

讨论论文

Discussion Paper

当 AI 越来越聪明： 技术冲击下知识型工作的分化与重组 —— 基于 Ide 和 Talamas (2025)

杨燕青、陈浩文、姜芮
2026 年 7 月

摘要

【摘要】本报告以 Ide 和 Talamàs (2025) 提出的知识层级理论为分析工具，考察人工智能能力持续提升和自主性增强背景下，知识型工作的重组方向及其分配后果。报告首先梳理三个正在显现的现实信号：专业服务业入门岗位收缩、AI 辅助工具对经验积累较少者的支持更为明显，以及“小团队、大产出”型企业涌现；继而在原模型设定下进行参数化求解，系统推演 AI 能力由弱到强过程中职业结构、工资分布与企业形态的演化。

分析表明，AI 并非简单替代人类劳动，而是根据其能力相对于劳动者知识分布的位置，重塑人与技术之间的分工，并随能力提升发生方向性翻转。当基础型自主 AI 尚处于一般生产任务的能力区间时，入门岗位和部分处于中间知识区间的专业人员面临较大调整压力，工资不平等可能扩大；当 AI 能力进一步提升并能够提供高质量顾问支持时，处于职业起步阶段、知识与经验积累相对有限的劳动者也有机会分享 AI 红利，不平等扩张随之放缓。与此同时，AI 自主性是影响收益分配的重要因素：在能力相同的情况下，非自主 AI 更有利于知识技能积累较少的劳动者，但其总产出低于自主 AI。

因此，AI 治理既不应将技术视为需要防范的外部威胁，也不能完全由市场机制分配转型成本。更可取的方向，是在积极拥抱 AI、释放其生产潜力的同时，完善职业培训、人才成长通道、责任规则和再分配机制，维护劳动者尊严，使技术进步转化为更加广泛、包容的发展机会。

【关键词】人工智能 知识型工作 劳动分工 收入分配 智能体自主性

Abstract: This report applies the knowledge-hierarchy framework developed by Ide and Talamàs (2025) to examine how rising AI capability and autonomy reshape knowledge work and affect the distribution of economic gains. It first reviews three emerging developments: the contraction of entry-level jobs in professional services, the stronger benefits of AI assistance for workers with less accumulated experience, and the rise of “small team, large output” firms. It then parameterizes and solves the original model to trace the evolution of occupational structure, wage distribution, and firm organization as AI capability increases.

The analysis shows that AI does not simply replace human labor. Its effects depend on where its capability lies relative to the workforce’s knowledge distribution, reshaping the division of labor between humans and technology and producing a directional reversal as AI improves. When basic autonomous AI remains within the capability range of ordinary production tasks, entry-level workers and some professionals in the middle of the knowledge distribution face greater adjustment pressure, and wage inequality may widen. As AI becomes more capable and provides high-quality advisory support, workers at earlier career stages, with relatively limited accumulated knowledge and experience, may also gain from AI, slowing the expansion of inequality. AI autonomy is another important determinant of distribution: at the same capability

level, nonautonomous AI is more favorable to workers who have lower knowledge in performing their jobs, but generates lower total output than autonomous AI.

AI governance should therefore neither treat technology as an external threat to be contained nor leave the costs of transition entirely to market forces. A more balanced approach is to embrace AI and unlock its productive potential while strengthening vocational training, career-development pathways, accountability rules, and redistribution mechanisms. Such an approach can protect human dignity and ensure that technological progress creates broader and more inclusive opportunities.

Keywords Artificial Intelligence; Knowledge Work; Division of Labor; Income Distribution; AI Autonomy

目 录

1 引言：AI 冲击知识型工作的两种对立叙事	4
1.1 现象：AI 能力的快速扩张	4
1.2 矛盾：两组相互冲突的经验证据	5
1.3 本报告的分析路径	6
2 事实：知识型工作正在发出的三个信号	6
2.1 信号一：入门岗位收缩	7
2.2 信号二：AI 使用收益的不均等	7
2.3 信号三：“小团队、大产出”企业的出现	8
2.4 现有解释的缺口	9
3 框架与推演：AI 能力提升过程中知识型工作的重组	10
3.1 理论骨架：知识分工、层级与 AI 的进入方式	10
3.2 职业版图的演化与“大翻转”	11
3.3 三类劳动者的命运曲线	12
3.4 不平等的阶段性：先扩大、后收敛	14
3.5 企业形态的重塑：组织收缩与“超级小企业”	15
4 延伸：AI 自主性的分配后果	17
4.1 “顾问”与“员工”：两种形态、两种格局	17
4.2 效率与公平的权衡	18
4.3 监管光谱的现实映射	20
5 结论与启示	21
5.1 核心结论	21
5.2 政策启示	21
5.3 分析的边界	22
参考文献	23

1 引言：AI 冲击知识型工作的两种对立叙事

讨论 AI 与工作的关系，不能只停留在“谁替代谁”的零和叙事中。AI 的真正意义，不应是压低人的价值，而应是扩展人的能力、释放人的创造力，并推动组织重新设计更有尊严、更有成长性的工作。问题的关键不在于是否拥抱 AI，而在于如何让 AI 的扩散过程更加包容：既释放新技术的生产潜力，也避免把转型成本过度压在处于职业起步阶段和承担调整压力较大的劳动者身上。

1.1 现象：AI 能力的快速扩张

过去三年间，人工智能从“能对话”走到了“能干活”。2022 年底 ChatGPT 发布时，大模型的主要形态还是问答助手；2023 年 GPT-4、Claude 等新一代模型已经能够胜任写作、编程、翻译、数据分析等相当广泛的认知任务；2024 年以 OpenAI o1 为代表的推理模型将多步推理能力推上新台阶；进入 2025 年，能够自主拆解任务、调用工具、连续工作数小时的“智能体”（Agent）密集落地，业界普遍将这一年称为“智能体元年”。一个直观的度量是：METR 的持续评测发现，AI 能够可靠完成的任务时长呈指数增长，其对 2023 年以来前沿模型的最新测算显示增速已加快至约四个月翻一番（以人类专业人员完成同一任务所需时间衡量）——从几分钟的代码修补，扩展到以小时计的完整工程任务。

这一能力扩张的经济含义在于：AI 正在进入以往自动化浪潮从未触及的领域。历史上的机械化、信息化主要替代程式化、可编码的任务，而知识型工作——依靠专业知识判断和解决非常规问题的的工作，如法律分析、金融研究、软件开发、管理咨询——长期被视为自动化的“安全区”。大模型改变了这一格局：知识一旦被编码进模型，就可以在任意规模的算力上同时复用，不再受专家个人时间的约束。据高盛（2023）估计，美国约三分之二的职业至少部分暴露于 AI 自动化；Eloundou 等（2024）发表于《科学》的研究进一步测算，约 46% 的岗位可能有超过一半的任务受到大模型影响。知识型工作恰恰是暴露度最高的部分。

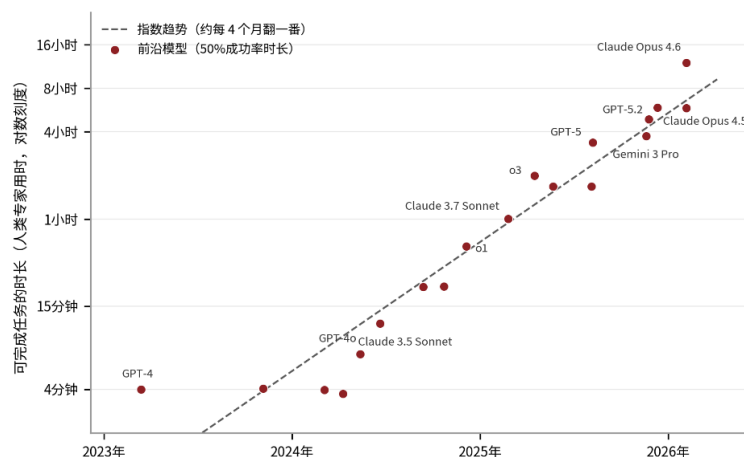


图 1 AI 能力的快速扩张：2022—2026

资料来源：METR 时长评测公开数据（Time Horizon 1.1，截至 2026 年 2 月），作者按 METR 官方方法拟合绘制。时长指模型以 50% 概率完成任务所对应的人类专家用时。

