

研究报告

AI 代理经济落地：真实收入、推理通缩与生产级纪律

十八个月内，AI 代理从发布会演示变成了一门收入生意：少数具名代理产品合计已产生约 29 亿美元年度经常性收入，纯代理 AI 公司 2025 年风投融资更以 64.2 亿美元创纪录。但市场结构已经反转——2026 年初资本同比多出 144%，交易笔数却少了 39%，资金全部流向已有客户的公司，而新创业公司数量几乎冻结归零。链杉研究院绘制这份报告：收入实据、驱动一切的推理成本通缩、区分「试水者」与「规模化者」的落地鸿沟，以及盈利优先型投资者应当真正看向哪里。

Industry Reports • 2026 年 10 月 3 日 • Chainfir Research

代理不再是演示：收入线已被跨过

2026 年关于 AI 代理最重要的事实，不再是某项能力基准，而是一本收入账。EigenCloud 2026 年 4 月的分析汇总了定价可验证的具名代理产品，合计年化经常性收入约 29 亿美元：Anysphere 旗下的 AI 代码编辑器 Cursor 于 2026 年 2 月跨过 20 亿美元年化收入——三个月翻一倍——此前它 2025 年 6 月破 5 亿、11 月破 10 亿美元 ARR。法律代理 Harvey 在 2025 年收入翻三倍后达到 1.95 亿美元 ARR。Bret Taylor 与 Clay Bavor 创立的企业客服代理平台 Sierra 突破 1.5 亿美元 ARR——其第一个 1 亿美元仅用七个季度。Replit 的代理做到约 1.5 亿美元，Cognigy 如今每年处理超过 10 亿次客户交互。

数字背后的深度比标题更重要。Cursor 披露 64% 的财富 500 强在使用其产品，覆盖 5 万多家企业，每天通过该平台写出的企业代码超过 1 亿行。这些不是等待提示词的助手：Cursor 端到端编排多步编码 workflow，Harvey 为九位数标的的案件起草和解条款，Sierra 的代理在财富 500 强的客户运营中执行跨系统动作。代理经济的收入第一章写在开发者工具与客服两个领域——那里工作天然数字化、可度量且海量。

具名 AI 代理产品：年化经常性收入 (百万美元)

