下载日志已记录，仅供内部参考

中泰证券 ZHONGTAI SECURITIES

从模型调用到任务交付：Agent 升级重构大模型价值

——2026 AI 应用系列深度报告（六）

传媒互联网

证券研究报告/行业深度报告

2026 年 10 月 23 日

评级：　增持（维持）

分析师：康雅雯

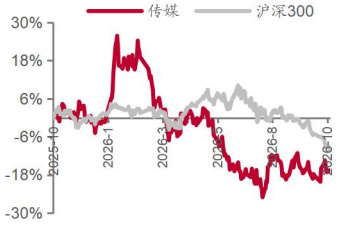
执业证书编号：S0740515080001

Email：kangyw@zts.com.cn

基本状况

上市公司数	131
行业总市值(亿元)	14,915.77
行业流通市值(亿元)	13,732.09

行业-市场走势对比



相关报告

1、《运动题材与影院场景双向拓展，“电影+体育”打开消费新空间》2026-10-19

2、《Meta Muse 专题：消费场景渗透，资源叙事更新与生态价值重估【中泰传媒互联网】》2026-09-29

3、《【中泰传媒互联网】2Q26 游戏财报总结：业绩高增，AI 催化，带动估值修复》2026-09-09

重点公司基本状况

简称	股价（元）	EPS 2024A	EPS 2025A	EPS 2026E	EPS 2027E	EPS 2028E	PS 2024A	PS 2025A	PS 2026E	PS 2027E	PS 2028E	评级
腾讯控股	337.2	20.1	24.8	26.0	27.8	30.5	5.1	4.2	3.7	3.4	3.1	买入
阿里巴巴-W	88.8	13.6	7.4	4.8	5.0	7.0	2.0	1.9	1.7	1.6	1.4	-
智谱	424.7	0.0	-9.0	-8.6	-6.6	0.6	0.0	150.5	20.5	7.7	3.9	-
MINIMAX-W	147.2	0.0	-27.2	-17.4	-19.2	-17.6	0.0	104.7	19.0	7.0	3.4	-

备注：股票截止自 2026 年 10 月 22 日收盘价，每股指标按照最新股本数全面摊销。未覆盖公司预测数据取自 IFind 一致预期，货币单位为人民币元

报告摘要

■ **Agent 升级拓宽模型需求，产业价值向任务交付与生态服务延伸。**

■ 模型能力提升与任务成本下降，正在使 AI 能够承接更多实际工作；Agent 通过多轮执行和持续跟进，进一步增加模型使用深度。企业采购、个人付费与长期协作需求共同发展，使模型厂商和应用产品的商业价值，更多体现为能够完成哪些任务、为客户创造多少价值，以及如何通过产品和生态承接付费。

■ **高水平智能的使用门槛降低，模型与执行系统共同决定任务经济性：**

➢ **前沿智能水平相当，单任务成本下降：**Artificial Analysis 评测显示，GPT-6.1 Sol 与 GPT-6 Astra 的智能指数分别为 52 和 53，单任务平均 API 费用分别为 0.72 美元和 3.26 美元，前者以低约 78%的成本实现了相近的综合表现；

➢ **模型与执行框架的不同组合，影响单任务费用：**Coding Agent 评测显示，在综合得分相同的情况下，不同模型与执行框架组合的单任务费用差距明显；

➢ **报价与配置同步优化，拓宽应用落地范围：**随着模型报价、推理配置及工具调用方式协同改善，开发者能够按需灵活分配预算，进一步拓宽 AI 的应用场景。

■ **Agent 加深模型调用，缓存复用为连续执行提供成本支撑。**

■ OpenRouter 样本显示，2026 年 9 月 Agent 周 Token 用量已达到人工交互类别的约 7 倍，且接近 90%的 Agent Token 属于缓存输入。Agent 围绕同一目标反复读取背景、调用工具并处理结果，使一次需求对应多轮模型调用；缓存则通过复用已经处理的内容，降低重复读取的费用。调用深度增加与复用成本下降相互配合，有望支持更多需要多步处理、持续跟进的任务。

