容方式从调整编制转向调整系统弹性, 企业按任务量动态扩展算力与API调用, 而非提前招人形成冗余。

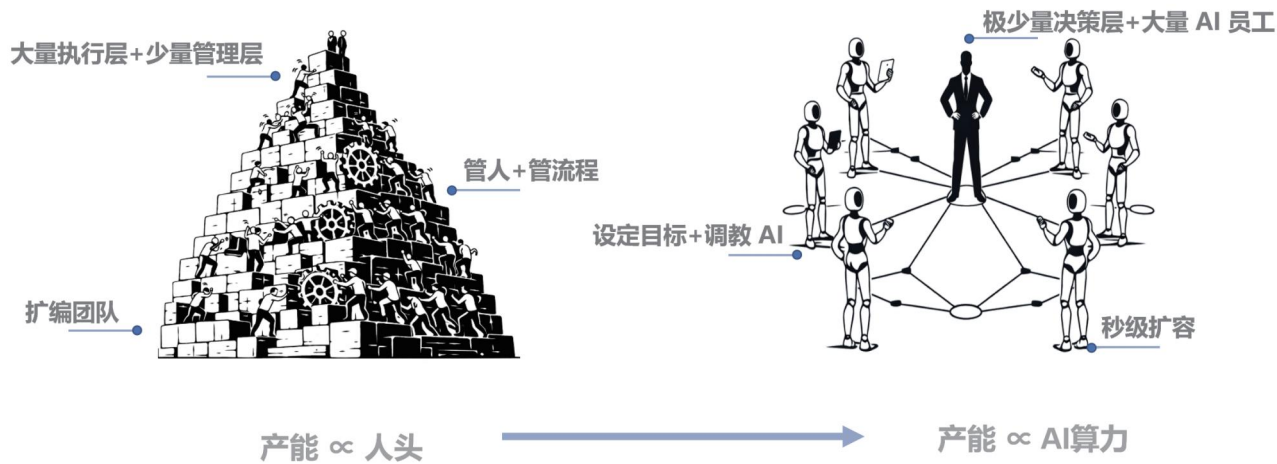


图 2-5 组织压缩：产能与人员规模和解耦

公开案例提供了早期证据：Salesforce 在内部部署 Agentforce 后，将客户支持团队从约 9,000 人调整至约 5,000 人，与此同时 AI 已承担约 50% 的客户对话，且满意度评分与人工组基本持平⁵。在 AI 原生创业领域，Anysphere（旗下 Cursor）以约数十至数百人的团队规模实现 20 亿美元 ARR，成为 SaaS 历史上最快从 0 增长至 10 亿美元 ARR 的公司⁶。两个案例分别从“被压缩的传统组织”与“天生轻量化的 AI 原生组织”，共同揭示了组织竞争力的核心变量，正在从拥有多少人转变为能调动多强的系统。

这一现象与企业理论的经典直觉构成有趣的张力。科斯（Ronald Coase）的企业边界理论指出，企业之所以存在，是因为内部协调成本低于市场交易成本；当内部协调成本进一步降低时，企业应当扩张边界。AI 确实大幅降低了内部协调成本——但它对市场交易成本的压缩更为剧烈。当 AI Agent 可以实时完成外部协作中的需求匹配、质量校验与交付闭环时，搜寻、签约、监督等传统市场摩擦同步压缩，外部协作的成本优势反超内部雇佣。科斯等式的两侧同时被改写，但市场侧的降幅更大——企业的最优边界因此不是扩张，而是向内收敛至不可外包的核心决策节点。产能与人员规模的脱钩，正是这一理论在 AI 生产范式中的映射。

5 Benioff, M. 在 The Logan Bartlett Show 播客访谈中表述: "I was able to rebalance my headcount on my support — from 9,000 heads to about 5,000, because I need less heads." 同期 Salesforce 官方声明: Agentforce 已承担约 50% 客户对话，人工组与 AI 组的客户满意度 (CSAT) 评分基本持平。来源: CNBC 报道 ("Salesforce CEO confirms 4,000 layoffs 'because I need less heads' with AI"), 2025 年 9 月 2 日; Salesforce 官方声明, 2025 年 9 月。

6 Anysphere 公司公告及多方报道, 2025–2026。Cursor 于 2025 年 1 月达到 1 亿美元 ARR, 2025 年 6 月达到 5 亿美元, 2025 年 11 月跨越 10 亿美元, 2026 年初达到约 20 亿美元 ARR。同期员工规模从约 50 人扩展至约 300 人, 人均年化营收约 600–700 万美元。来源: TechCrunch、Bloomberg、Anysphere 公司公告。具体数字以官方最新披露为准。

需注意的是，产能与人数的脱钩并非没有边界。当人机分工的合理界限被越过，代价将由业务质量与客户体验支付——这条边界如何划定，是第三章的核心命题。

2.5 四维压缩的叠加:两个核心涌现

时间、空间、价值链、组织结构——四重压缩沿因果链条层层传导、交叉叠加，涌现出两个迥异于旧范式的突变：

第一，能力的去在场化。当专业任务的完成时间被骤降至分秒级（时间压缩），且能力可以云端分发或低成本私有化部署（空间压缩），二者叠加的结果是：专业能力首次大规模脱离了物理载体而存在，实现了能力供给与地理位置绑定关系的彻底解耦。

第二，最小生产单元的重定义。当中间层塌陷（价值链压缩）且产能与人头脱钩（组织压缩），复杂价值创造的最小容器不再必须是企业主体。“少量关键决策者+ 一组AI系统”正在成为新的最小可行生产单元，能够承载过去需要完整部门体系才能实现的业务闭环。

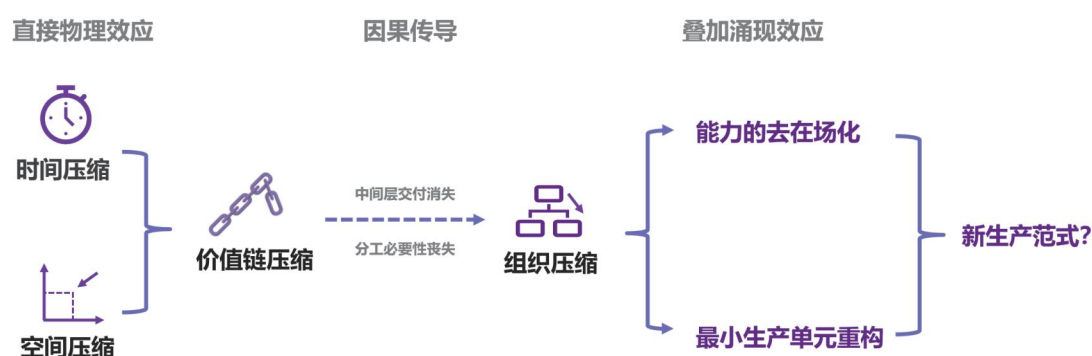


图 2-6 四维压缩的叠加：两个核心涌现，催生新生产范式

两个涌现效应指向同一结论：旧生产关系的核心假设已逐一失效，围绕“AI能力最大化”设计的新范式正在成型。

案例2-3：美国经典跑车OPC · 四维压缩叠加

1个人，4条产品线，0个新增岗位

时间、空间、价值链、组织——四重压缩如何同时发生在一个人身上

1→4

产品线扩展
散热风扇 → 四品类并行

1人

组织规模始终不变
零新增岗位

0

采购/跟单/翻译
后台被基础设施吸收

● 时间：沟通周期从数周缩至实时

● 空间：跨国制造远程可验证

● 价值链：寻源/翻译/跟单整体折叠

● 组织：1人调度多供应商多品类

— 背景

美国一家聚焦经典跑车的利基改装件企业，创始人一人兼任产品工程师、销售和品牌经营，技术文章在车友论坛被视为权威。限制他的不是技术想法，而是后台重量——跨国供应链的语言、验厂、图纸修改、物流追踪均需人工介入督办。传统模式下扩展产品线，几乎等于先复制出一套组织后台。到2021年，越来越多时间花在维护官网和追踪琐事上，而不是做最擅长的事。

— 发现

四重压缩后最值得注意的变化：创始人的注意力终于回到了最值钱的地方——技术判断和产品开发。一个侧面佐证：官网几乎停止更新，但论坛活跃度明显上升，技术文章被更多人引用。认知资源被重新分配了。这正是2.5节“两个核心涌现”的微观注脚——能力去在场化与最小生产单元重定义，同时发生在一个人身上。

03

新生产范式： 围绕 AI 能力最大化的重构

Overview

当生产系统开始围绕AI能力而非人力流程进行设计时，一种新的生产范式便开始形成。本章将围绕这一重构过程，首先确立两条根本原则——以AI能力最大化为第一性的静态原则，与人机边界随AI能力提升而持续上移的动态原则；随后分析两条原则共同驱动的三个衍生结果——资产形态的转变、经济结构的重塑、评价体系的建立；最后讨论新范式的内生约束。