1.2 矛盾：两组相互冲突的经验证据

AI 究竟在冲击谁？围绕这一问题，目前存在两组结论相反、但各自都有扎实数据支撑的证据。

第一组证据来自受控实验和现场实验，结论指向“AI 辅助工具更明显支持经验较少者，并缩小组内差距”。Brynjolfsson 等（2025）对一家财富 500 强企业客服中心的研究发现，配备 AI 辅助工具后，新手客服的生产率提升幅度最大，而资深客服几乎不受影响；Noy 和 Zhang（2023）发表于《科学》的写作实验显示，原本成绩较差的参与者使用 ChatGPT 后进步更明显，组内绩效差距显著收窄；Dell'Acqua 等（2023）针对波士顿咨询公司数百名顾问的实验同样发现，能力低于平均水平的顾问从 AI 中获得的提升（约 43%）明显大于高水平顾问（约 17%）。跨越客服、写作、编程、咨询等多个场景，实验证据高度一致：AI 辅助工具具有提高效率、缩小组内差距的潜力。

第二组证据来自真实劳动力市场，结论则指向另一种机制：当 AI 能够独立承担部分任务时，入门岗位更容易被压缩，而高专业积累岗位的相对需求上升（CEISD，2026）。Berger 等（2024）利用美国高频招聘数据发现，ChatGPT 发布后，基础性白领岗位（如初级文员）的招聘量明显下降，而管理与专家职位等高专业积累岗位的招聘量不降反升；法律科技公司 Litera（2022）对逾 200 名并购律师的调查显示，AI 接管文书审阅等例行工作后，初级律师的传统“练手”机会正在减少；Beane 和 Anthony（2024）对投资银行的田野研究也记录了类似现象：资深合伙人越来越多地直接使用 AI 完成过去交给初级分析师的工作，新人的核心业务参与机会受到压缩。在企业的真

实用人决策中，AI 看起来是一种接管部分基础性任务、同时互补高专业能力的技术。

两组证据不可能通过“谁的数据更好”来裁决——它们测量的对象本来就不同：实验研究考察的是给定岗位上配备 AI 辅助工具的效果，而招聘数据反映的是企业在 AI 可以独立完成部分工作时的组织重构。要把二者放进同一个逻辑框架，需要一个能够同时刻画“AI 辅助人”与“AI 接管任务”、并允许其后果随 AI 能力变化的理论。

1.3 本报告的分析路径

基于上述现实，Ide 和 Talamas（2025）将 AI 引入劳动经济学中经典的“知识层级”模型，系统刻画了不同能力、不同自主程度的 AI 进入生产组织后，职业选择、企业形态与收入分配的均衡变化¹。该框架的独特价值在于：它不把 AI 笼统地视为一种“技术进步”，而是明确刻画 AI 的两个关键属性——能力水平（能解决多难的问题）和自主程度（能否独立工作），并证明这两个属性共同决定了“谁受损、谁受益”。

本报告属于 Discussion Paper 系列，定位不是评述论文本身，而是把论文的理论当作分析工具来使用：我们在原模型设定下自行选取参数、求解均衡，将 AI 能力从低到高的整个区间完整推演一遍，把论文中分散于各命题的结论串成一条“能力时间线”，再对照现实中正在发生的信号，回答三个问题：其一，当前观察到的对立证据处于演化序列的哪个阶段；其二，随着 AI 能力继续提升，哪些趋势会延续、哪些会翻转；其三，AI 自主性的监管选择将如何改变收益分配的格局。

下文结构安排如下：第二章梳理知识型工作领域正在发生的三个典型事实；第三章介绍理论框架并给出完整推演，这是报告的主体；第四章讨论 AI 自主性的分配后果及其监管含义；第五章总结全文并给出政策启示。

2 事实：知识型工作正在发出的三个信号

在进入模型之前，本章先把需要解释的现象摆出来。我们将其归纳为三个信号：入门岗位收缩、AI 使用收益不均等、“小团队、大产出”企业出现。三者分别对应劳动力市场的需求端、生产率的微观证据和企业组织形态，共同构成检验任何 AI 劳动理论的“事实清单”。

¹该文属于 Garicano（2000）开创的知识层级研究传统分析范式。本报告全部模型图形均由作者在原模型设定下自行参数化求解绘制，非论文原图；文中转述的理论结论均以论文正式命题为准。

2.1 信号一：入门岗位收缩

第一个信号来自劳动力市场的需求端：知识密集型行业的入门级岗位正在系统性收缩。在海外，这一趋势已有较为完整的数据刻画。除前述 Berger 等（2024）基于招聘广告的证据外，多家研究机构对美国应届毕业生就业市场的追踪均显示，2023 年以来软件开发、金融分析、法律服务等领域的初级职位发布量降幅显著大于同行业的资深职位；部分头部科技企业和咨询公司公开缩减了校园招聘规模，并明确将“AI 承担初级工作”列为原因之一。

互联网、金融等行业校园招聘规模的调整，以及招聘平台上初级白领岗位供需关系的变化，都指向入门通道收窄的方向。值得强调的是，入门岗位收缩并不等于总用工量下降：在多数观察中，同期高级职位的需求是稳定甚至上升的。换言之，变化的重心不是“要不要人”，而是“要什么层级的人”——这正是后文理论框架所要解释的组织重构。以美国最新的 ADP 高频工资单数据为例，2022 年 10 月至 2025 年 9 月间，22-25 岁入门劳动者在 AI 高暴露岗位上的就业量下降约 6%，是所有年龄×暴露度分组中唯一收缩的一组；同一年龄段在低暴露岗位仍增长约 8.5%（图 2）。

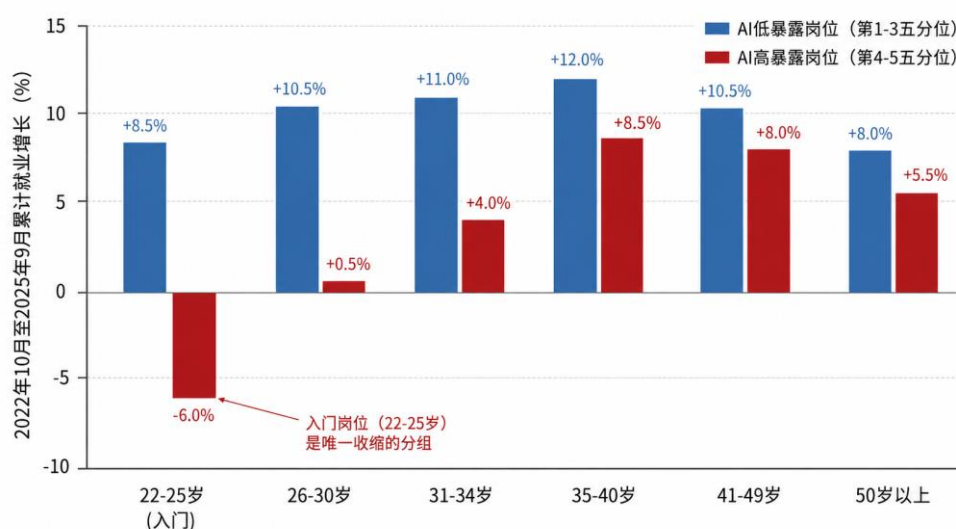


图 2 知识密集型行业初级岗位与资深岗位招聘量对比

资料来源：Brynjolfsson, Chandar & Chen (2025), Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence, Stanford Digital Economy Lab。基于 ADP 工资单数据构建；AI 暴露度按 Eloundou 等（2024）度量的五分位分组，第 4-5 五分位为高暴露、第 1-3 五分位为低暴露。