■ **企业应用逐步形成采购依据，收费方式更加贴近业务交付。**

■ Apollo 统计显示，2026 年第二季度，69%的标普 500 公司提及已上线 AI 部署，29%提供了量化结果；微软 Microsoft 365 Copilot 付费席位超过 3000 万，ServiceNow AI 年度合同价值超过 10 亿美元。部署、成效与付费数据共同反映，AI 正进入企业实际业务流程。随着产品将多轮调用组织为具体功能和服务，按席位、执行动作及约定结果收费的模式持续发展，客户价值也有了更明确的商业承接方式。

■ **个人助手走向长期协作，持续服务拓宽模型与应用空间。**

■ 记忆、主动跟进与权限管理，使个人助手能够保留用户背景和任务进度，在约定时间或收到新信息后继续处理事务。我们认为，这类能力有望将单次使用延伸为持续服务，既增加模型调用需求，也通过节省时间、提高任务完成效率，为用户留存和续费提供支持。个人助手与办公 Agent 共享相关基础能力，也为从办公付费需求切入长期服务提供了路径。

■ **投资建议：关注模型增长与个人 Agent 生态两条主线。**

➢ 模型侧：建议关注**智谱、MiniMax**，重点跟踪模型迭代、调用量增长及商业收入兑现；

➢ 模型与 Agent 生态侧：建议关注**腾讯、阿里巴巴**，重点观察用户入口、办公与生活服务的整合，以及持续使用和付费转化；

➢ 垂类 Agent 与终端个人助手侧：建议关注**迈富时**其企业 Agent 相关业务落地与**国脉文化**关注其云智手机个人助手布局。

下载日志已记录，仅供内部参考

- 整体而言，模型能力与产品系统、业务生态的协同，有望共同拓宽 AI 的应用范围与商业化空间。
- 风险提示：行业政策与监管风险、技术研发不及预期风险、市场竞争加剧风险、盈利不及预期风险、测算偏差和数据更新不及时的风险等。

下载日志已记录，仅供内部参考

内容目录

低价智能拓宽应用边界，任务效率重塑竞争维度 4
模型能力与成本协同改善 4
模型+Agent 形成能力与成本的多种组合 5
模型调用规模与支出分布分化 7
Agent 连续执行提高模型使用强度 8
缓存降上下文复用的成本 9
企业采用进入业务流程，经济价值开始形成采购支撑 9
企业 AI 部署与成效推动采购规模落地 9
收费单位更贴近工作交付 11
个人付费持续扩大，专业工具获得更强需求 11
付费用户规模扩张，存量用户支出持续提升 11
通用助手扩大使用，专业工具服务具体工作 13
Agent 走向长期协作，记忆与主动性拓宽应用边界 15
个人 Agent 访问量扩张，需求方向明确 15
模型能力与产品系统推动 Agent 走向长期协作 16
投资建议 19
投资建议与标的 19
风险提示 19
行业政策与监管风险、技术研发不及预期风险、市场竞争加剧风险、盈利不及预期风险、测算偏差和数据更新不及时的风险等 19

图表目录

图表 1： 模型价格与能力调整价格指数 4
图表 2： 主流前沿大模型智能指数及价格对比图 5
图表 3： 主流前沿大模型 Coding Agent 能力与单任务成本对比图 6
图表 4： TokenEconomizer 测试原理机制解析 7
图表 5： OpenRouter 上用户对不同模型厂商的支出占比与平台 Token 用量 8
图表 6： OpenRouter 上的 Agent 与人类 Token 调用量对比 8
图表 7： OpenRouter 上的 Agent 与人类 Token 使用构成 9
图表 8： 标普 500 公司企业 AI 部署情况 10
图表 9： AI 的成效目前主要体现在降本 10
图表 10： AI 大模型收费模式 11
图表 11： AI 消费支出增长集中于最高支出群体 12
图表 12： 不同 AI 客户群体的支出指数 12
图表 13： 美国家庭订阅 AI 服务占比 13
图表 14： 全球 AI 网站访问量榜单前列产品 14
图表 15： 中国 AI 网站访问量榜单前列产品 15
图表 16： 不同个人 Agent 访问量趋势 15
图表 17： 个人助手如何把用户目标变成持续执行 16
图表 18： 个人助手定时心跳与事件唤醒图解机制 17
图表 19： 个人助手记忆系统图解机制 17
图表 20： 个人助手与办公 Agent 的共同基础 18
图表 21： 企业 AI Agent 的演进 18