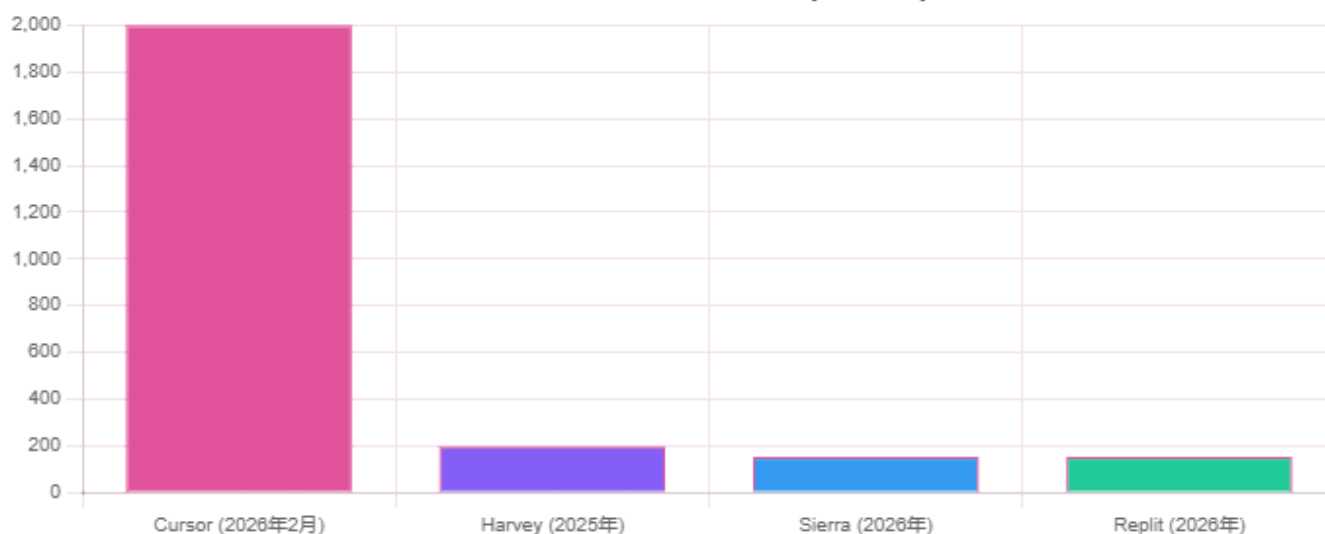


图 1：头部代理产品可验证 ARR。Cursor 从 10 亿美元（2025 年 11 月）三个月翻倍至 20 亿+（2026 年 2 月）；Harvey 2025 年内翻三倍至 1.95 亿美元；Sierra 上线八个季度达 1.5 亿美元。具名代理合计约 29 亿美元 ARR。数据来源：EigenCloud 分析（2026 年 4 月）；公司公开披露。

底层那台通缩引擎

没有计算史上最陡的成本曲线，这一切经济性都不成立。2023 年 3 月 GPT-4 发布时，前沿智能的价格是每百万输入 token 30 美元；到 2024 年 7 月，GPT-4o mini 以 0.15 美元提供了 GPT-4 级质量——十六个月下降 200 倍，据 tokencost.app 维护的 AI 价格指数（追踪各主流厂商每一次重大调价）。若以「最便宜的 GPT-4 级模型」为口径，价格底从 2022 年底的约 20 美元/百万 token，降至 2024 年四季度的 0.88 美元、2025 年四季度的约 0.40 美元——三年 50 倍，出自中泰证券行业深度报告。同期 PC 计算成本仅降约 3.4 倍、互联网带宽约 2.6 倍。智能的通缩速度比此前任何通用投入品快一个数量级。

两股结构性力量让曲线继续下行。其一，开放权重竞争：Meta 的 Llama 3.1 与 DeepSeek 的 V3（2024 年 12 月发布，每百万输入 token 0.14 美元，约为 GPT-4 首发价的 1%）反复重置价格底，倒逼所有西方厂商重定价。其二，前沿本身也在变便宜：Anthropic 在 2025 年底一夜之间把 Claude Opus 降价 67%；Andreessen Horowitz 估计，同等能力的推理成本自 2021 年以来每年下降约 10 倍。对代理开发者而言，这意味着复利：去年价格下不经济的工作负载，今年毛利率转正；而代理产品当前 40–70% 的毛利率（传统 SaaS 为 70–90%）会随最大成本项的通缩机械扩张。