2.2 信号二：AI 使用收益的不均等

第二个信号来自微观生产率证据：同样使用 AI，不同经验与专业积累的劳动者获益差异明显，且方向一致——经验或专业积累越少，获益越大。表现在数据上，是引

入 AI 辅助工具后组内绩效分布的整体压缩：客服场景中新手的问题解决速度向老手靠拢（Brynjolfsson 等，2025）；写作任务中原本表现较低组与较高组的差距收窄（Noy 和 Zhang，2023）；编程场景中，Peng 等（2023）发现使用 GitHub Copilot 的开发者完成任务的速度平均快约 56%，且经验较浅者受益更明显；咨询场景中，原本低于平均水平的顾问获得了两倍于高水平顾问的绩效改善（Dell'Acqua 等，2023）。

这组证据在传播中常被概括为“AI 具有缩小差距的潜力”。但需要注意其共同的适用范围：上述研究考察的几乎全部是“辅助型”AI——AI 以副驾（copilot）形态嵌入既有岗位，为在岗者提供建议和参考，最终决策与执行仍由人完成。它们回答的是“给定组织结构不变，AI 辅助谁的效果更好”，而没有回答“当 AI 足以独立完成部分工作时，组织会如何重新配置岗位”。后一个问题正是信号一所反映的现实。两个信号并置，提示“辅助”与“自主执行”可能对应截然不同的分配后果——这一直觉将在第四章得到理论确认。

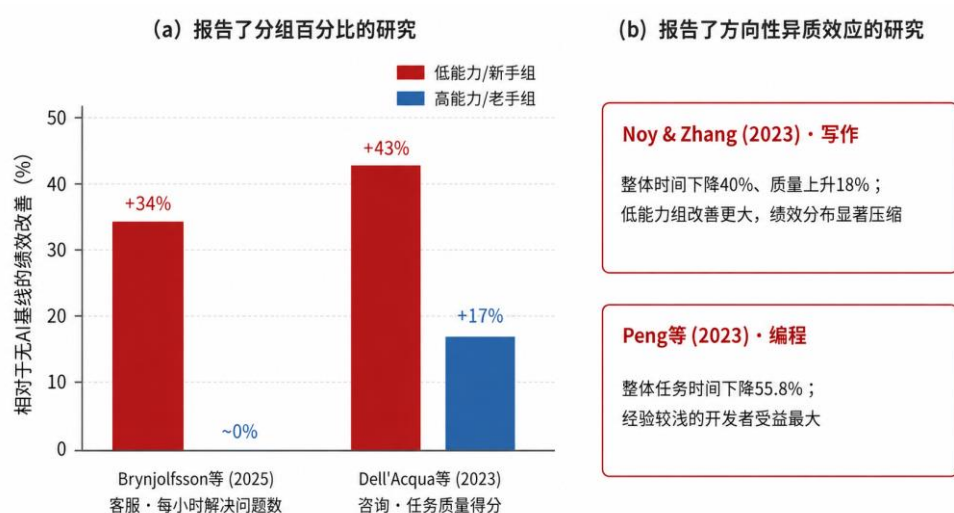


图 3 AI 辅助工具的生产率提升情况

资料来源：（a）Brynjolfsson, Li & Raymond (2025), *Generative AI at Work*, QJE 140(2); Dell'Acqua et al. (2023), *Navigating the Jagged Technological Frontier*, HBS WP 24-013。（b）Noy & Zhang (2023), *Science* 381(6654); Peng et al. (2023), arXiv:2302.06590。作者依据各论文正式发表版本的原文数字整理绘制。

2.3 信号三：“小团队、大产出”企业的出现

第三个信号来自企业组织形态：一批员工规模极小、产出或市场影响力却极大的企业正在出现，且集中于 AI 应用最深的领域。图像生成公司 Midjourney 在员工仅约 40 人时，据报道年收入已达数亿美元；DeepSeek 以百余人规模的研发团队推出性能对标国际头部模型的产品，引发全球关注；金融科技公司 Klarna 宣布其 AI 客服系统承担了

相当于 700 名人工客服的工作量。这类企业的共同特征是：极少数高水平专家搭配大规模算力，中间层级和基础执行层高度精简。

传统企业理论中，产出规模与雇员规模大体同步扩张，“有规模、无人手”（scale without mass）的形态历史上仅零星出现于平台型企业。而在大模型时代，这种形态呈现出向一般知识生产领域扩散的趋势²。它对既有的企业规模分布、劳动收入份额乃至反垄断分析都构成新问题：如果最有生产力的组织形式不再需要多少雇员，增长的收益将通过什么渠道分配给劳动者？

表 1 “小团队、大产出”典型案例（初步整理，数据待核实更新）

公司名称	员工规模	产出/估值指标
Midjourney	约 40—60 人（公开估计口径存在差异）	2025 年收入/ARR 公开估计约 3 亿—5 亿美元，显示极高人均收入。
DeepSeek	约 160 人（媒体报道口径）	DeepSeek-V3 为 671B 总参数、37B 激活参数模型；技术报告称完整训练约 2.788M H800 GPU 小时，按 2 美元/GPU 小时折算约 557.6 万美元。该数值不等于公司全部研发与硬件成本。
Cursor（Anysphere）	300 人以上（2025 年报道口径）	2025 年 11 月完成 23 亿美元 D 轮融资，投后估值 293 亿美元；多家媒体报道其年化收入已超过 10 亿美元。
Klarna（AI 客服）	公司员工约 3,800 人（2024 年 8 月报道口径）；AI 客服相当于约 700 名全职客服工作量	AI 客服上线首月处理约三分之二客服聊天，平均解决时间由 11 分钟降至 2 分钟；公司预计 2024 年利润改善约 4,000 万美元。

资料来源：作者根据公司公告、技术报告及公开媒体报道整理。由于部分私营公司未披露经审计收入或员工数，表中涉及收入、估值、员工规模的信息均按公开报道口径列示；DeepSeek 训练成本仅指技术报告披露的模型训练算力折算成本，不代表公司全部研发、硬件和运营投入。

2.4 现有解释的缺口

三个信号放在一起，对理论提出了不低的要求。流行的解释框架大致有二。其一是任务模型传统（Acemoglu 和 Restrepo，2018）：技术接管部分任务、创造新任务，劳动者在任务间重新配置。该框架刻画“哪些任务被接管”十分有力，但企业在其中近似一个任务集合，难以解释信号一中“初级岗位收缩而高级岗位扩张”的组织层级变化，也难以解释信号三的企业形态变异。其二是“技能偏向型技术进步”的直觉：AI 接管部分基础性任务、互补高专业能力。它与信号一吻合，却与信号二正面冲突——实验数据显示经验较少者获益更大。

缺口的根源在于：多数框架把 AI 当作一种同质的“技术冲击”，没有区分 AI 的能力水平，更没有区分 AI 是以“辅助”还是“自主执行”的形态进入生产。而现实中，这两个维度恰恰在快速变化——模型能力每隔数月上一个台阶，产品形态从副驾工具

²本段企业案例数据来自公开报道的初步整理，员工规模与收入数字均为报道口径，正式稿将逐一核实并在表 1 中标注资料来源。

演进为自主智能体。一个合格的分析框架，必须让“AI 影响工作的方向和分配后果”成为 AI 能力与自主性的函数，而非一次性的定性判断。下一章引入的知识层级框架，正是为此而生。

3 框架与推演：AI 能力提升过程中知识型工作的重组

3.1 理论骨架：知识分工、层级与 AI 的进入方式

Ide 和 Talamas（2025）的模型建立在 Garicano（2000）开创的知识层级传统之上。这一传统对知识型生产的刻画可以用三句话概括。第一，生产的本质是解决问题：每个人在工作中不断遇到难度不一的问题，能解决就有产出，解决不了就没有。第二，人与人的差别在于知识水平：知识水平决定了一个人能独立解决的问题范围。第三，时间是最稀缺的资源：知识再高的专家，一天也只有二十四小时。

在这三个前提下，分工与企业会自发形成：让多数人在一线从事常规生产，遇到解决不了的难题时向少数知识更高的人求助——前者可称为“生产者”，后者可称为“解题者”。求助要占用解题者的时间，因此一名解题者只能支撑有限规模的团队，这就内生出了企业的层级结构和管理幅度。均衡中，知识水平较低的人担任生产者，知识水平较高的人担任解题者，且越优秀的解题者带越大的团队、匹配越优秀的生产者。通用汽车前掌门斯隆的名言“例行事务到不了我这里，我只处理例外”，正是这一结构的经典写照。