3.1 AI生产范式的两条设计原则

静态原则:AI能力最大化为第一性原理

新范式的起点,是设计原则的变化。工业时代以来,组织设计的默认前提是“人是唯一通用的执行主体”——先定义岗位,再划分部门,最后用流程串联。当 AI 在速度、并发、跨域综合上展现出跨量级优势时,这一默认前提不再成立。生产过程的设计逻辑,必须从“围绕人如何分工”转向“围绕 AI 在哪些边界内可以最大化价值创造”。



图 3-1 静态原则：AI 能力最大化为第一性原理

传统范式中,节点之间存在交接损耗。过渡时代将部分节点替换成AI,节点效率提高,但依然延续传统流程,因而组织整体效能提升很有限。新范式以 AI 价值最大化为核心, 零基设计重构人机分工边界。人类职能最终收敛到定义目标、处理例外、承担责任三项核心任务。

因此, AI 时代的正确起点不是“给现有员工配备 AI 工具以提升效率”,而是零基设计 (Zero-Based Design): 对每一项业务任务进行反向推演——哪些环节应由 AI 系统端到端完成,哪些环节必须保留给人类。前者决定了流程的压缩极限,后者定义了组织不可替代的核心壁垒;切分本身,即是当下的人机边界。

本文编写团队的实证研究表明,当软件开发组织从传统的水平分层(职能分工)转向 AI 驱动的垂直整合(端到端交付)时,资源消耗可压缩 8 至 33 倍,核心效率来源于人与人之间的协调开销被消除⁷。a16z 的分析同样印证了这一判断: AI 让个体效率提升了 10 倍,但没有一家企业因此变得 10 倍有价值,因为高效个体拼不出高效组织⁸; 1890 年代美国工厂用电动机替换蒸汽机,此后三十年生产率几乎未动;直到围绕电力重新设计生产线,增速才翻了四倍——触发翻转的不是更好的电动机,是重新设计工厂。

案例3-1: 储能运营 · 零基设计 · AI新生产范式



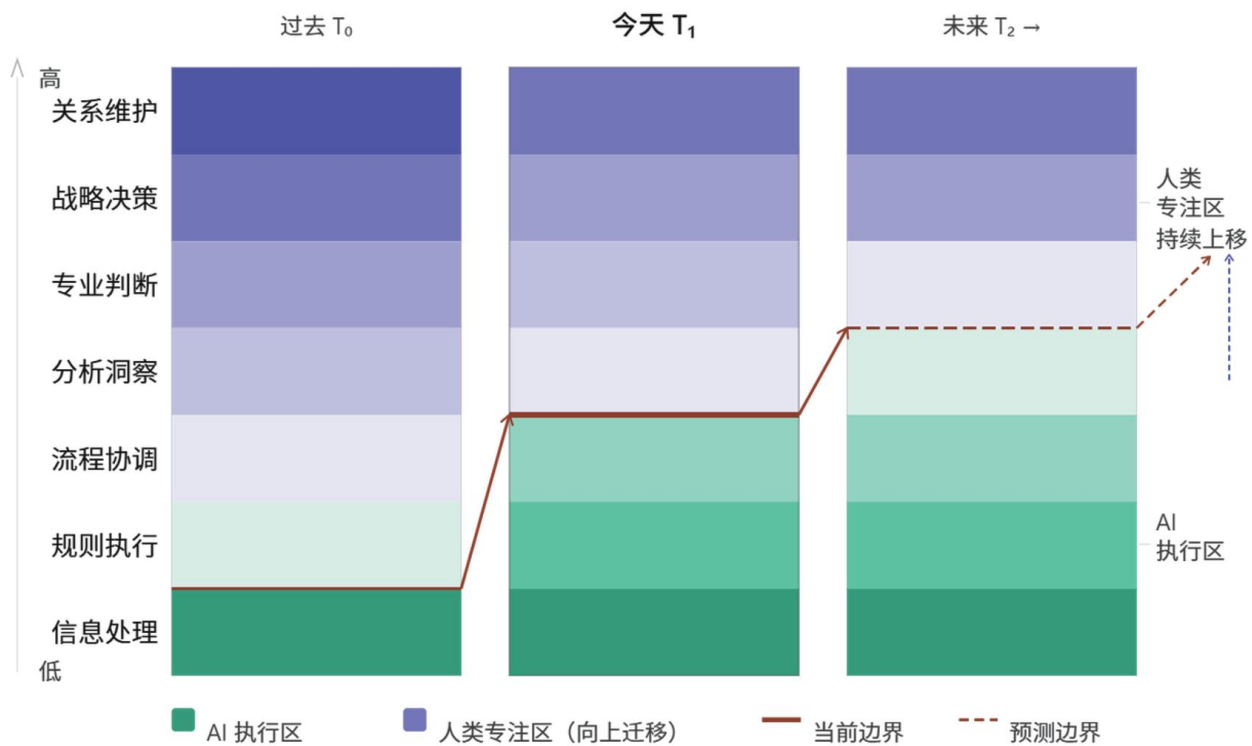


图 3-2 动态原则：人机协作边界动态上移

颜色即状态：同一工作层在三列中的颜色变化直接显示其归属演变，紫色越深代表人类越稳固掌握的高价值领域，青绿越深代表AI越彻底接管的基础执行层。

箭头指方向：对角连接箭头显示边界在各截面间的跃升幅度。右侧虚线延伸箭头与“持续上移”标注，表达边界尚未停止的核心命题。

这一动态上移的终局指向何处？当标准化执行、流程协调、乃至部分分析洞察被逐层抽离后，人类的不可替代性将收敛于三项核心职能：

定义目标。洞察真实世界的模糊需求，将其转化为AI可执行的清晰目标函数。机器可以优化任何目标函数，但无法判断哪个目标值得优化。

处理例外。利用常识与跨领域联想，化解系统在长尾场景中的逻辑死锁。模型在分布内表现优异，但现实世界的边界条件永远超出训练分布。

承担责任。完成最终的风险兜底、信任背书与利益分配。AI无法签署合同、承担法律后果或为破产负责——责任的不可让渡性，是人类在生产关系中最后的锚点。

人机边界的上移，直接重构了管理职能的内核。在旧组织中，管理的核心是分工与督办——安排谁来做、检查做得如何。在新范式下，管理者越来越少地安排谁来做，越来越多地定义系统应当在什么边界内优化。管理能力的核心，从驱动人的执行力，转变为：定义目标函数、设计系统护栏、配置人机断点、以及持续运营AI效能。

案例3-2：移民留学· 人机边界动态上移

500人销售团队的人机重构

前4步AI逐步独立承接，后2步始终由人主导——边界随数据沉淀持续上移

6→4月

新人独立上手周期
AI副驾加速培养

+30%

关单率预期提升
Phase 1+2 叠加

70%

销售时间回归关键决策
AI承接前置环节

背景

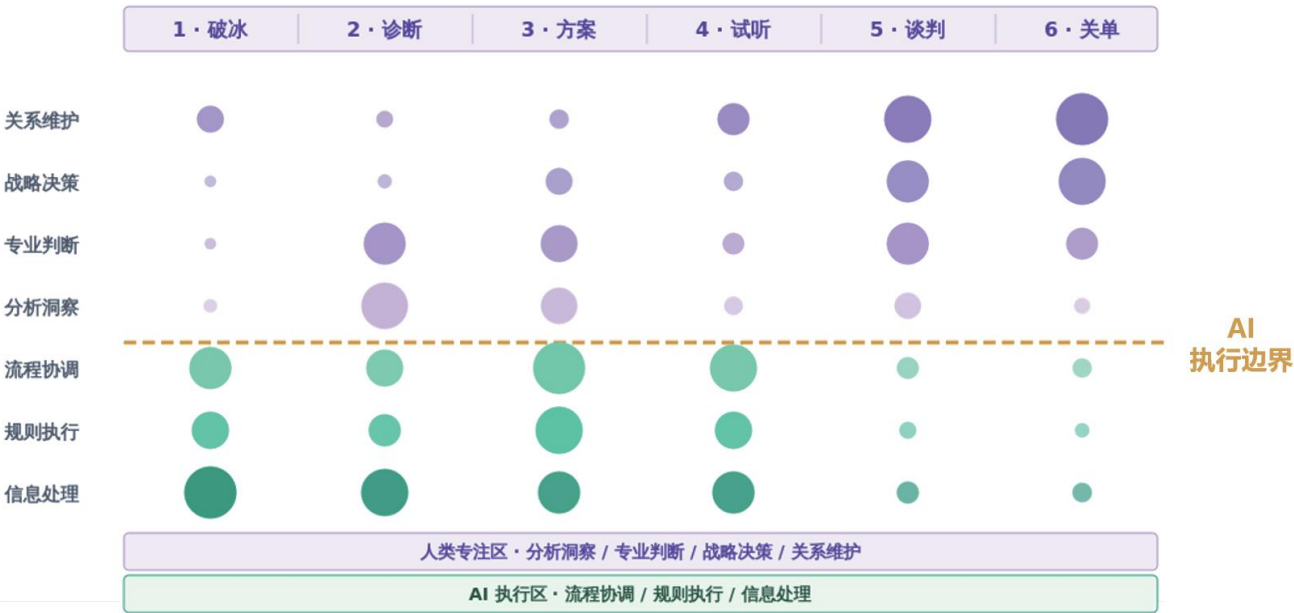
留学语培头部机构，1200万注册用户、500人销售团队。高客单（2-4万元）、焦虑驱动、一锤子买卖——销售70%时间做关系工作（信任/决策/谈判/关单），纯信息工作只占10%。传统“AI替代信息搬运”思路最多释放10%时间，不足以突破规模化瓶颈。真正的切入点：重构人机分工，让AI独立承接销售顾不上、做不动的环节。