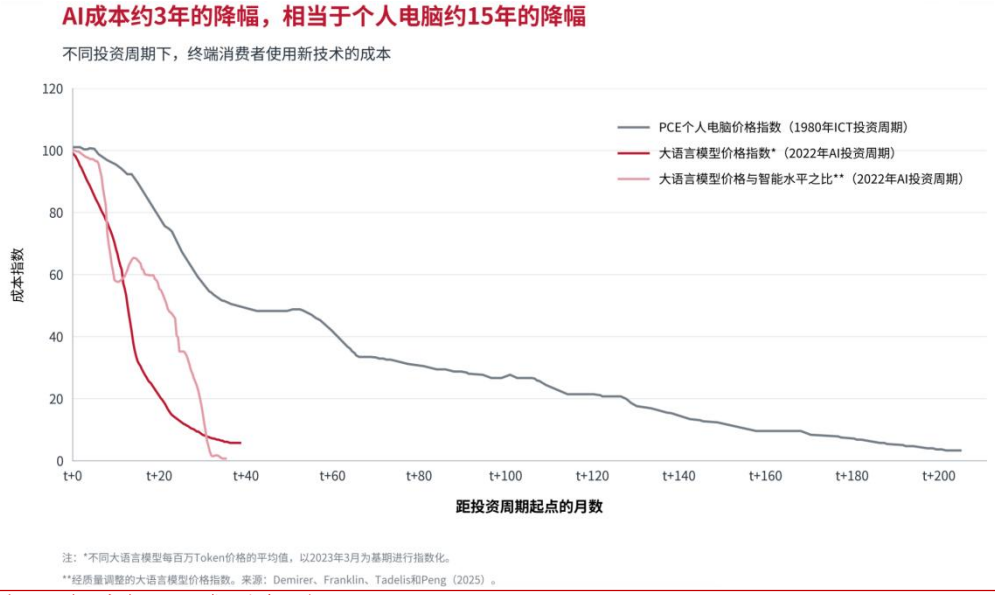
下载日志已记录，仅供内部参考

低价智能拓宽应用边界，任务效率重塑竞争维度

模型能力与成本协同改善

- **大模型能力与调用成本正在实现协同改善，高水平智能得以覆盖更广泛的预算区间，优质智能的使用门槛进一步降低。**根据美国商务局与彭博关于不同投资周期下终端消费者使用新技术成本的统计，2022 年开启的 AI 投资周期中，大语言模型价格与智能水平之比在约 3 年内的降幅，已相当于 1980 年 ICT 投资周期中个人电脑（PCE）约 15 年的降幅。