AI 如何进入这个世界？论文抓住现代 AI 区别于以往技术的关键特征：知识一旦编码进模型，就可以在全部可用算力上无限复制——AI 相当于“可以随意复印的劳动者”。据此，模型把 AI 刻画为两个参数：一是能力水平，即 AI 能独立解决的问题范围，对应现实中模型能力的持续提升；二是自主程度，自主的 AI 既能独立承担一线生产（作为“数字员工”），也能为人答疑（作为“顾问”），而非自主的 AI 只能担任顾问角色。企业可以租用算力，用 AI 替换任一层级上的人。本章先分析自主 AI 随能力提升的完整后果，第四章再讨论自主与非自主的对比³。

推演方法上，本报告在论文的基准设定下（知识水平在人群中均匀分布、求助成本取论文基准值）自行编程求解竞争均衡，将 AI 能力参数从零到一的整个区间扫描一

³模型的其他设定还包括：全社会共享同一前沿模型；算力可按需租用且相对人力时间充裕；企业最多两个层级。论文证明这些简化不改变主要结论的方向。

遍，绘制职业结构、工资分布与企业形态的连续变化图。需要说明的是：下文全部模型图形均为本报告自行求解绘制，并非论文原图；推演不改动原模型的任何假设，图形的具体形状依赖参数取值，但本章引述的定性结论均为论文正式证明的命题，不随参数改变。

3.2 职业版图的演化与“大翻转”

推演的第一个结果，是 AI 能力提升过程中职业结构的完整演化图景。以横轴表示 AI 能力、纵轴表示劳动者的知识水平，可以画出一张“职业版图”：给定 AI 能力，不同知识层级的劳动者分别处于一线生产、专业化解题以及与 AI 配合完成任务等不同位置。图中蓝色区域表示一线生产者，红色区域表示解题者；斜线阴影区域则表示与 AI 形成直接配合的人类劳动者，其中上方阴影部分是“匹配 AI 的解题者（自主 AI 辅助生产）”，下方阴影部分是“匹配 AI 的生产者（自主 AI 顾问支持）”。这张图的核心信息，是 AI 能力提升会引发一次方向相反的两阶段变化。

第一阶段，AI 能力较低，相当于均衡中一名普通生产者的水平，论文将其称为“基础型”自主 AI。此时，企业首先发现，用 AI 从事一线生产比雇佣人类更便宜，因此大量算力率先进入生产环节。图中可以看到，在这一阶段，知识分布高位出现了“匹配 AI 的解题者（自主 AI 辅助生产）”区域，含义是最优秀的一批解题者开始转而支持 AI 完成生产任务；与此同时，原本知识水平较高的一部分一线生产者，因 AI 接管部分任务而退出原有生产岗位，并转向解题者角色。因此，随着 AI 能力上升，一线生产者区域逐步收缩，解题者区域相应扩张，但其平均质量被稀释，处于较低位置的人类生产者只能从经验和知识积累相对有限的解题者那里获得帮助。直观地说，基础型 AI 推动部分劳动者向解题岗位转移，但这种“晋升”并不意味着福利改善，后文的工资分析将进一步说明这一点。

第二阶段，当 AI 能力越过生产者与解题者之间的分界线之后⁴，职业结构发生翻转，论文将其称为“先进型”自主 AI。此时，AI 不仅是高效的生产者，也开始成为高质量、低成本的顾问。图中新增的下方阴影区域“匹配 AI 的生产者（自主 AI 顾问支持）”表明，处于知识阶梯较低位置、并能熟练调用技术的人类劳动者开始直接借助 AI 完成原本需要向人类解题者请教的问题；相应地，一部分原先承担基础解题任务的

⁴ 该分界线是指如果没有 AI，一个员工被划分为一线生产者和解题者之间的知识水平分界线。如低于该水平则为一线生产者，高于则为解题者。该分界线同样会导致 AI 影响产生翻转的临界点。

人类劳动者，其原有比较优势下降，并转向一线生产环节。因此，在这一阶段，一线生产者区域重新扩大，而解题者区域相对收缩；但与前 AI 时代不同的是，这些生产者背后已不再主要依赖人类解题者，而是由 AI 提供顾问支持。图中的竖虚线所标示的临界点，恰好对应前 AI 时代生产者与解题者的分界线，这也意味着，判断现实究竟处于哪个阶段，关键就在于比较当前 AI 的能力与现实组织中中层解题者的能力高低。

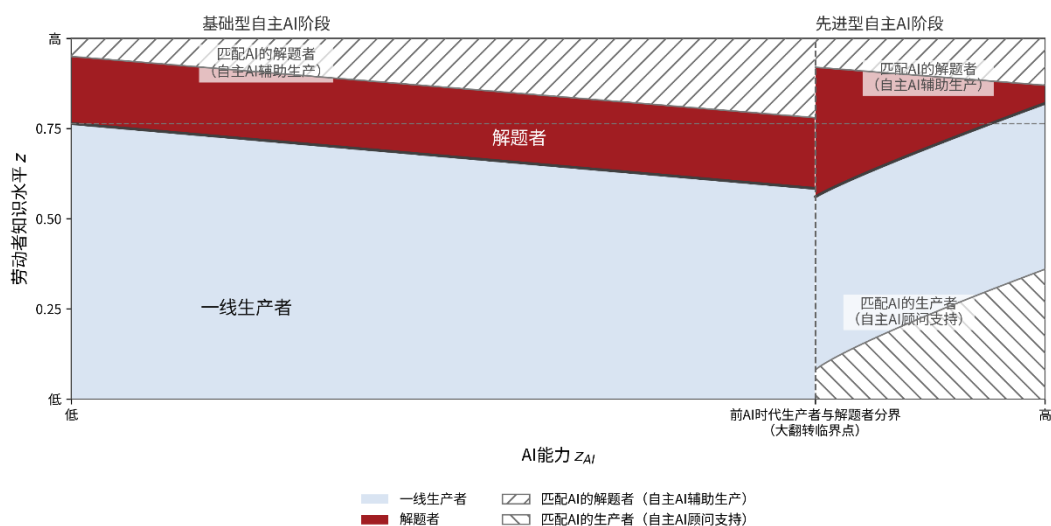


图4 职业版图随 AI 能力提升的演化与“大翻转”

资料来源：作者依据 Ide 和 Talamas（2025）知识层级模型整理绘制。在保持论文基准设定 $G(z) = z$ 、帮助成本 $h = 1/2$ 不变的基础上，将论文中基础型自主 AI 与先进型自主 AI 两种情形下的职业分层结果整合为连续职业版图。其中，匹配 AI 的解题者（自主 AI 辅助生产）和匹配 AI 的生产者（自主 AI 顾问支持）分别对应论文中的 S_a 与 W_a 类型。

3.3 三类劳动者的命运曲线

职业版图刻画“人在哪里”，工资曲线刻画“人得到什么”。沿 AI 能力轴追踪三类代表性劳动者——处于职业起步阶段的“新人”、位于中间区间的“骨干”、接近知识分布高位的“专家”——可以得到三条明显分化的工资轨迹。图 5 的横轴是 AI 能力，纵轴是相对于无 AI 基准的工资变化，阴影区表示工资低于无 AI 基准的调整压力区间。与图 4 刻画职业结构重组相对应，图 5 进一步展示了这种重组如何转化为不同知识层级劳动者的收入后果。

首先，专家的曲线最为平稳，也最接近论文的核心结论。根据论文的命题，只要 AI 尚未达到最顶尖人类的知识水平，知识分布高位始终存在一部分受益者。因此，在图 5 中，专家的工资曲线整体位于零轴之上，并随 AI 能力提升而继续上升。其机制在于，自主 AI 的可复制性放大了稀缺知识的杠杆：在基础型自主 AI 阶段，顶尖知识者可以利用廉价 AI 生产者替代部分人类团队，从而以更低成本扩展自身知识的应用范围；