发现

人机边界不是按“信息vs关系”一刀切，而是沿业务流程逐步映射、逐步上移。前4步以信息处理和流程协调为主，AI从副驾到独立承接；后2步关系密度陡升，AI仅提供策略支持。上移的触发条件不是技术单方面决定，而是数据沉淀、用户验证、组织配套、合规策略四者共同就绪。

方法论应用：6步销售流程 × 7层能力级别 · 人机边界精确映射

方法论应用（圆圈大小表示价值程度，颜色深浅表示人或AI的参与程度）



3.2 资产形态:从个人经验到系统复利

静态原则与动态原则共同驱动的第一个衍生结果,是组织核心资产形态的质变。

传统组织的脆弱性在于:知识寄存在人身上,经验随人员流动而流失。一个资深员工的离职,可能意味着一整套未被文档化的业务判断逻辑的永久丧失。组织能力本质上是个体经验的简单加总,既无法复制,更无法复利。

AI原生 workflows 改变了这一底层逻辑。在新范式中，任务执行过程天然就是数字化、可记录、可回溯的。当执行结果持续回灌到提示词模板、知识库与 workflow 配置中，每一次执行都在为下一次积累优化条件。组织能力不再随人员流动而归零，而是沉淀为可迭代的系统资产。

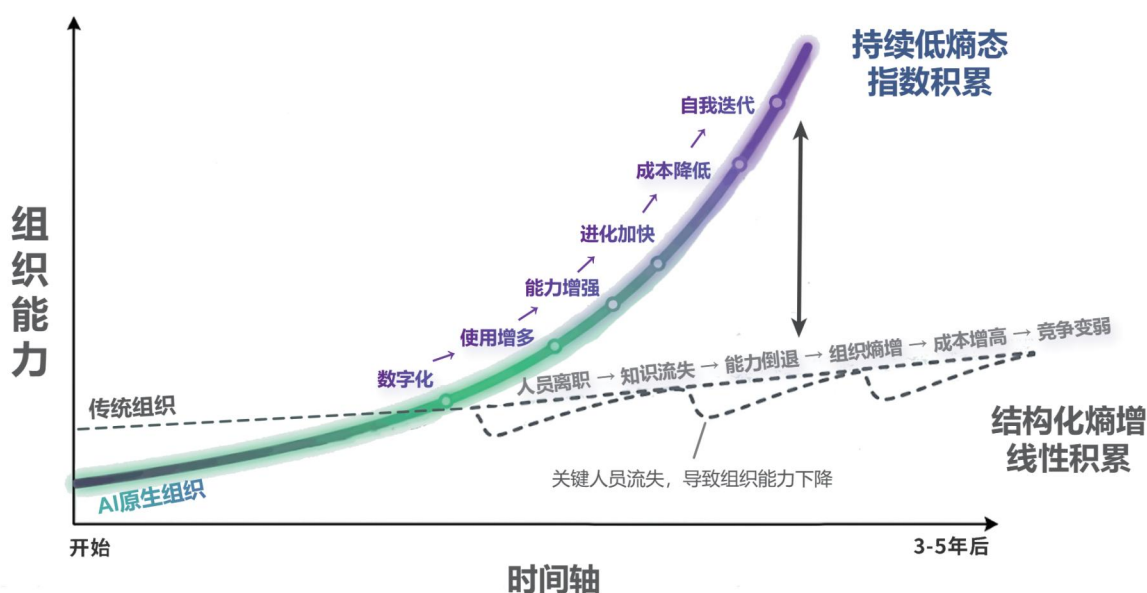


图 3-3 组织能力进化：AI 原生 vs 传统组织

传统组织受到关键人才流失、知识耗散与结构化熵增等制约, 组织能力缓慢线性增长甚至出现下降, 天花板低; AI 原生组织由于天然的数字化基因, 过程数据及业务结果持续反哺业务引擎, 保持组织低熵状态, 实现复利增长。

“运行即沉淀”使组织知识首次具备了复利特征：系统资产的价值随使用次数增长而非折旧，且边际积累成本趋近于零。这意味着，新范式下的组织竞争力不再取决于拥有多少经验丰富的人，而取决于系统资产的积累深度与迭代速度。常设团队将变得更小，但对目标定义的精度与系统运营的质量要求将空前提高。

由此产生的竞争后果是组织进化速度的代际分化：传统组织的能力增长受限于人员经验的线性积累，且始终面临人员流失导致的能力倒退风险；而系统资产驱动的 AI 原生组织，其能力随每一次执行自动沉淀、持续复利，进化曲线将从线性转向指数型。两种组织形态之间的能力差距，更将随时间呈现加速扩大的趋势，导致传统范式的组织会迅速丧失市场竞争力而被淘汰。

案例3-3：国企软件 · 系统资产复利

15人的活，4人干完了

执行能力如何从个体经验迁移至系统资产——一家500人国企的火种团队实验

8.3×

资源投入压缩
100人月 → 12人月

90→10%

代码执行时间占比
从“做”转向“督”

6→1

职能层级坍塌
水平分层 → 端到端

旧范式 · 能力寄存于个体

- 01 知识存在员工脑中，难以传承
- 02 能力 = 个体经验 × 人数
- 03 经验随人员流动而流失
- 04 扩张必须线性增员

→
能力
迁移

新范式 · 执行层沉淀为系统资产

4人火种团队 = 15人标准配置产出

AI接管代码生成、测试与部署。工程师从“职能边界内执行”转为“架构审查与交付担责”。执行能力从个体经验迁移至可复用的系统资产——每次执行即沉淀，每次调用即复利。

背景

500人规模国有软件企业内部实验，年人力成本超8000万元。长期沿用水平分层开发模式，以“人力投入换项目收入”为商业逻辑——组织能力完全依附于个体经验的线性堆叠。面临竞争力根基动摇时选择最小切口：精选4名工程师组建“火种团队”，经AI原生沉浸训练后以企业级平台为验证战场。

发现

8.3×压缩的真正来源不是“用更少的人”，而是“执行能力从个体经验迁移至系统资产”。火种成员的复盘——“过去九成时间写代码，现在倒过来”——正是这一迁移的注脚。与AI原生团队33倍压缩的案例对照：差距不在人，而在组织起点。历史包袱越重，压缩比越接近下界；但即便下界，8倍也已是代际跃迁。组织能力不再取决于拥有多少资深员工。

3.3 经济结构：成本连续化与定价迁移

两条原则驱动的第二個衍生结果，是组织经济结构的重塑。

成本结构的连续伸缩。旧范式的特征是价值按任务实现，成本按资源预置。企业必须预先支付固定薪资和软件订阅费，天然包含大量结构性浪费——人力闲置、为软件冗余功能付费、为需求波动预留产能冗余。当认知劳动被大规模改写为按使用量（Token 或算力）计费的计算服务时，成本的基本单位从岗位 / 席位转向了 workflow 执行。企业的成本结构首次可以随业务负载连续伸缩，价值曲线可以无限趋近于成本曲线。