图表 1：模型价格与能力调整价格指数

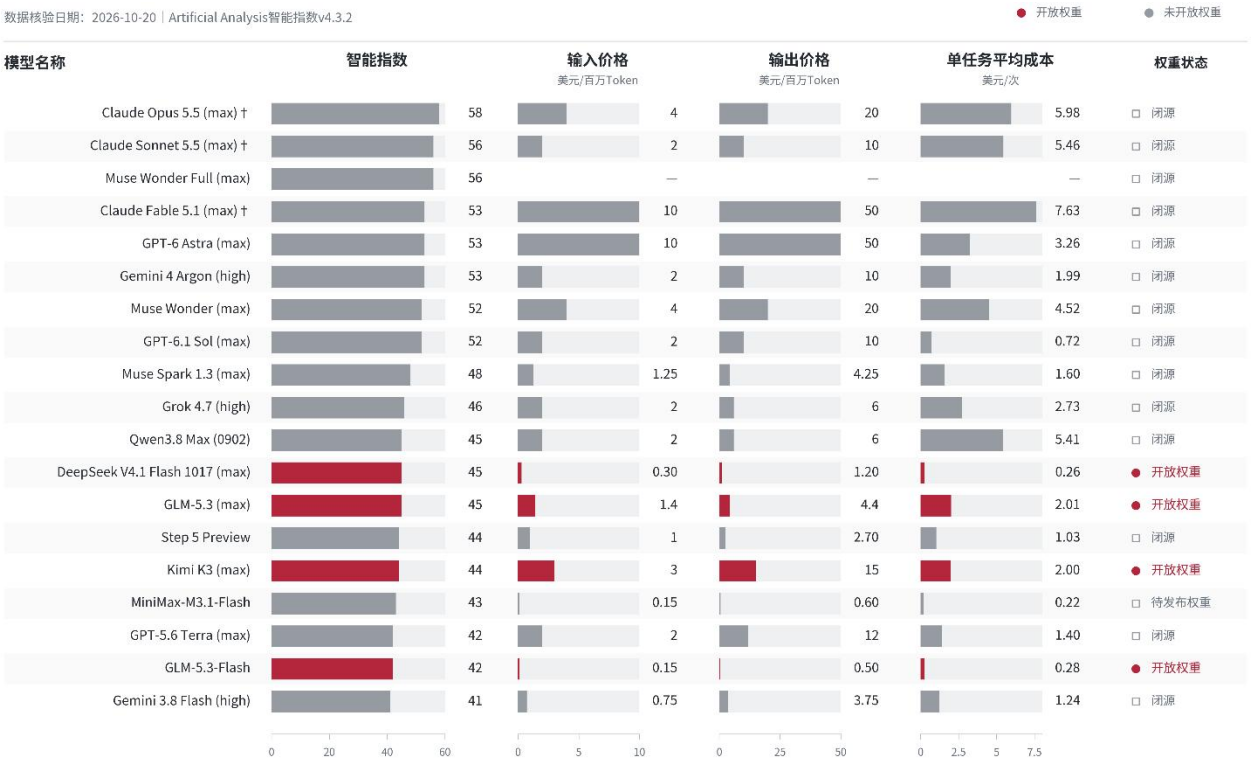


来源：美国商务局，彭博，中泰证券研究所

- 这种成本下移并非单纯的账面调价，而是从“模型报价、执行效率、分层供给”三个维度同步向实际工作传导：
  - **前沿模型的单任务调用成本下探，高水平智能使用门槛降低**：在 Artificial Analysis 最高推理档位的智能指数评测中，GPT-6 Astra 得分为 53，平均每任务 API 费用为 3.26 美元；而 GPT-6.1 Sol 得分为 52，平均每任务 API 费用仅为 0.72 美元。两者智能指数仅相差 1 分，但后者单任务费用低约 77.9%。这组结果反映，接近前沿水平的综合表现已经能够通过更低的调用预算获得，模型选择开始拥有更充分的能力与成本组合。
  - **报价下调与 Token 消耗结构共同压降实际负载费用**：Anthropic 的 Opus 5.5 较 Opus 5 将输入、输出 Token 报价降低 20%，缓存读取报价降低 60%；在公司测试的默认设置与典型负载下，整体使用费用预计降低 40%。单 Token 价格变化与完整工作负载费用变化幅度不同，说明模型完成任务所消耗的 Token 结构和数量，已经成为影响使用成本的重要因素。
  - **高性价比供给进一步丰富模型服务层次**：DeepSeek V4.1 Flash 1017 最高推理档位的智能指数为 45，输入、输出价格分别为每百万 Token 0.30 美元和 1.20 美元，智能指数评测平均每任务费用低至 0.26 美元；MiniMax-M3.1 Flash（智能指数 43）单任务平均成本亦仅为 0.22 美元。结合 GPT-6.1 Sol 和 Opus 5.5 的效率改善，不同能力与价格区间均出现了更具经济性的模型供给，应用开发者因而能够在满足具体任务质量要求的前提下，将模型预算灵活分配至不同工作环节。