进入先进型自主 AI 阶段后，他们又继续受益于 AI 承担更多例行性或辅助性任务。

其次，新人的曲线呈现“先承压、后改善”的特征。基础型自主 AI 阶段，处于职业起步阶段的劳动者一方面在一线生产环节面对 AI 接管部分任务的压力，另一方面其可获得的人类解题支持又因解题者队伍扩张而被稀释，因此工资往往低于无 AI 基准。随着 AI 能力继续提升，特别是越过关键门槛之后，情况发生逆转：这些劳动者可以以较低成本获得高质量 AI 顾问支持，从而处理超出自身既有经验的问题，工资开始回升并可能转为正增长。论文对此给出的严格结论是：只有当 AI 能力高于某一门槛时，知识分布低位区间才会出现受益者；而一旦 AI 能力足够高，这部分受益者的范围会进一步扩大。

最后，骨干的曲线最为颠簸，也最值得政策关注。图 5 中的“骨干”指的是位于知识分布中间区间的代表性劳动者。按照论文结果，AI 带来的收益主要出现在知识分布两端，而中间区间往往更容易承受持续调整压力。因此，图中的骨干曲线大体处于零轴以下：在基础型自主 AI 阶段，他们可能在一线生产中受到 AI 接管部分任务的影响，也可能转向质量较低的解题岗位；在先进型自主 AI 阶段，他们又更容易受到 AI 顾问的替代压力。尤其值得强调的是，知识水平恰好与 AI 相当的劳动者在模型中承受较高调整压力，因为 AI 构成了其能力的可复制替代品。也正因如此，“与 AI 同水平”可以理解为模型中的高风险转型位置，而随着 AI 能力持续上移，这一压力位置也会沿着知识阶梯不断上移。

总的来看，图 5 传达的不是“AI 让所有人同步受益或同步受损”，而是一个需要主动治理的分化过程：高专业积累者更容易率先受益，处于职业起步阶段的劳动者在 AI 足够强之后也可能获得改善，而中间层劳动者则更可能承受转型成本。这也意味着，AI 的收入效应并不只是技术替代问题，更是组织分工重组、知识价值再定价和社会支持体系能否及时跟上的问题。

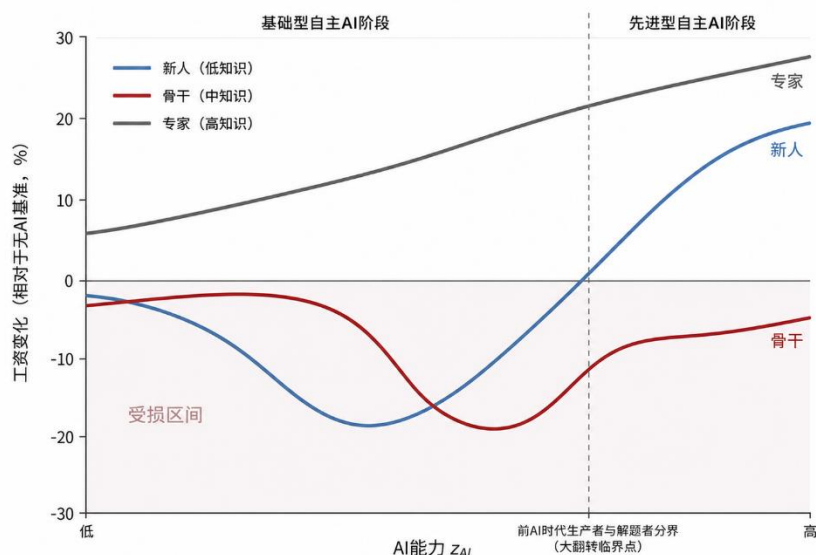


图5 新人、骨干与专家：工资随AI能力提升的分化轨迹

资料来源：作者依据 Ide and Talamàs（2025）知识层级模型整理绘制。在保持论文基准设定 $G(z) = z$ 、 $h = 1/2$ 的前提下，结合论文 Figure 5、Proposition 5 以及关于匹配效应和收入分配效应的相关命题，将知识分布底部、中部和顶部的代表性个体分别概括为“新人”“骨干”和“专家”。图中纵轴百分比主要用于表现方向和相对变化，并非论文原始数值模拟结果的逐点复刻，而是依据论文关于工资分配效应的核心命题，对三类代表性劳动者所作的示意化整理。

3.4 不平等的阶段性：先扩大、后收敛

将全部个体的工资变化加总，可以刻画收入分配的整体走势。图6以高低知识分位工资比作为不平等的示意指标，展示了AI能力提升过程中工资差距的阶段性变化。需要说明的是，图中的临界点不是一个固定数值，而是指AI能力从前AI时代生产者区间进入解题者区间的分界。其具体位置取决于知识分布、帮助成本等参数设定，因此图中仅标示其经济含义，而不标注具体数值。

在基础型自主AI阶段，AI更像一种“基础型数字员工”，首先进入一线生产环节，并接管部分标准化、可重复的初级任务。由于高专业积累者可以利用AI放大自身知识，而职业起步阶段劳动者仍面临任务调整和匹配质量下降的双重压力，工资差距随AI能力提升而扩大。图6左侧曲线的上升部分对应这一机制，也更接近当前基于招聘数据得到的结论：AI作为基础型数字员工时，入门岗位更容易受到压缩。论文也指出，基础型自主AI会使知识积累较少的劳动者承受更大调整压力，而高专业积累者能够利用AI以较低成本放大自身专业能力。

当AI能力越过前AI时代生产者与解题者之间的分界后，进入先进型自主AI阶段。此时AI不仅能从事生产，也能作为顾问帮助基础岗位劳动者解决更复杂的问题，因此职业起步阶段劳动者开始分享AI红利，工资差距的扩张放缓，并出现一定程度的收敛。

图 6 右侧曲线的下降部分对应这一过程，也更接近通过实验数据得到的结论：当 AI 以顾问形态介入工作时，经验较少者往往获得更明显的生产率提升。与此同时，曲线并未回到无 AI 基准，表明即便职业起步阶段劳动者开始受益，自主 AI 的收益分配仍然偏向知识分布高位；只要 AI 尚未达到人类知识的绝对上限，顶尖知识者仍能通过 AI 的可复制性持续放大自身优势。

因此，图 6 对前文两组经验证据提供了一个统一解释：招聘数据中看到的“入门岗位受压缩”，更接近基础型自主 AI 作为数字员工进入生产环节的情形；实验研究中看到的“经验较少者受益”，更接近 AI 作为顾问嵌入工作流的情形。二者并不矛盾，而是 AI 能力与 AI 自主性处于不同位置时的不同表现。真正需要警惕的不是 AI 本身，而是不能把实验中“AI 缩小差距”的结论简单外推到完全自主 AI 主导的未来；当 AI 从工具进一步成为可以独立完成的工作伙伴，收入分配逻辑会随组织结构一起发生改变。