OpenAI 旗舰模型每百万 token 输入价格 (美元) , 2023–2026

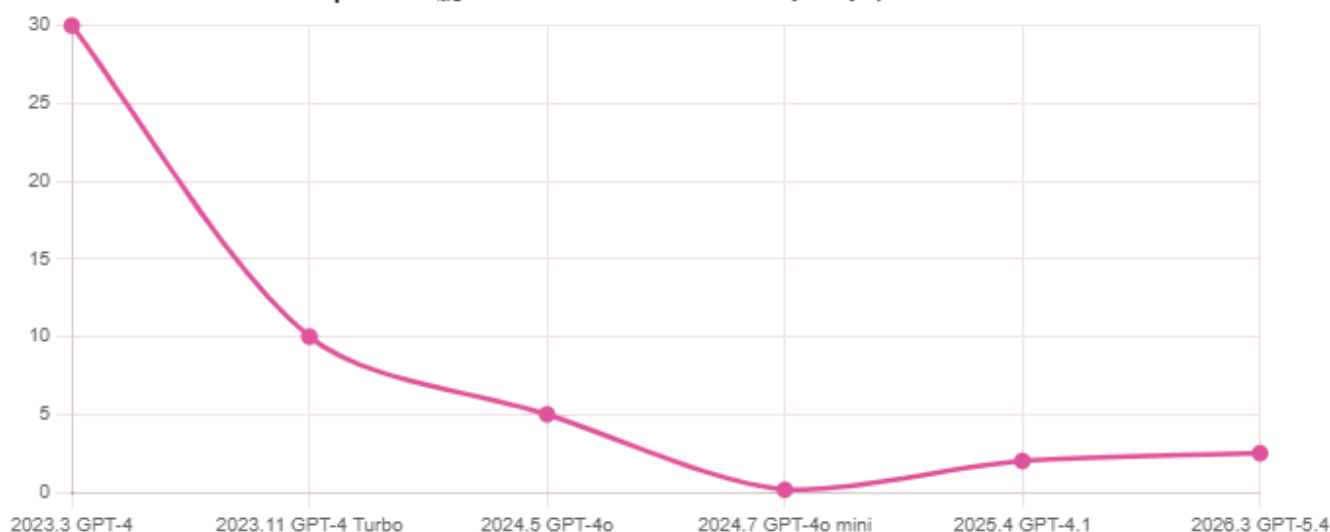


图 2: GPT-4 级质量的智能价格在 16 个月内从每百万 token 30 美元降至 0.15 美元 (200 倍) ; GPT-4 级模型价格底三年下降约 50 倍, 2025 年四季度约 0.40 美元。数据来源: tokencost.app AI 价格指数 (各官方 API 定价页) ; 中泰证券行业深度报告。

资本信号：更少、更大、更后期

风投已经投票——投的是「集中」。据 AgentMarketCap 汇编的 Tracxn 赛道数据, 纯代理 AI 公司 2025 年融资 64.2 亿美元创纪录, 十年累计 242 亿美元。但 2026 年的形态恰好反转: 2026 年 1–4 月, 代理 AI 公司融资 26.6 亿美元、仅 44 笔; 而 2025 年同期是 10.9 亿美元、71 笔——资本多出 144%, 交易少了 39%。平均单笔从约 1,500 万美元升至约 6,000 万美元, 近期最大 15 笔轮次均值达 1.55 亿美元, 几乎是 2025 上半年均值 (8,200 万) 的两倍。最刺眼的数字是: 2023 年新成立的代理 AI 创业公司有 245 家, 2026 年第一季度 Tracxn 实时赛道图上恰好只有 1 家。市场不是冷了, 是在分拣。

New Market Pitch 对 2025 年 8 月至 2026 年 9 月每一笔纯代理 AI 融资的逐笔研究 (39 笔、36.8 亿美元) 显示同样的压缩: 前十大交易吸走 69.1% 的资本, B 轮及以后拿走 82.4%, 39 笔中 32 笔是投向已融过资公司的追加轮。地理分布同样集中: 北美 33 笔 (26.6 亿美元)、欧洲 6 笔 (10.2 亿美元)。例外是中国: 毕马威记录, 2025 年中国代理式 AI 融资 294.7 亿美元、同比增长 39%, 2026 年一季度单季即达 94.5 亿美元、同比增 129.4%——置于全年 406.4 亿美元 (+58.5%)、2,336 笔的中国 AI PE/VC 总盘子之内。各地的资本都在为生产系统付费, 而不是为原型付费。