下载日志已记录，仅供内部参考

图表 2：主流前沿大模型智能指数及价格对比图



*数据截止 2026/10/20

来源：Artificial Analysis，中泰证券研究所

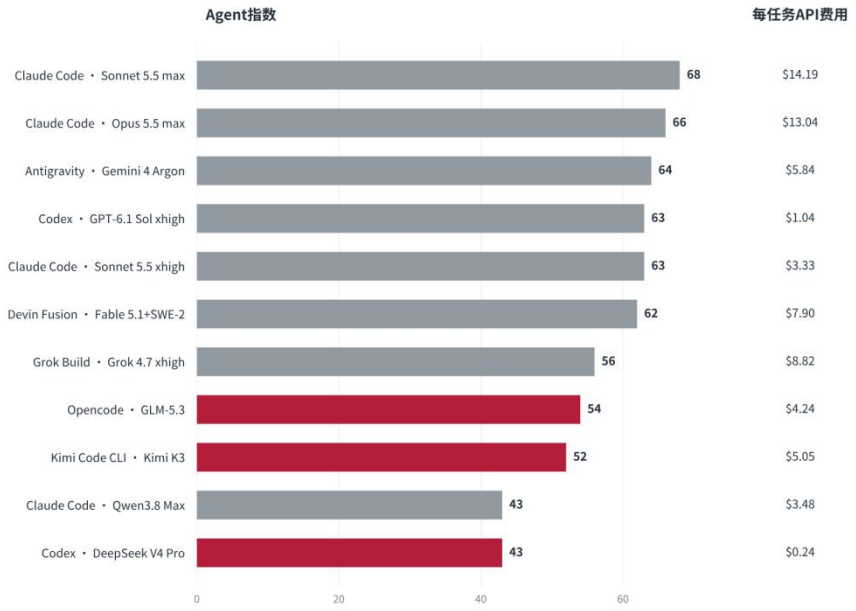
- 模型供给改善的产业意义，在于增加能够以合理成本完成的实际工作。将单位价格下降与效用提升共同视为 AI 扩大采用的重要条件：前者提高可负担性，后者增加用户能够获得的服务价值。**我们认为**，当前模型迭代中的报价优化、缓存折扣和任务效率提升，正沿着这一机制改善应用开发与使用条件；供给改善与商业化之间的连接，已由单纯的“单位智能价格”进一步延伸至“单位任务所产生的经济性”。

模型+Agent 形成能力与成本的多种组合

- 在 Agent 产品落地过程中，最终的交付效率不仅取决于底层大模型，更体现为模型与执行系统的共同表现。开发者既可以通过**调整单一模型的推理档位**适配预算，也可以通过**优选不同的模型与执行框架组合**实现能力与成本的平衡：
  - **不同模型与执行框架组合形成明显的任务成本差异**：根据 Artificial Analysis Coding Agent Index v1.5 评测，Codex 搭配 GPT-6.1 Sol xhigh 的 Coding Agent 指数为 63，每任 API 费用仅 1.04 美元；而 Claude Code 搭配 Sonnet 5.5 xhigh 同样取得 63 分，每任务费用却达 3.33 美元。与此同时，Claude Code 与 Sonnet 5.5 max 组合以较高费用（14.19 美元）获得最高综合得分（68 分），国产开放权重模型组合如 Opencode·GLM-5.3（54 分，4.24 美元）、Kimi Code CLI·Kimi K3（52 分，5.05 美元）、Codex·DeepSeek V4 Pro（43 分，0.24 美元）亦各具成本与能力特征。**我们认为**，模型与 Agent 的不同组合分别提供了能力上限与成本效率方面的选择；在复杂任务和高频任务具有不同质量要求、时间要求与预算条件时，这种分层供给有助于形成更丰富的 Agent 产品结构，扩大 Agent 可以承接的任务范围。

下载日志已记录，仅供内部参考

图表 3：主流前沿大模型 Coding Agent 能力与单任务成本对比图



*数据截止 2026/10/15

来源：Artificial Analysis，中泰证券研究所

- 此外，**外部执行控制系统的优化正进一步挖掘既有模型的成本潜力**。根据 2026 年 10 月 15 日发布《TokenEconomizer》研究，在 Claude Code、Codex 和 GitHub Copilot CLI 等封装智能体与运行框架上测试了外部黑盒成本控制机制。该机制通过“控制器 + 学习器”闭环对三个环节进行干预：
  - 微技能注入（知识层）：从已完成的执行轨迹中提炼指令与检测器，合成微技能置于系统与运行框架指令前缀中，依据证据决定是否启用；
  - 工具裁剪（接口层）：在每项任务开始前，屏蔽近期任务未使用或被判定无关的工具（如未加载 web、jupyter、browser、mcp:*），仅保留 shell、read、edit、grep、git、test 等核心工具集；
  - 失败循环干预（控制层）：学习当前仓库的交互轮次预算，在实时执行过程中一旦检测到重复失败（如反复执行 edit -> test 报错），立即引导一次方向调整或中止无效重复执行，并将完成后的执行轨迹回流至学习器更新元策略。
- 在 277 项 SWE-Bench 任务、6 个代码仓库的测试中，TokenEconomizer 在包含控制器自身开销的情况下，将平均每任务费用降低 18.4%—30.9%，并保持了相近的任务成功表现。**我们认为**，这项实验说明，执行系统能够通过减少重复工作和改善工具组织，明显提高既有模型的任务经济性，为 Agent 产品的效率竞争提供了依据。