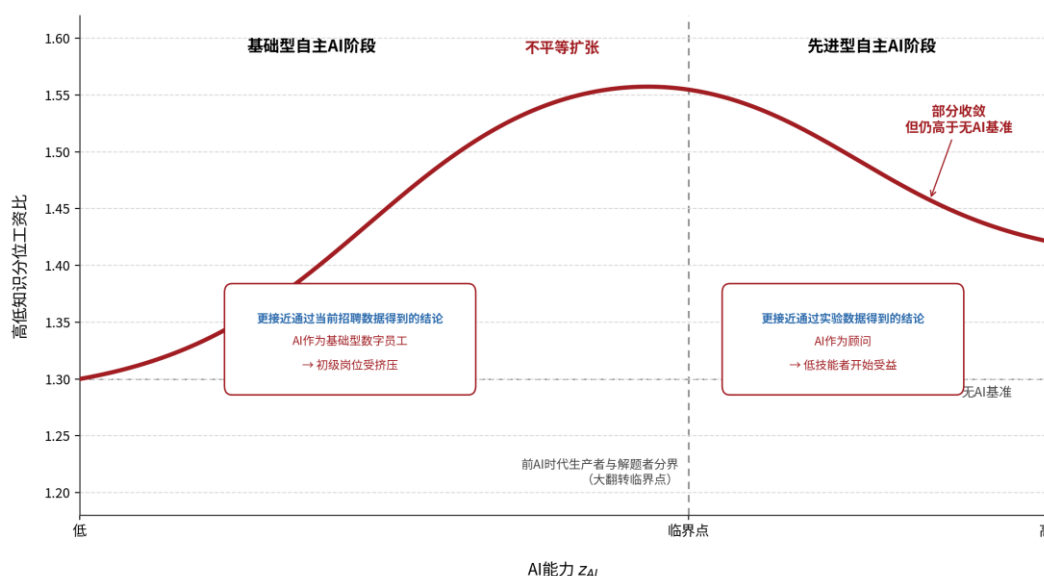


图 6 不平等的阶段性：工资差距随 AI 能力的先扩大、后收敛

资料来源：作者依据 Ide 和 Talamas（2025）知识层级模型整理绘制。图中曲线为基于论文关于自主 AI、非自主 AI 及收入分配效应的命题所作的示意化表达，重点刻画工资差距随 AI 能力提升的阶段性变化，并非论文原始数值模拟结果的逐点复刻。图中临界点表示 AI 能力从前 AI 时代生产者区间进入解题者区间的分界，其具体位置依赖模型参数设定。

3.5 企业形态的重塑：组织收缩与“超级小企业”

模型中企业结构是内生决定的，因此 AI 冲击不仅改变劳动者的职业位置，也会同步改变企业形态。图 7 以人类雇员规模为横轴，示意 AI 进入前后企业分布的变化：灰色虚线表示无 AI 基准，蓝色实线表示 AI 进入后的新分布，浅蓝色区域表示新生或扩

张的组织类型，浅红色区域表示收缩或消亡的组织空间。图的核心含义不是所有企业都会单向变大或变小，而是 AI 会重塑不同规模企业的相对优势，并使企业形态在两个阶段呈现方向相反的变化。

在基础型自主 AI 阶段，AI 首先作为低成本生产者进入一线生产环节。由于最顶尖的解题者转向支持 AI 生产，依赖人类协作的企业面临解题者队伍被稀释、匹配质量下降的问题，原有中大型组织的优势因此受到削弱。图 7（a）显示，在这一阶段，中大型组织空间收缩，而小企业数量相对增多。但这里增加的小企业，并不是“超级企业”，而更多是小规模、低管理幅度的小型组织。换言之，基础型 AI 主要冲击的是原有依赖多层人类协作的大型、高分权组织，而新增的小企业并不一定更强，反而往往反映出组织能力的阶段性下降。

进入先进型自主 AI 阶段后，企业形态发生翻转。AI 不仅能够承担生产任务，还能够作为顾问支持一线工作者解决更多问题，使得企业的管理幅度重新扩张，部分工作可以在更少人类雇员的情况下完成。图 7（b）显示，在这一阶段，少人化企业明显增多，中型组织空间则受到挤压。这里的“小”不再意味着规模小、层级少，而是意味着少量高专业积累人才配合大规模 AI 能力即可支撑强大的生产活动。因此，这一阶段出现的，是“少人但强”的新型企业形态：它们对庞大人类团队的依赖下降，却具备更强的组织能力和更高的扩张潜力。

这一机制也为第二章“小团队、大产出”的现实案例提供了理论解释。Midjourney、DeepSeek、Cursor、Klarna AI 客服等案例的共同特征，不是企业简单变小，而是少量高专业积累人才与大规模 AI 能力结合，使过去需要通过雇佣、培训和层级管理才能扩张的知识，开始通过模型和算力实现复制。由此，AI 对企业形态的影响不是简单压缩，而是结构性重塑：传统中型组织的空间受到挤压，而分布左端则可能持续出现“少人化、高能力”的新型企业。

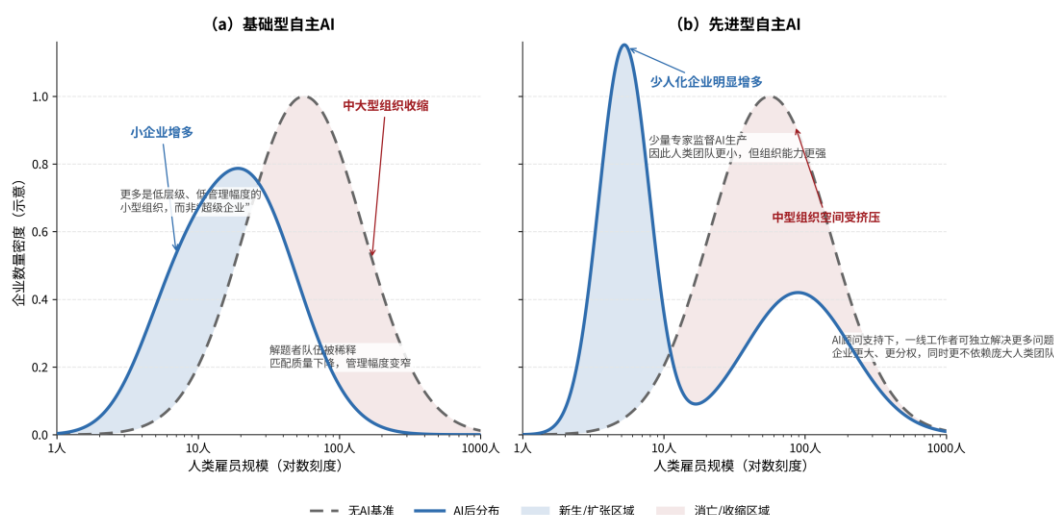


图 7 企业规模分布的重塑与“超级小企业”的出现

资料来源：作者依据 Ide 和 Talamas（2025）主文及在线附录第 4.1.2 节整理绘制。本图是在论文关于 AI 重塑企业结构、企业规模分布和管理幅度变化的分析基础上，结合本报告第二章案例所作的示意化表达。

4 延伸：AI 自主性的分配后果

4.1 “顾问”与“员工”：两种形态、两种格局

第三章的推演假定 AI 是自主的，即既能独立生产，也能提供咨询。但现实中大量 AI 应用并不具备这样的自主性：出于技术或监管原因，它们往往只能以“副驾”或“顾问”的形态嵌入人的工作流，为人提供建议，却不能独立交付任务。论文对此作了专门比较，并给出了与自主 AI 截然不同的分配预测。图 8 展示的正是在同等能力下，无 AI、自主 AI 和非自主 AI 三种情形的工资分布对比。其核心结论是：当 AI 只能充当顾问时，最大的受益者不再主要是高专业积累者，而转为处于职业起步阶段、原本较难获得高质量指导的劳动者；相反，当 AI 既能生产又能咨询时，受益更明显的则是高专业积累者。

其背后的机制有两层。第一，非自主 AI 不能独立进入生产环节，因此基础岗位劳动者不会像在自主 AI 情形下那样，在一线直接面对 AI 接管任务的压力。第二，非自主 AI 的机会成本更低：它不能独立创造产出，因此其咨询价格更低，职业起步阶段劳动者可以以较低成本获得高水平建议，把 AI 创造的红利更多留在自己手中。图 8 左侧蓝色阴影区正对应这一逻辑：在知识分布低位区间，非自主 AI 曲线位于自主 AI 和无 AI 基准之上，说明副驾式 AI 更有利于需要更多工作支持的劳动者。

相比之下，自主 AI 下的受益格局几乎倒转。由于自主 AI 既能充当顾问，又能直接从事生产，高专业积累者能够利用它接管例行工作、放大自身知识杠杆，因此在知

识分布高位，自主 AI 曲线明显高于非自主 AI 曲线。图 8 右侧红色阴影区所示的，正是“高知识区间：自主 AI 受益更大”的分配格局。与此同时，非自主 AI 对高专业积累者的帮助相对有限，因为它不能替他们独立完成生产任务，反而可能在咨询市场上与他们形成竞争。