代理 AI 风投融资：资本 +144%，笔数 -39%

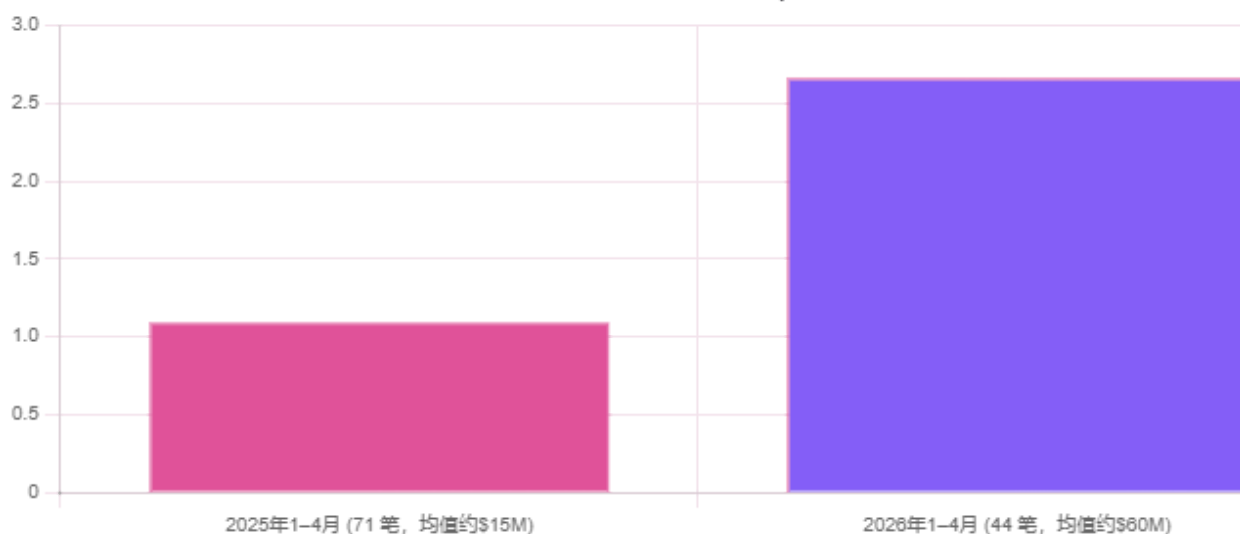


图 3：纯代理 AI 融资同窗口对比：资本增 144%、笔数减 39%，平均单笔约翻两番。2025 全年以 64.2 亿美元创纪录。新创业公司数量从 2023 年 245 家坍缩至 2026 年一季度 1 家。数据来源：Tracxn 赛道数据，经 AgentMarketCap（2026 年 4 月）。

单位经济：按结果定价取代按席位定价

商业模式的创新与技术本身同样重要。Intercom 的 Fin 代理按「每张解决的客户工单 0.99 美元」收费——不按席位、不按消息、按结果——对比人工每单 15–25 美元的综合成本：15–25 倍的成本压缩，EigenCloud 称之为「经济上的必然」。Klarna 披露其内部 AI 部署替代了 700 个全职客服当量的工作，年化节省约 6,000 万美元——并把这一叙事带进了 IPO 进程。当软件对着人力预算而非 IT 预算定价时，可及市场扩大一个数量级：仅全球客服一项就是数千亿美元级的人力开支。

效率同样体现在组织架构上。Anysphere 每员工对应约 1,370 万美元 ARR，而 Salesforce 约 29 万美元——是这家全球最大企业软件公司收入效率的 47 倍（EigenCloud 分析）。Harvey 一年收入翻三倍而人数不翻三倍。这就是「数字劳动力」的实际含义：代理不是组织偶尔取用的工具，而是组织如今赖以运行的流程——最佳估值真正定价的正是这种工作流所有权，而不是模型先进程度。

落地鸿沟才是真正的市场地图

在意图与生产之间，横着 2026 年最具决定性的鸿沟。MuleSoft 与 Deloitte Digital 对 1,050 名 IT 负责人的调查显示，93% 计划在两年内引入自主代理。McKinsey 2025 年 11 月的《State of AI》发现 62% 的组织正在试验 AI 代理——但仅 23% 在至少一个业务职能中规模化，跨多职能规模化的不足 10%。Gartner 基于 3,412 家组织的调研预测，到 2027 年底超过 40% 的代理 AI 项目将被取消，理由是成本攀升、ROI 不清与风控不足——尽管它同时预测到 2026 年底 40% 的企业应用将嵌入任务型代理（2025 年不足 5%）。两件事同时为真：采用近乎全民化，而其中大多数无法规模化。

失败模式在 McKinsey、Gartner 与 Deloitte 的研究中高度一致——是组织性的，而非技术性的：部署前没有治理框架、没有可观测工具看清代理在做什么、试点前没有采集基线指标（于是 ROI 永远无法被证明）、没有担责的业务负责人。51% 的客服负责人表示安全顾虑推迟或限制了 AI 计划。成功少数派共享四个特征：部署前先投

基础设施、治理文档先行、试点前采集基线、设专职负责人背负绩效。对投资者的实务解读：落地鸿沟不是对这个赛道保持谨慎的理由——它是持久价值向垂直专家、以及弥合鸿沟的治理/身份/可观测基础设施集中的原因。

价值在哪里汇聚

资本已经画好了答案。在 New Market Pitch 的数据集中，垂直 AI 代理——为网络安全、医疗、法律、金融、客户运营等特定领域构建的系统——贡献了 39 笔中的 22 笔、30.4 亿美元，占全部披露资本的 82.6%。Tracxn 更长窗口显示，仅客服与销售一类就吸走 2022–2025 年代理 AI 融资的约 37%。估值纪律也在回归：Finro 2026 年一季度覆盖 214 家 AI 代理公司的数据集显示平均倍数约 30 倍 ARR，离散度更多由工作流位置而非增速解释——拥有关键任务流程的代理（Sierra 约 67 倍）比租用注意力的代理更值钱。市场规模估算因口径而异，但方向一致：Grand View Research 将企业级代理 AI 定为 2025 年 36.7 亿美元、2030 年 245 亿美元（CAGR 46.2%）；更宽口径下 AI 代理总市场 2026 年 109 亿美元、2030 年 503 亿美元。