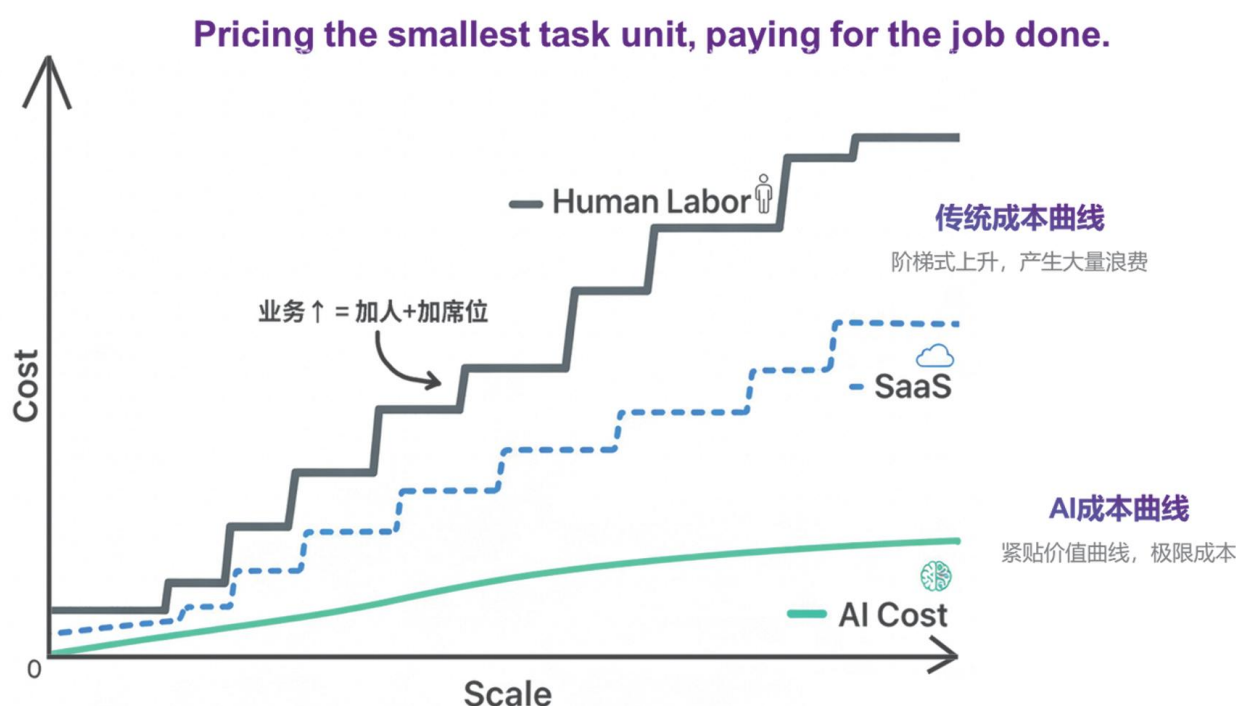


图 3-4 经济结构：AI 成本曲线 vs 传统成本曲线

传统人力与SaaS的成本曲线呈现出阶梯式增长，导致成本投入无法避免的浪费。今天的AI能力可以实现按最小任务单元交付和计价，从而实现价值曲线与成本曲线的无限贴合，极大提高要素生产率。

定价逻辑向“交付结果”迁移。成本结构的变化必然向外传导至商业定价。传统专业服务（法律、咨询、设计）的定价公式是“专家单位时间成本 × 任务时间 + 地理溢价”。当 AI 将任务时间压缩数个数量级并消除地理溢价时，以投入量（Billable Hours）为锚的定价模式将全面失效。服务提供商将被迫转向以交付成果为锚的定价逻辑，即 Results as a Service（RaaS）：客户为结果付费，而非为过程投入付费。

案例3-4：上海律所 · 定价模式重构





图 3-5 AI 绩效四维度评估拓扑图

该框架不仅是对内的资源配置工具——决定哪些 Agent 可以托管更高价值的业务，更是对外以 RaaS 模式交付时的可审计依据。具体的指标设计详见附件。

3.5 约束条件:技术局限与治理底线

“AI 能力最大化”必须与“AI 风险可治理”并重。新范式的内生约束主要来自两个层面，必须在系统设计中前置应对：

第一，概率生成与确定性交付之间的张力。大模型的本质是概率预测，而商业交付要求确定性。这不是模型“还不够好”的阶段性问题，而是技术原理层面的结构性张力。应对策略不是等待模型完美，而是建立容错架构：多模型交叉验证、RAG 知识挂载降低幻觉、强制置信度拦截机制——将概率性技术包装为确定性交付。

第二, 系统自治与责任追溯之间的悖论。系统越自治, 决策路径越黑盒, 错误归属越模糊。当 AI Agent 自主完成了一笔错误的交易或一份有瑕疵的合同审查, 责任归属于谁? Klarna 的 AI 客服在标准化退款等规则明确的场景中表现优异, 但当系统被授权独立处理订单纠纷等需要情境判断与情绪回应的场景时, 用户陷入循环应答而无法触达真人, 最终迫使公司重新招聘人工客服——这正是系统自治边界设定失当的典型样本: AI 在忠实的执行任务, 只是任务被错配, 且没有设置人在回路的断点⁹。应对策略的核心原则是: 最终责任不可让渡。无论系统自治程度多高, 高风险决策必须设置人在回路 (Human-in-the-loop), 且全部决策路径必须可回溯、可审计。管理重心从事前审批转向事中监控与事后追溯。

这两项约束是 AI 新范式与生俱来的特质, 构成了其运行的边界条件。正如工业生产需要安全规程, AI 生产范式需要将容错架构与责任机制内置于系统设计之中, 而非作为事后补丁。

AI 新范式已成型, 但不会均匀席卷所有行业。

9 Siemiatkowski, S. 接受Bloomberg采访 (2025年5月8日), 承认此前过度以成本为导向削减人工客服导致服务质量下降, 宣布重新招聘人工客服。来源: Bloomberg, "Klarna Turns From AI to Real Person Customer Service," 8 May 2025.

04

行业扩散： 新范式的不均匀渗透

Overview

本章从通用目的技术改造生产范式的历史规律出发,解析 AI 生产范式的扩散机制,构建阻力-动力模型,评估其在不同行业的扩散路径、时间表与利润池迁移方向。

4.1 通用目的技术的叠加扩散

经济史学家将通用目的技术定义为：具有广泛适用性、持续改进潜力、并能催生互补性创新的基础技术。过去两个多世纪以来，经济史学界普遍认定的几项重大通用目的技术包括：蒸汽机、电力、信息通信技术，以及当前的人工智能。由通用目的技术引发的范式迁移从来不是均匀漫灌——电力首先改造了城市照明，再渗透到工业生产，之后才逐步实现家用电器化；以互联网为代表的信息通信技术先颠覆传统媒体，再重塑零售和社交网络。

每一代通用目的技术的扩散都不是在空白画布上展开，而是叠加在前代技术所构建的基础设施之上。这种叠加效应产生两类截然不同的结果：

正外部性叠加——波峰叠波峰。当新一代通用目的技术的能力与前代技术的基础设施形成协同增强时，涌现出单一技术无法实现的突破性应用。典型案例是自动驾驶与智能网联汽车：没有互联网提供的实时通信与高精地图基础设施，AI 的感知与决策能力无处施展；没有 AI 的实时环境理解与路径规划，互联网连接的车辆仅是移动信息终端。类似的波峰叠加正在医疗（互联网远程诊疗基础设施 + AI 影像诊断能力）、教育（在线学习平台 + AI 个性化辅导）、金融（数字支付网络 + AI 实时风控）等领域也在发生。

负外部性叠加——波谷叠波谷。当各代技术的固有缺陷相互放大时，治理挑战也以非线性速度放大。最突出的案例是 AI 生成虚假信息的治理困境：互联网的设计基因是信息趋近零边际成本的复制与瞬时分发，AI 的能力是以极低成本批量生成高度逼真的图文及音视频内容。两者叠加，是任何一代技术单独存在时都不会达到的危害量级。类似的负向叠加还包括：AI 增强的网络攻击（互联网的攻击面 × AI 的自动化渗透能力）、工业级的个性化操控（社交媒体的用户画像 × AI 的精准内容生成）等。

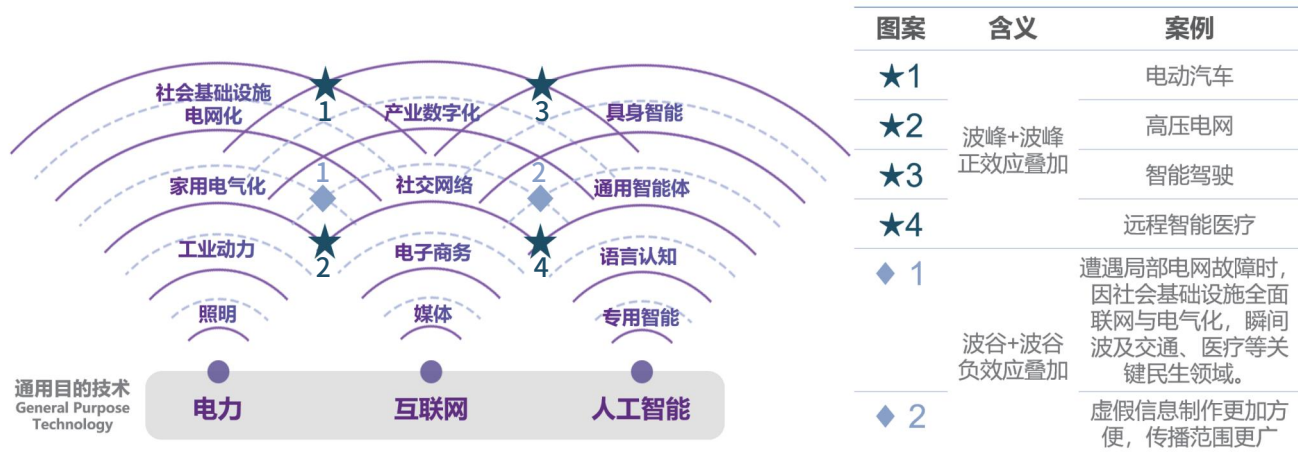


图 4-1 通用目的技术的叠加扩散效应

电力、互联网、AI 等通用目的技术沿特定路径(商业动力 VS 应用阻力)扩散且影响相互叠加，形成正向效应与负向效应叠加的涌现。

叠加扩散效应解释了 AI 在不同行业的扩散差异：凡是信息通信技术已深度渗透的行业（数字化程度高、数据基础设施完善），AI 叠加的阻力小而增益大，因而快速扩散；反之，数字化基础薄弱的行业，AI 缺乏叠加对象，导致扩散显著滞后。