这一定理性的对比，也为第二章信号二所展示的经验事实提供了直接解释。实验研究中的 AI，本质上几乎都是“副驾驶型”或“顾问型”AI，因此往往表现出“更明显支持经验较少者”的特征；而招聘市场和组织调整中所反映出的，则更接近 AI 作为“数字员工”进入生产流程后的效果。两类证据并不矛盾，它们对应的是两种不同 AI 形态的分配逻辑。政策上的含义也因此非常明确：不能把当前副驾驶 AI 的“缩小差距”效果，简单外推到未来更高自主性的 AI 阶段，因为一旦 AI 从顾问进一步变成可独立完成工作任务的工作伙伴，受益格局就会系统性转向高专业积累者。

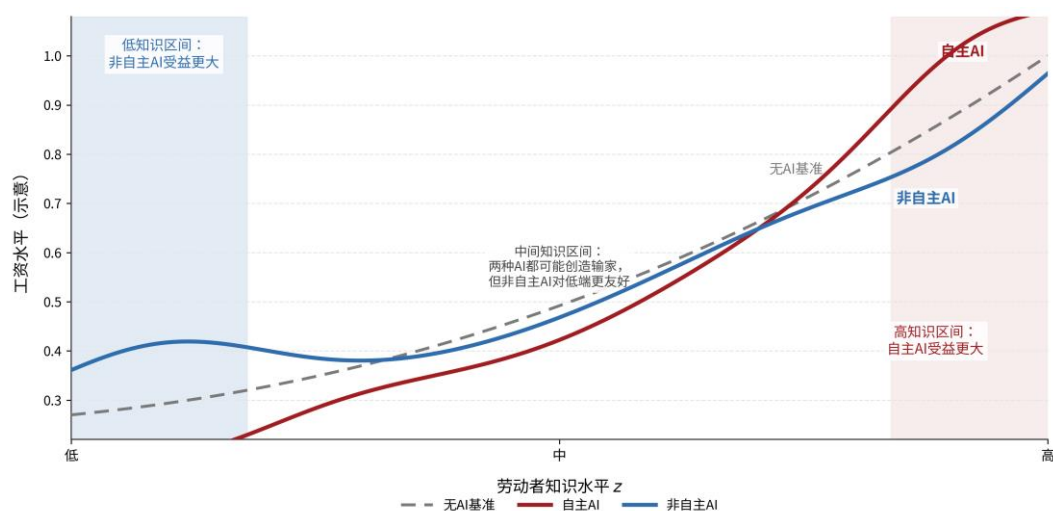


图 8 “顾问”与“员工”：两种 AI 形态下的受益格局对比

资料来源：作者依据 Ide 和 Talamas（2025）主文第 VI 部分、Proposition 6 和 Figure 9 整理绘制。论文主文直接比较了无 AI、自主 AI 与非自主 AI 三种情形下的工资分布；本图是在其理论结论基础上所作的重绘，并非论文原图复刻。

4.2 效率与公平的权衡

既然非自主 AI 更有利于普通劳动者，是否应当限制 AI 自主性？模型的回答是：可以，但有真实代价。对于同等能力的 AI，如果允许其自主生产，它既能充当“顾问”，也能充当“员工”，从而创造更高的总产出；如果限制其自主性，使其只能作为顾问嵌入人的工作流，那么普通劳动者的相对处境会改善，但整体产出会下降。换言之，限制自主性并不是没有成本的再分配工具，而是一种典型的效率—公平权衡。

图 9 将这一权衡表示为一个“监管可行集”。横轴表示工资不平等程度，纵轴表

示总产出水平。随着 AI 能力提高，自主 AI 和非自主 AI 都会推动经济向更高产出、更高不平等的方向移动，但两条路径并不重合，自主 AI 相对于非自主 AI 是右上方：产出更高，但不平等也更高；限制自主性则是从右上方向左下方移动，换来更平等，但牺牲产出。图中的纵向差距表示限制自主性的产出代价，横向差距表示限制自主性带来的再分配收益。

这一图示至少说明两点。第一，权衡是真实存在的，不存在既能最大化产出、又能自动改善分配的“免费午餐”。如果选择放开自主性，社会会得到更高的总产出，但收益会更多流向知识分布高位；如果选择限制自主性，普通劳动者受益会更多，但整体蛋糕会变小。第二，随着 AI 能力继续提高，这种权衡会变得更加尖锐：限制自主性的效率代价会上升，而其再分配收益未必同步扩大。因此，从模型含义看，自主性监管更适合作为转型期的缓冲工具，而不是长期依赖的根本性分配制度。

由此，智能体监管不能只被理解为安全问题，也应被看作一种分配政策选择。短期内，如果政策目标是缓冲职业起步阶段劳动者和承担较高调整压力人群受到的冲击，可以通过强化人类监督、限制关键环节的自主部署等方式，暂时压低 AI 自主性；但从长期看，如果 AI 能力持续上升，单纯依靠限制自主性来维持公平，将付出越来越高的产出代价。更稳妥的政策组合，仍应是在逐步释放 AI 产出潜力的同时，通过税收转移、社会保障、职业培训和教育体系调整来承接分配压力。

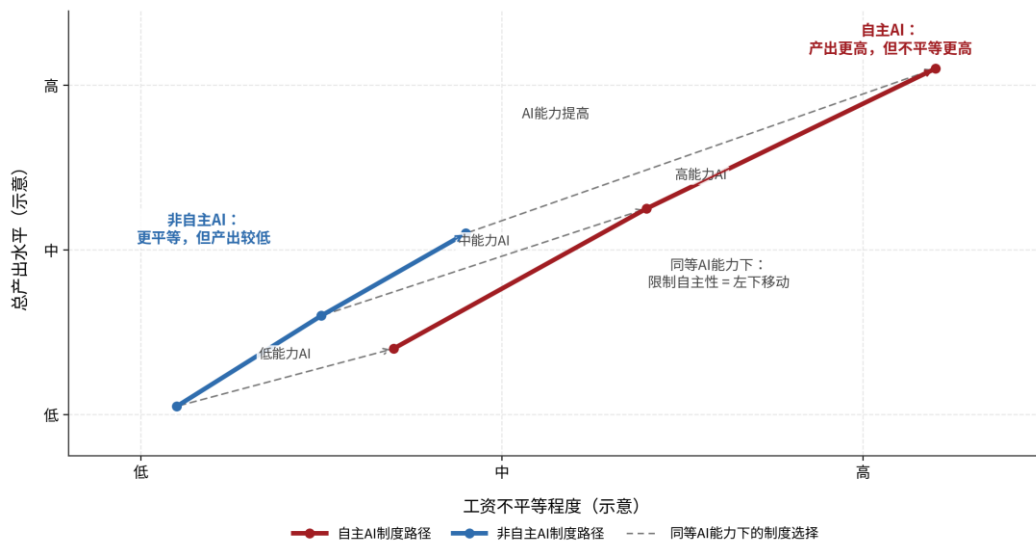


图9 自主性监管的效率—公平权衡

资料来源：作者依据 Ide and Talamàs（2025）主文第 VI 部分和 Proposition 6 整理绘制。论文表明，在相同 AI 能力下，自主 AI 的总产出高于非自主 AI，但不平等程度也会增加，具体见三条虚线箭头。本图据此将 AI 自主性监管概括出的示意图，非论文原图复刻。

4.3 监管光谱的现实映射

以“对 AI 自主性的实际约束强度”为标尺，可以将主要经济体的监管取向粗略放入前述效率—公平权衡框架之中。欧盟更接近约束较强的一端。《人工智能法案》采取风险分级监管路径，对高风险 AI 系统设置人类监督、透明度、风险管理等合规义务，这实际上抬高了 AI 在敏感场景中自主运行的制度门槛。就本报告的框架而言，欧盟的隐含取向更接近“以部分产出弹性换取安全、可问责性与分配稳定”。