对盈利优先型投资者，我们在这个市场重点盯五件事：

- **在强监管领域已有生产收入的垂直代理。**近期 82.6% 的资本流向垂直专家；法律、金融、医疗与合规代理按结果变现，而非按席位。
- **治理与可观测层。**62% 对 23% 的落地鸿沟就是一张基础设施采购单：代理的身份、授权、审计与监控，是卖铲子的位置。
- **单位经济已被验证的按结果定价企业。**按单/按任务定价把推理通缩直接转化为毛利率扩张。
- **亚太被低估的创业形成。**近期 39 笔纯代理融资北美拿走 33 笔；除中国 295 亿美元的代理融资浪潮外，亚太创业形成仍然稀薄——这是区域资本的结构性的入场窗口。
- **复利型公司，而非冲刺型公司。**平均倍数向 30 倍 ARR 回归、40% 以上项目预计 2027 年前被取消的环境里，持久的工作流所有权胜过演示速度。

纯代理 AI 融资按类别分布，2025.8–2026.9（十亿美元）

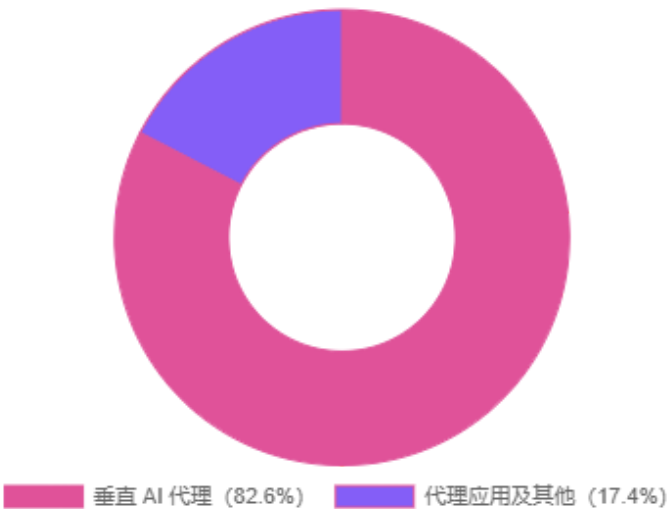


图 4：39 笔纯代理 AI 融资合计 36.8 亿美元中，垂直 AI 代理（法律、金融、网络安全、客户运营等领域专用系统）占 22 笔、30.4 亿美元——资本的 82.6%。数据来源：New Market Pitch 代理 AI 融资分析（2026 年 9 月）。

盈利优先的读法

代理经济的第一章，与我们自 2018 年以来承销过的每一轮基础设施周期同韵：最喧闹的资本去前沿，但持久的回报归于「能力遇到付费客户」的那一层。智能价格 200 倍的崩塌没有摧毁价值——它把价值从「卖智能的人」手里，转移到「把智能用到过去由人完成的工作上」的人手里。如今在复利生长的公司——Cursor、Sierra、Harvey——与我们研究的每一家持久型科技企业共享同一特质：如果产品消失，客户会立刻在钱上感觉到。

这正是我们的纪律所施加的过滤器。链杉资本只投已盈利企业——至少连续八个季度正运营贡献——约 90% 出资来自自有资产负债表，并把资本、人才与资源整合为同一头寸打出去。在一个 62% 的组织在试验、仅 23% 在规模化的市场里，我们没有兴趣资助试验；我们感兴趣的是站在这个比例正确一侧的公司——已经在生产、已经在赚钱、已经在复利的那些——遍布亚太的 AI、Web3 与前沿科技版图。

这一轮的算术异常干净。智能每年通缩 10 倍；按结果定价把通缩转成毛利；垂直工作流所有权把毛利转成护城河。代理经济不再需要被相信，它需要被承销。

「两年前的问题是 AI 代理能不能赚到钱。那场争论已经结束了——收入是真的，单位经济被验证了，智能成本还在以每年 10 倍下降。剩下的唯一问题是纪律：投靠生产工作流赚钱的公司，而不是做演示的公司。这个过滤器自 2018 年起定义了链杉，而这一轮周期对它的奖赏超过以往任何一轮。」——链杉资本创始人兼 CEO Brian Tian