4.2 AI生产范式扩散：动力-阻力模型

AI生产范式的产业化扩散，受制于两股力量的博弈：技术落地的阻力与商业变现的动力。我们将这两个维度定义为X轴与Y轴。

X轴：应用阻力——AI进入一个行业必须克服的客观摩擦

阻力越小，渗透越快。主要由四个要素构成，前三项为技术性阻力，第四项为非技术性阻力：

- 1. 数据就绪度。**行业核心知识是否已高度数字化、结构化？代码、金融文本等高度结构化的应用阻力最低；依赖线下隐性经验的领域应用阻力最高。
- 2. 容错率空间。**业务场景对概率性错误的包容度。营销文案容错率高，医疗诊断容错率极低。
- 3. 标准化程度。**业务流程是否具备清晰的输入输出边界与客观的评价标准。标准化程度越高，AI端到端交付的闭环越容易建立。
- 4. 监管壁垒与社会接受度。**技术能力就绪不等于可以落地。医疗、法律、金融等行业的强监管框架和公众信任阈值，构成了独立于技术之外的扩散摩擦力。技术的指数级演进与监管的线性制定之间存在结构性时差，这一时差在高监管行业尤为显著。

Y轴:商业动力——企业推动AI落地的商业势能

动力越强,转型意愿越高。主要由三个要素构成:

- 1. 认知劳动成本占比。行业利润是否被庞大的信息转译与加工人力成本严重侵蚀?认知劳动成本占比越高的行业,中间层压缩释放的经济价值越大,企业的转型动机越强。
- 2. 供需响应频率。业务是否面临海量、高频、即时的并发需求?高频并发场景(如海量客服、实时风控)从产能约束的“人手数→算力供给”转换中获益最大。
- 3. 规模化瓶颈。行业是否长期受制于优质专家产能有限而无法实现非线性增长?长期被专家产能瓶颈锁死的行业,AI带来的规模化突破动力最强。



图 4-2 AI 生产范式扩散的动力 - 阻力模型

AI生产范式扩散的本质,是商业动力与应用阻力之间的持续博弈。行业位置基于本章论述的七个评估要素(三类商业动力+四类应用阻力)定性推导,反映当前阶段的总体特征而非精确坐标。同一行业内部不同环节可能分属不同象限(例如“教育领域的题库生成”属于低阻力场景,而个性化教学涉及高信任与强监管,内槛显著更高)。随着技术迭代与监管演进,行业位置将持续变化。

4.3 四象限：行业扩散的时间表

将扩散门槛与采用驱动力交叉，得到 AI 生产范式的行业扩散四象限。这不仅是一张演进时间表，更是创业者与投资机构进行赛道选择的战略地图。

象限一：快速渗透区（高商业动力 × 低应用阻力）

该区域包括了 AI 扩散最成熟的行业。其数据往往高度数字化、容错率相对宽容、标准化程度高；同时认知劳动成本占比大、并发需求密集。AI 生产范式在这些行业的渗透从是否发生变成了已经发生到什么程度。

典型行业：零售 / 客服（如跨境电商）、营销 / 内容创作（如网络小说与自媒体写作）、代码编写 / 测试、基础财务与税务处理

演化路径：这些行业正处于从 AI 辅助人类执行向 AI 端到端交付、人类监督例外的快速跃迁中。率先完成闭环的企业将通过数据飞轮累积优势，后来者的追赶成本会急剧上升。

案例4-1：券商投研 · 快速渗透区 · 行业扩散

2天变20分钟：中型券商的产出能力跃迁

数据高度数字化 × 认知成本占比极高 × 并发需求密集 —— 投研是AI范式扩散最快的行业切面之一

200×

初稿生成时间压缩

2-3天 → 20分钟

10×

信息覆盖密度跃升

5-10份 → 50-100份来源

当晚

财报窗口期响应

次日出稿 → 当晚出稿

旧范式 · 人力串行四步流程

- 01 数据下载与研报检索
- 02 信息整理与行业对比
- 03 图表制作与分析撰写
- 04 复核校对与格式输出



AI
压缩

新范式 · AI自主研究 + 人定框架

AI多轮自主研究 → 分析师审视判断

AI自主拆解研究任务，并行检索财报、研报与市场数据，经多步推理与交叉验证输出结构化报告。分析师聚焦于定义框架、验证逻辑、提炼投资洞见。

— 背景

一家30余名分析师的中型券商研究所。投研天然AI即时爆发区的典型行业——数据高度数字化、认知劳动成本占比极高、财报季并发需求密集。头部券商百人团队当晚推送解读，中型所1-2人独立完成全流程，80%时间被锁死在信息搬运而非专业判断上。差距的本质不是判断力，而是信息加工的人力规模。

— 发现

AI抹平的不是判断力差距，而是信息搬运的规模差距。当初稿生成和数据整理被AI接管后，区分研究质量的唯一变量回到了分析师定义问题与识别逻辑漏洞的能力——这恰恰是头部券商垄断不了的东西。中型机构的突围路径第一次从“堆人力”变成“拼判断力”。这正是即时爆发区的竞争特征：先完成AI闭环的机构，通过数据飞轮建立累积优势。

象限二：高潜突破区（高商业动力 × 高应用阻力）

这里包括了最具战略价值但渗透最困难的行业。强烈的商业动力（专家产能严重受限、认知成本极高）与极高的应用阻力（低容错率、强监管、数据敏感）形成对峙。虽然面临技术落地难的挑战，但AI一旦突破应用阻力，将会带来极高的潜力与极高价值的回报。

典型行业：先进制造、医疗（如临床诊断）、金融（如风控与量化交易）、AI科研（如新药与新材料的研发）

演化路径：渗透以“切片式”推进——先在容错率相对较高的子场景（影像初筛、合同条款检索、风险预警）建立滩头阵地，逐步积累监管信任与临床/司法验证，再向核心决策环节扩展。突破的关键不仅是模型能力，还包括垂直领域的合规护栏与人机协作机制的成熟度。一旦突破，壁垒本身即转化为护城河。这些行业将成为大爆发中体量最大、利润最厚的价值高地。

象限三：规模普惠区（低商业动力 × 低应用阻力）

AI进入的应用阻力不高，但行业自身的商业动力有限，比如服务同质化程度高、客户付费意愿低，所以虽然AI能大量涌入这些行业并普及化，但客户付费意愿太低。

典型行业：基础办公、标准化教育（如K12基础教育）、通用聊天机器人

演化路径：AI以工具增强而非范式替代的方式渐进渗透。效率提升真实但温和，不会引发行业结构的剧烈重组。较低的应用阻力可以带来较大的规模，低动力低利润则带来普惠的属性，成为人人可得的效率工具。

象限四：长期探索区（低商业动力 × 高应用阻力）

该区域包括了应用阻力高但商业动力低的行业。这些行业往往数据未充分数字化，强依赖物理世界操作、人际信任或“人味”，或由于涉密或极重大利害关系（如核武器控制权）等非技术原因排斥AI的使用。

典型行业：传统非遗手工业、高端奢侈品定制、涉密政务核心决策

演化路径：短中期内AI渗透有限，行业结构保持相对稳定。但需注意：象限边界不是永久固定的。随着多模态能力、具身智能和监管框架的演进，今天的蛰伏区可能在五到十年后被重新归类。

案例4-2：英国科研设备 · 高潜突破区 · 能力去地理化

7年不发单，24小时破冰

高门槛精密制造的扩散临界点——当“必须物理接近才可信”的前提被AI改写

7 yr → 24h

从观望到破冰
首单24h内获十余家报价

~40

两年内项目数
跨3种工艺路径并行

11

跨国工厂网络
本地2-3家 → 全国11家

旧范式 · 物理接近才可信

- 01 依赖本地2-3家老供应商
- 02 按行业标签寻找同行业工厂
- 03 实地验厂建立信任链
- 04 需专职采购+质量+供应商管理

→
AI
压缩

新范式 · 按能力路由匹配

需求拆解为能力条件 → 全网络路由匹配

AI将需求拆解为精度、材料、批量、工艺复杂度等能力维度，在制造网络中寻找最匹配的能力节点。匹配逻辑从“同行业”变为“同能力”——工厂可能从未做过科研仪器，但精密加工维度恰好满足全部要求。