美国则更接近约束较弱的一端。当前美国联邦层面的 AI 政策更强调创新、基础设施建设和产业领先，而不是建立统一的、专门针对智能体自主部署的事前许可制度。白宫《美国 AI 行动计划》明确强调移除阻碍 AI 发展和部署的繁重监管、加快私营部门采用，这意味着其政策重心更偏向释放自主 AI 的产出潜力。放在图 9 的权衡框架中，美国的隐含取向更接近“以更大的分配压力换取更高产出和技术领先”。

中国的位置更接近光谱中段，并呈现出“应用端鼓励、责任端收紧”的双轨特征。一方面，国务院关于深入实施“人工智能+”行动的意见明确鼓励人工智能与经济社会各行业深度融合，推动智能经济和智能社会新形态形成，这意味着产业政策层面对智能体落地和自主化应用持积极态度。另一方面，《生成式人工智能服务管理暂行办法》以及《人工智能生成合成内容标识办法》又对服务提供者设置了内容安全、标识、备案、处置等责任要求，从合规端提高了完全自主运行的责任成本。

以本报告的框架观察，中国的这种组合并不是简单地站在权衡曲线某一端，而是试图在“释放应用潜力”和“压实平台责任”之间寻找中间位置。其政策效果取决于责任规则最终是否会实质性抑制自主形态的部署：如果责任规则形成较强约束，那么分配上可能更有利于普通劳动者，但需要承担部分产出损失；如果责任规则主要发挥事后追责和风险控制作用，而不显著抑制自主 AI 扩散，那么就必须正视自主 AI 红利更可能向知识分布高端集中的再分配压力。

因此，智能体监管不能只被理解为安全治理问题，也应被理解为分配政策问题。欧盟、美国和中国的差异，表面上是监管强弱不同，本质上是对“产出效率—分配稳定—安全可控”三者权重的不同选择。对中国而言，更可行的政策方向不是简单复制欧盟的强约束，也不是完全采取美国式的弱约束，而是在鼓励智能体应用扩散的同时，把责任规则、劳动转型支持和再分配政策同步设计好。否则，一旦 AI 从“工具”进一步变成“员工”，由此带来的将不只是技术替代问题，而是企业组织、劳动收入和社会

会分配格局的系统性重塑。

5 结论与启示

5.1 核心结论

本报告借助知识层级理论，把 AI 冲击知识型工作的讨论从“是否替代”推进到“何时、影响谁、以何种方式”。核心结论可以概括为三句话。第一，方向取决于位置：AI 对岗位的影响，由 AI 能力相对于人群知识分布的位置决定——能力低于组织中层时，更多劳动者转向解题岗位；越过中层门槛后，AI 顾问又会支持更多一线生产者独立完成复杂任务。第二，证据并不矛盾：实验中观察到的“AI 更明显支持经验较少者”，对应非自主的副驾形态；招聘数据中观察到的“入门岗位压缩”，对应基础型自主形态，二者是同一演化过程中不同机制的表现。第三，自主性是分配开关：同等能力的 AI，作为“员工”时总产出更高、红利更容易集中于高专业积累者；作为“顾问”时更有利于普通劳动者，但总产出较低。效率与公平的权衡，内嵌于 AI 的技术形态本身。

5.2 政策启示

第一，识别并支持转型期的重点人群。模型提示，最需要政策关注的并不只是知识分布低位区间，而是“与 AI 能力相当”的中间层——当前阶段主要包括各行业的初级专业人员和承担标准化知识工作的劳动者。随着 AI 能力上移，这一转型压力位置也会移动。就业政策与培训资源应当从单纯“托底”转向“托底与护中层并重”，帮助初级专业人员向 AI 监督、AI 协作、复杂问题判断等新角色转型。

第二，为人才成长通道设计制度替代。入门岗位历来承担着在干中学、传帮带的人才培养功能；当 AI 接管入门工作，市场自发提供的培养机会将系统性不足，高端人才的供给反而可能在长期受损。可考虑的方向包括：以财政或税收工具分担企业保留培养性岗位的成本、在职业资格体系中嵌入 AI 协作实训环节、由高校和公共部门承接部分原由企业承担的实务训练职能。

第三，把分配视角纳入智能体监管。自主性规则不仅关乎安全，还直接决定 AI 红利在劳动者之间的分配。监管设计应当明示其在效率—公平权衡上的立场，并为再分配安排预留衔接：若选择放开自主性以争取产出与技术领先，则需同步强化税收转移与社会保障；若以责任规则实质限制自主形态，则应评估并公开其产出代价。

5.3 分析的边界

最后需要说明本推演的适用边界。其一，模型中劳动者的知识水平外生给定，未刻画学习与人力资本积累，因此关于培养通道的讨论只能在模型之外展开；其二，模型假定算力相对人力充裕、全社会共享单一前沿模型，若算力约束收紧或模型市场高度分割，结论需作相应调整；其三，两层组织的简化虽经论文证明不影响主要结论的方向，但现实中多层级组织的重构细节会更复杂；其四，模型是静态比较，“翻转”的现实节奏还取决于调整成本与制度摩擦，宜将本报告的时间线理解为方向性刻画而非时点预测。这些边界并不削弱核心结论，但提醒读者：本报告提供的是一副观察 AI 与工作关系的理论透镜，而非一份预测报表。

归根到底，AI 治理的目标不应是把技术挡在门外，也不应让市场竞争单独决定所有人的命运。更好的方向，是在积极拥抱 AI 的同时，把劳动者的尊严、成长通道和转型支持放在制度设计的中心。AI 可以成为提高生产率的工具，也可以成为扩展普通人能力的基础设施；关键在于，社会能否及时调整教育、培训、企业责任和再分配机制，让技术进步真正转化为更广泛、更包容的发展机会。

参考文献

- [1] Ide E, Talamàs E. Artificial Intelligence in the Knowledge Economy[J]. *Journal of Political Economy*, 2025, 133(12): 3762-3800.
- [2] Garicano L. Hierarchies and the Organization of Knowledge in Production[J]. *Journal of Political Economy*, 2000, 108(5): 874-904.
- [3] Garicano L, Rossi-Hansberg E. Organization and Inequality in a Knowledge Economy[J]. *The Quarterly Journal of Economics*, 2006, 121(4): 1383-1435.
- [4] Acemoglu D, Restrepo P. The Race Between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment[J]. *American Economic Review*, 2018, 108(6): 1488-1542.
- [5] Brynjolfsson E, Li D, Raymond L. Generative AI at Work[J]. *The Quarterly Journal of Economics*, 2025, 140(2): 889-942.
- [6] Noy S, Zhang W. Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence[J]. *Science*, 2023, 381(6654): 187-192.
- [7] Dell'Acqua F, McFowland III E, Mollick E R, et al. Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality[R]. *Harvard Business School Working Paper No. 24-013*, 2023.
- [8] Peng S, Kalliamvakou E, Cihon P, et al. The Impact of AI on Developer Productivity: Evidence from GitHub Copilot[EB/OL]. *arXiv: 2302.06590*, 2023.
- [9] Berger P G, Cai W, Qiu L, et al. Employer and Employee Responses to Generative AI: Early Evidence[R]. *SSRN Working Paper No. 4874061*, 2024.
- [10] Beane M, Anthony C. Inverted Apprenticeship: How Senior Occupational Members Develop Practical Expertise and Preserve Their Position When New Technologies Arrive[J]. *Organization Science*, 2024, 35(2): 405-431.
- [11] Eloundou T, Manning S, Mishkin P, et al. GPTs are GPTs: Labor Market Impact Potential of LLMs[J]. *Science*, 2024, 384(6702): 1306-1308.
- [12] Autor D. Applying AI to Rebuild Middle Class Jobs[R]. *NBER Working Paper No. 32140*, 2024.
- [13] Goldman Sachs. Generative AI: Hype, or Truly Transformative?[R/OL]. *Goldman Sachs Global Investment Research*, 2023-07-05.
- [14] Litera. Litera Technology in M&A Report[R/OL]. 2022.
- [15] 上海科技大学 CEISD. AI 时代的中国劳动力市场——基于中国版动态 C-O*NET 数据库[R/OL]. 上海科技大学教育、创新和可持续发展研究中心, 2026-07-20.