— 背景

英国一家成立约40年、团队仅十余人的蛋白质结晶设备企业，服务顶尖高校与科研机构。几十年依赖本地老供应商——不是不知道中国成本更低，而是不相信跨国陌生供应链能支撑高门槛研发制造。语言、验厂、知识产权，每一项对十余人的团队都是不可承受的组织成本。这是典型的高扩散门槛行业：数据未充分数字化、强依赖物理验证、信任建立成本极高。转折来自自身——创新加速后，老供应商承载力开始不足。

— 发现

高门槛不等于不可渗透——关键是找到突破扩散门槛的临界条件。当AI把供应商能力变成可远程验证的数字档案，“必须物理接近才可信”的旧前提被改写。空间压缩没有降低苛刻要求，反而让这些要求第一次能在跨国网络中被快速匹配。十人团队无需建采购部即可调度全球制造网络——一旦门槛被突破，高壁垒行业的扩散速度同样可以极快。

4.4 利润池迁移：价值向两端集中

随着 AI 生产范式从象限一向其余象限依次扩散，宏观产业的利润池正在发生不可逆的再分配。第二章已论证了价值链中间层塌陷的机制，本节聚焦于塌陷之后——释放出的价值正在流向何处。

利润的再分配沿三个方向同时发生：

第一，中间层利润向终端用户溢出。中间服务层（流程外包、初级咨询、代理商）被压缩后，其原本截获的价差并不会凭空消失。一部分以价格下降的形式直接让渡给终端消费者和中小企业——过去支付高额专业服务费的需求方，将以极低成本获得同等甚至更优的交付质量。

第二,基础设施层截获“系统租金”。掌握算力集群、基础大模型与云服务的平台,通过API调用和Token计费,稳定截获全行业数字化运转的基础租金。其商业地位类似工业时代的电网与铁路——不直接参与具体业务,但所有业务的运转都必须经过其基础设施。

第三,意图发起层获得杠杆化回报。拥有核心业务场景、独有数据与深度行业Know-how的企业(或超级个体),无需再供养庞大的中间执行团队,即可将业务意图直接对接AI基础设施完成交付。中间成本的压缩,直接转化为意图发起者的利润率跃升。

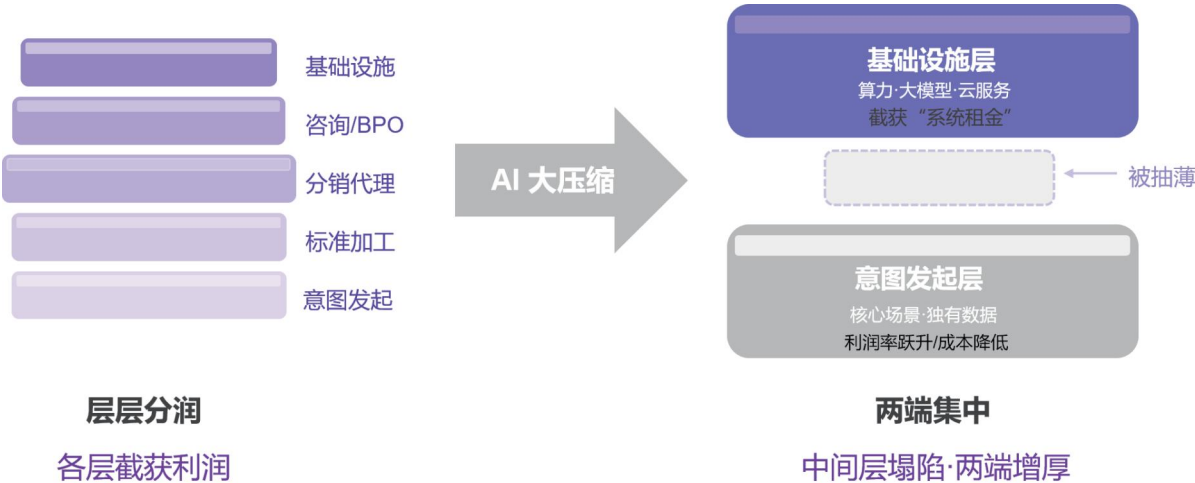


图 4-3 AI 大压缩导致传统价值链的利润池向两端集中

随着价值链中间层的塌陷,过去层层分润的结构被AI压缩,利润向基础设施层与意图发起层(客户场景)双向汇聚,中间层被抽离。

三重再分配叠加,宏观利润结构正在从传统的层层分润转向两端集中——价值向提供算力与模型的基础设施层和定义问题与拥有场景的意图发起层汇聚,中间的信息转译与标准加工层被持续抽薄。这一结构与产业经济学中的“微笑曲线”高度同构——施振荣在1990年代描述的制造业价值链两端高、中间低的利润分布,正在被AI大压缩从硬件制造推广至一切认知劳动密集型行业,且中间层的彻底塌陷使曲线的弧度被急剧加深。

05

大爆发：

被 AI 大压缩释放的新价值与创新机会

Overview

技术史反复证明，每一次大压缩的终局不是萎缩，而是爆发——正如电力压缩了能源获取路径，催生了电气化的工业生产与家庭生活；互联网压缩了实体传播渠道，催生了电商与创作者经济；AI 大压缩正在消除认知劳动的中间摩擦，而这将带来更大的价值爆发。



图 5-1 大压缩带来创新机会

随着价值链中间层的塌陷,过去层层分润的结构被AI压缩,利润向基层层与意图发起层(客户场景)双向汇聚,中间层被抽离。

5.1 释放出来的两类价值

两类长期被认知成本封锁的价值空间正在被打开:过去无力支付的买方首次进入市场(需求侧),过去无法独立交付的个体与小型组织首次获得全链路生产能力(供给侧)。

需求侧价值:被经济性压抑的需求被激活。

过去,私人医生、专属理财顾问、一对一法律咨询是少数人的特权——不是因为需求不存在,而是因为专家产能受限于人脑的生物学瓶颈,供给成本居高不下。当AI将专业能力输出的边际成本压缩多个数量级后,“极致个性化”从奢侈品变为可规模交付的基础服务。每一个普通消费者和小微企业,都成为此前只有高净值客户才能触达的专业服务的潜在买家,形成全新增量市场的爆发。

供给侧价值:受资源门槛抑制的生产力被激活。

过去,一个人拥有好的创意或深度的行业洞察,却无法独立完成从产品开发到全球交付的完整链路——代码编写、多语言适配、合规审查、营销投放,每一环都需要专业团队。认知链路的压缩第一次使一人公司(OPC)或数人团队具备了端到端的全链路生产能力。由此释放的不仅是一批新的采购需求,更是一个全新的生产者群体——创业门槛的骤降激活了大量拥有行业洞察却曾经缺乏资源来执行的个体,他们既是AI原生工具链与交付基础设施的使用者,更是新价值的直接创造者。

案例5-1：自行车 · 超级个体 · 大爆发

核心件\$7.5，终端售价\$200

当创业门槛降到一个人一张信用卡——大压缩释放的超级个体全链路生产需求

\$7.5

核心件制造成本
终端售价 \$200+

8_h

首批配件授盘
从询盘到收到报价

1_人

全链条管理
零新增岗位

旧范式 · 先建组织才能做产品

- 01 创始人个人寻源 · 行业内搜索
- 02 逐一比价 · 跨语言沟通图纸
- 03 反复试验验证加工能力
- 04 跟踪打样 · 确认交期与质量



新范式 · 基础设施替代组织后台

不必先建组织，即可调度工业级制造网络

AI将需求拆解为能力条件，在制造网络中按能力标签路由——难件分发给高精密节点，简单件分给成本最优节点。比价、试错、沟通成本同时塌缩。一个人完成多供应商、多工艺的跨国采购全链条。

— 背景

英国极致精简的一人公司——1名董事=1名股东=1名实控人。创始人拥有航空航天与F1赛车工程背景，产品聚焦高精度自行车中轴系统。技术能力不缺，瓶颈在后台：要把精密传动系统从图纸变成产品，必须先复制出一套采购、跟单和跨境协调的组织后台。这正是大压缩前“超级个体”被锁死的典型困境——有洞察、有技术，却无法独立完成从设计到全球交付的完整链路。

— 发现

对超级个体而言，AI平台不是效率工具，而是基础设施——它让一个人不必先建组织，就能调度工业级制造网络。这正是大压缩释放的第二类需求：当认知链路被折叠，创业门槛降到极低，被激活的不是一批新员工，而是一个全新的创业群体。组织没变大，能力边界被彻底拉宽——这是大爆发中“超级个体全链路生产需求”的活标本。

5.2 价值捕获的生态位

两类价值空间的打开，催生了大量的市场机会。市场参与者沿资源禀赋与历史包袱分布在一连续光谱上，光谱两端是两类典型形态——AI 原生创业者与传统行业龙头，大多数企业则处于二者之间的过渡带。两端的差异最具结构性意义，可作为分析坐标。

表 5-1 价值捕获的两类典型生态位

维度	AI原生创业者	传统行业龙头
主场价值空间	超级个体生产力释放	高认知密度行业深度解锁
延伸价值空间	标准化场景的普惠化:法务初审、基础税务、模板化营销等	高信任场景的普惠化:医疗初筛、金融产品推荐等
核心资产	可快速结构化的轻量 Know-how + 场景切片能力,数据驱动的快速迭代飞轮	深度行业 Know-how + 合规与信任体系,高质量数据驱动的深度优化飞轮
关键转型动作	选择反馈回路短的垂直场景,尽快建立工作流闭环	知识资产化:把散落在文档与资深员工头脑中的 Know-how 转化为可调用的系统资产
主要风险	基础大模型能力升级直接覆盖其工作流	转型意味着主动重组曾创造巨大价值的中间层,首先是利益再分配,其次才是技术实施
典型商业模式	卖工作流/卖结果(RaaS)	能力输出 (API / 数据 / 结果)

光谱中段的企业——已有部分行业积累、又在进行 AI 原生改造的中小企业,同时面对两端的挑战与机会,需要根据自身资源禀赋决定向哪一端倾斜。

无论身处光谱何处,胜负由两件事决定:一是**系统资产的积累速度**——先跑起来的一方,每一天都在拉大与观望者的能力差距;二是**共生接口的建立**——最大的商业机会出现在两端的交叉地带:龙头通过数据授权与 API 开放能力, AI 原生企业通过工作流闭环完成最后一公里交付。封闭的数据不会自动变现,孤立的工作流无法突破信任壁垒。

06

结语： 压缩即创造

AI 压缩了旧世界，同时触发了新价值的大爆发。压缩的发生，正是新价值涌现的前提条件。前五章论证了这场重塑在产业层面的近景；但压缩释放出的图景，远不止于此。

近处，价值链被重塑：**新价值**——压缩释放的经济价值，正在向基础设施层与意图发起层重新汇聚；**新工作**——人类劳动向目标定义、例外处理与责任承担三项不可替代能力收敛；**新机遇**——一人公司得以完成过去需要数十人协同的业务闭环，垂直 Know-how 第一次能被直接交付。

远处，人类生存与组织的形态本身正在被重新构想。

新空间——当 AI 持续理解每个个体的偏好与意图，数字空间将从被动的内容容器演变为以个体为中心、主动响应的生存场域；实体与数字之间的边界，正在被重新划定。

新体验——千人一面的标准化曾是工业化的副产物。当 AI 与个人共同生成内容并随情境实时流动，标准化将让位于千人千面的实时生成。

新意义——当 AI 承接了认知链条中可标准化的部分，人得以从重复中抽身，重新投向机器无法替代的活动：值得追求的目标、需要勇气的判断、需要真诚的关系、需要灵魂的作品。

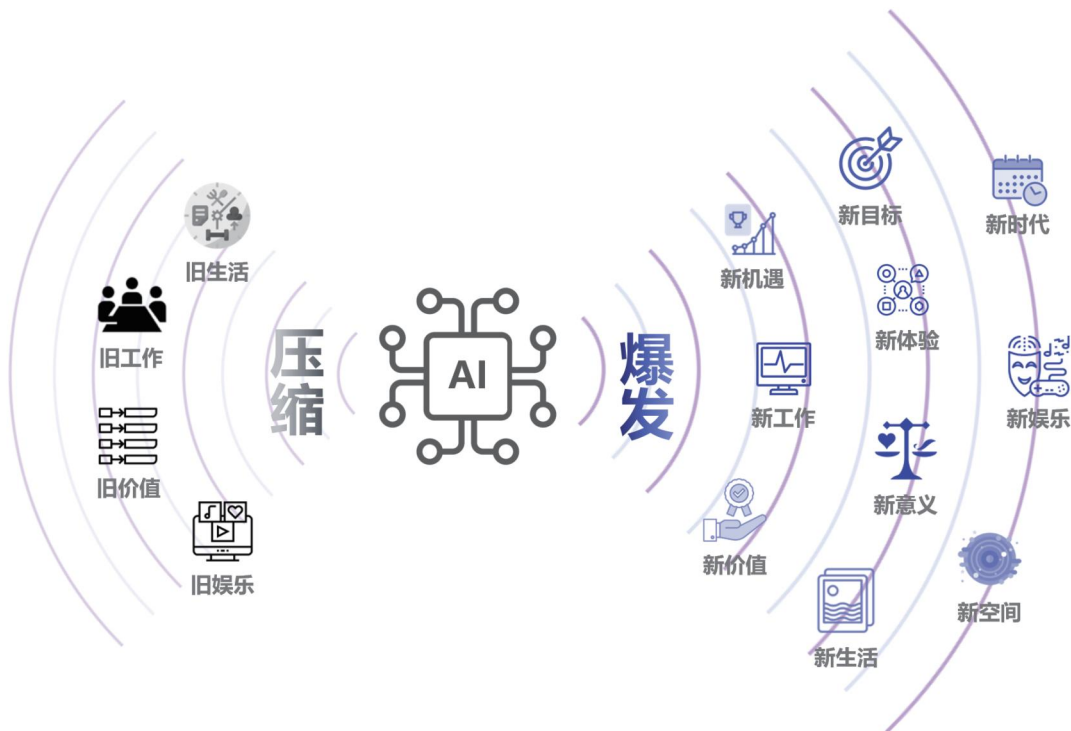


图 6-1 AI 大压缩触发新价值的大爆发

AI 大压缩消解了旧范式的工作与生活形态，同步爆发出新价值、新工作、新空间等更多元的可能性——压缩的另一面，是创造。

近处的图景已在发生，远处的图景才刚露端倪。但二者共同指向同一个判断：这场变革的尺度，远比产业重组本身更深远——它同时触及组织形态、认知方式、个体生存的同步迁移。任何只把它当作效率工具升级来理解的视角，都会错过更大的图景。

新空间、新体验、新意义、新时代、新生活、新目标、新娱乐——压缩消除的固然多，但爆发后释放的更多。压缩的是人类并不擅长且不享受的部分；爆发的创造刚刚开始，在我们每一个人面前。

(全文完)

附录

Appendix

附录1: AI大压缩行业扩散及对劳动岗位的影响

随着AI大压缩在各行业的梯次渗透,劳动力市场正在经历一场超越传统自动化周期的结构性重塑。这种重塑并非简单的“机器换人”,而是对传统岗位中“认知执行环节”的彻底剥离与重组。

1. 核心机制:任务解构与中间层的“空心化”

在评估AI对就业的影响时,最大的认知误区是将“岗位(Job)”视为最小分析单元。实际上,岗位是多个“任务(Task)”的集合,AI大压缩的本质是**任务解包与重新分配**。

“人类API”的失效:在传统企业架构中,大量初中级白领(如初级分析师、基础程序员、文案策划)扮演着“人类API”的角色——负责在不同系统或部门间搬运、翻译和汇总信息。当大模型具备跨模态的意图理解与生成能力后,这部分高频、标准化的信息处理任务被直接压缩。

组织结构的“折叠”:随着中间执行层的价值被AI抽离,传统的金字塔型组织架构开始向“沙漏型”甚至“扁平网络化”折叠。企业不再需要庞大的中层团队来向下拆解任务和向上汇报进度,管理跨度(Span of Control)被极大地拓宽。

2. 价值转移:从“执行者”到“编排者”的跃迁

旧岗位的消亡必然伴随新价值锚点的确立。在AI大压缩的范式下,劳动者的核心竞争力从“高效执行”向“系统编排与结果判别”转移。

超级个体的规模化:借助AI Agent,单个具备深厚行业Know-how的专家可以指挥一个由多个模型组成的“虚拟部门”。个人的产出上限不再受限于物理时间和精力,而是取决于其提出问题(Prompting)和调动算力的能力。

“品味”与“判断力”的溢价:当生成内容的边际成本趋近于零时,内容的“生产”不再稀缺,稀缺的是对优质内容的“筛选”与“把控”。在医疗、法律、核心代码等高容错成本领域,具备最终签字权和责任兜底能力的领域专家,其市场溢价将进一步走高。

3. 岗位演化矩阵与结构性重塑

为了更清晰地预判不同职业的未来走向,我们将现有岗位按其在AI大压缩中的受冲击程度与演化方向,划分为以下三大阵营:

岗位阵营	核心特征(底层逻辑)	典型代表	演化终局与趋势
被压缩类 (结构性缩减)	高信息密度、规则明确、缺乏物理交互与情感连接	初级代码编写、基础翻译、客户服务、数据录入	编制冻结,存量岗位被AI Agent直接替代,转化为按需调用的API
被增强类 (赋能与杠杆)	强战略决策、复杂物理操作、深度情感与信任建立	架构师、心理咨询师、高级销售、外科医生	剥离繁杂案头工作,专注核心高价值环节,人均产能实现指数级跃升
新生类 (范式伴生)	围绕AI系统的训练、编排、治理与合规	提示词工程师、AI工作流架构师、模型伦理审计员	成为新一代基础设施的建设者与维护者,填补技术与业务之间的断层

通过该矩阵可以看出,纯粹的“认知体力劳动者”将面临最严峻的出清风险,而能够将业务逻辑转化为AI工作流的“复合型架构人才”将迎来历史性的红利期。

4. 组织与个人的应对策略

面对不可逆的 AI 范式转移,企业和个人都不能再被动观望,必须主动调整价值模型,跟上生产力变革的节奏。

企业端:告别“按人头扩编”,转向“算力+核心大脑”的高效模式

预算与招聘计划上,建议把原本给初级执行团队的运营支出,分一部分到 AI 基建采购、内部模型微调和端到端自训练上。同时要注意:初级执行团队的调整不是简单裁员,要兼顾企业社会责任,通过技能再培训,把员工从“单纯执行者”变成“AI 的监督者、调优者”。建 AI 中台不是唯一解法,按需选择最适合业务的落地方式就好。

个人端:别只做“解答者”,要做会和 AI 对话的“对话者”

个体劳动者应主动跳出标准化技能的内卷竞争,将核心精力投入到跨学科视野构建、复杂问题拆解能力提升,以及对 AI 生成结果的批判性判断力培养上。在未来的劳动力市场中,提出高质量问题、与 AI 协同解决复杂问题的能力,将比输出标准化答案更具价值。

附录2:四维度10指标AI绩效评估体系

维度	核心问题	指标	定义	建议达标基准
交付	能不能用?	① 一次交付率	首次执行即正确且可用,无需人工修改或重试	根据具体场景、基模能力、实施阶段、综合确定且动态调整(如初始阶段≥95%)
		② 最终交付率	含重试和人工补救后最终交付的比例	根据具体场景、基模能力、实施阶段、综合确定且动态调整≥99%
		③ 平均执行时间	一次交付成功任务的平均耗时,按类型分层	按任务类型设SLA
成本	划不划算?	④ 单任务算力成本	每个任务实际消耗的算力资源费用(COGS)	按类型设上限,趋势不上升
		⑤ 投入产出比	每投入1元AI总成本产出的可量化业务价值	根据具体场景、基模能力、实施阶段、综合确定且动态调整(如初期≥5:1)
		⑥ 护栏覆盖率	高风险操作已部署拦截机制的比例	100%
安全	放不放心?	⑦ 异常拦截率	运行时异常被自动识别并阻断的比例	根据具体场景、基模能力、实施阶段、综合确定且动态调整(如初期≥95%)
		⑧ 日志完整率	操作日志记录完整五要素的比例	100%
可控	好不好管?	⑨ 人工介入率	需计划外人工补救才能交付的任务比例	趋势下降。但需根据具体场景、基模能力、实施阶段、综合确定且动态调整(如初期≤5%)
		⑩ 调优有效率	每次调优后目标指标确实改善的比例	根据具体场景、基模能力、实施阶段、综合确定且动态调整(如初期≥70%)

附录3：核心概念术语表

概念	解释
认知劳动	依赖人类脑力完成的知识性工作, 包括理解意图、转译信息、分析判断、编排执行等。在传统价值创造链条中, 认知劳动承担着“业务意图→工具操作→结果输出”全链路的核心中间环节。
AI大压缩	AI对认知劳动中间环节的系统性压缩。区别于此前能源、材料、信息领域的压缩, 本次压缩的对象首次从物理世界的外部约束转向认知活动本身。
零基设计 (Zero-Based Design)	一种组织设计方法论, 不从现有岗位和流程出发做渐进改良, 而是以AI能力边界为起点反向推演, 先确定AI应端到端完成哪些环节, 再确定人类必须保留哪些职能。
AI原生 workflows	从设计之初就以AI Agent为默认执行主体的业务流程。区别于在传统流程上叠加AI工具的增量式改造, AI原生 workflows 的流程结构、人机分工与数据流向均围绕AI能力边界重新设计。
系统资产	在 AI 原生 workflows 中随任务执行自然沉淀的组织能力资产, 包括但不限于: 提示词模板、知识库、工作流配置及执行记录等, 构成 AI 原生企业新的 Knowhow。与随人员流失而耗散的个人经验不同, 系统资产可记录、可追溯、可持续迭代。

概念

解释

人在回路
(Human-in-the-loop)

在 AI 自主执行流程中设置的人类审核断点。当任务涉及高风险决策、模糊情境判断或责任归属时，系统在关键节点暂停并交由人类确认，确保最终责任可追溯。

AI大压缩:智能经济时代的生产范式转移

The Great AI Compression: A Production Paradigm Shift in the Age of Intelligence

【联合发布】

香港大学人工智能、管理与组织研究中心 (HKU CAMO)

中国信息通信研究院华东分院

模驭人工智能科技(上海)有限公司

【学术指导专家】

李 晋¹

【特别顾问】(按姓名首字母排序)

陈雪频 ⁷	董 明 ²	付 伟	余 莹 ⁸	杨晶晶 ⁶
钟俊浩 ⁵				

【编写团队】(按姓名首字母排序)

常永波 ³	贺宇琪 ¹²	胡康薇 ⁸	李泽涵 ^{* 4 11}	马海兵 ^{4 10 11}
孙泽浩 ⁹	邬华伟 ⁸	许怀谷 ¹¹	袁静劼 ¹²	张 弛 ¹¹
仲子骞 ¹¹	祝 维 ¹¹			

【指导机构】

中国创造学会人工智能专委会

上海市人工智能行业协会

1 香港大学人工智能、管理与组织研究中心 (HKU CAMO)
2 上海交通大学安泰经济与管理学院
3 中国信息通信研究院华东分院
4 中国创造学会人工智能专委会
5 上海市人工智能行业协会
6 上海大模型生态发展有限公司 (模速空间)
7 上海融智文化传播有限公司

8 上海海智在线网络科技有限公司
9 上海大学社会学院
10 上海中侨职业技术大学
11 模驭人工智能科技(上海)有限公司
12 上海通茂律师事务所
* 通讯作者

【版权声明】

© 2026 香港大学人工智能、管理与组织研究中心、中国信息通信研究院华东分院、模驭人工智能科技（上海）有限公司。本报告内容受版权保护。

本报告允许在非商业用途下转载、引用与传播，但须以下列方式注明来源：

来源：《AI 大压缩：智能经济时代的生产范式转移》，
香港大学人工智能、管理与组织研究中心、中国信息通信研究院华东分院、模驭人工智能科技（上海）有限公司联合发布，2026 年 5 月。

未经联合发布单位事先书面许可，不得以任何形式将本报告内容用于商业用途，亦不得对内容进行实质性修改后再行传播。

2026 年 5 月 28 日 v1.0

【免责声明】

本报告基于公开资料、行业访谈及主创单位研究分析形成，仅供行业研究与信息参考之用，不构成投资建议、法律意见或业务决策依据。

报告所引之数据、案例及第三方资料均来源于公开渠道，编写团队已尽合理努力核实，但不对其准确性、完整性或时效性作出保证。文中涉及的企业与案例仅作为分析支撑，不构成任何评价或推荐。报告中的趋势判断、分析框架及前瞻性观点，反映发布时点的认知，可能随技术演进、市场变化及政策调整而发生变化。读者应结合自身情况独立判断。

本研究报告结合产业实情与社会发展态势综合分析，力求视角全面。囿于研究认知有限，社会影响因素繁杂，报告难免存有疏漏与观点欠缺之处。我们谨怀谦逊之心，恳请各界同仁多多指正，我们将虚心吸纳意见，持续优化完善研究内容。

【生成式 AI 使用声明】

本研究报告在撰写过程中使用了生成式 AI 工具辅助文献检索、数据整理、文稿起草与多语言翻译。所有 AI 生成内容均经过编写团队的人工审核、事实校验与专业判断，最终文本的观点、结论及任何错误由编写团队承担责任。

我们认为，公开披露 AI 的参与方式是知识生产诚信的基本要求，也与本文倡导的“人机协作、人承担最终责任”原则一致。

联系我们

香港大学人工智能、管理与组织研究中心
(HKU CAMO)



CAMO
HKU Centre for AI,
Management and Organization



camo@hku.hk

<https://camo.hku.hk/>

中国信息通信研究院
华东分院



中国信息通信研究院
华东分院



infoeastchina@caict.ac.cn

<http://www.ecit.org.cn/>

模驭人工智能科技
(上海)有限公司



模驭AI
moximize.ai



MKT@moximize.ai

<https://moximize.co/>