下载日志已记录，仅供内部参考

图表 4：TokenEconomizer 测试原理机制解析



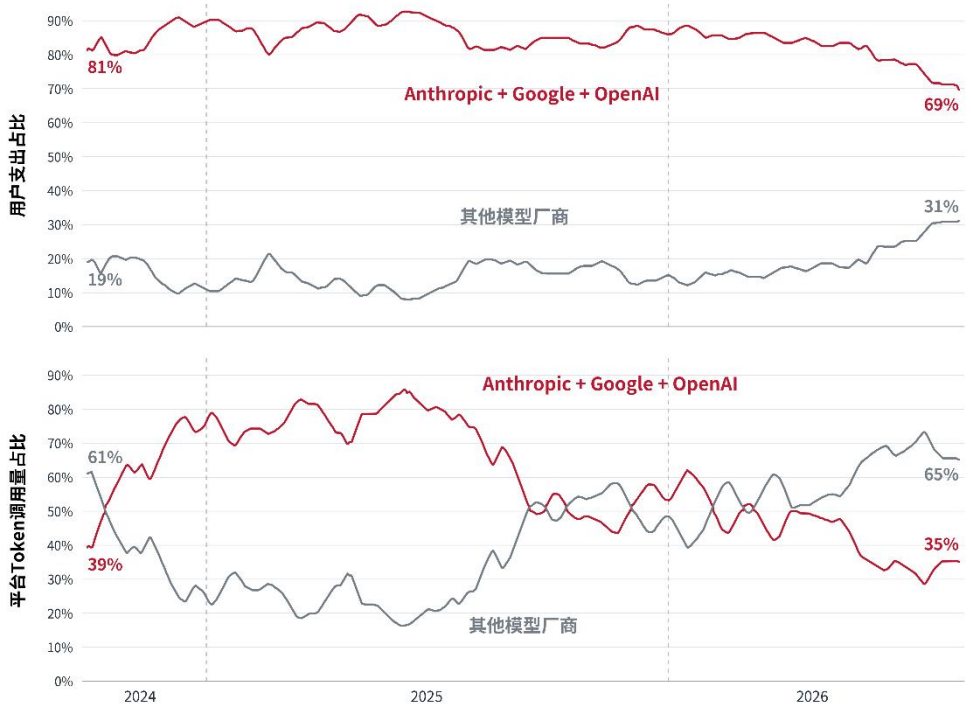
来源：《TOKENECONOMIZER: Experience-DrivenBlack-Box Cost Control for Coding Agents》论文，中泰证券研究所

模型调用规模与支出分布分化

- **前沿模型的商业溢价与低价模型的调用扩张可以同时存在。** 截至 2026 年 8 月 28 日，OpenRouter 的样本中，用户对 Anthropic、Google 与 OpenAI 合计产生约 69%的模型支出和 35%的 Token 调用份额。支出份额高于 Token 份额，表明**这三家厂商在该平台样本中具有更高的平均 Token 变现水平**。我们认为，分层的模型表现与价格为企业增加用例提供了更多选择，前沿模型仍可在前沿问题上获得溢价。
- **需求扩张由不同层次的模型供给共同承接。**一方面，高水平模型以更低任务成本进入更广的预算范围；另一方面，客户模型支出仍向部分前沿模型集中。两项事实共同支持模型市场的分层发展：**基础任务与复杂任务可以采用不同投入强度，调用规模和付费规模也会形成不同分布。模型厂商的商业价值与其能够承接的任务类型、实际使用结构及产品服务方式相关。**

下载日志已记录，仅供内部参考

图表 5：OpenRouter 上用户对不同模型厂商的支出占比与平台 Token 用量



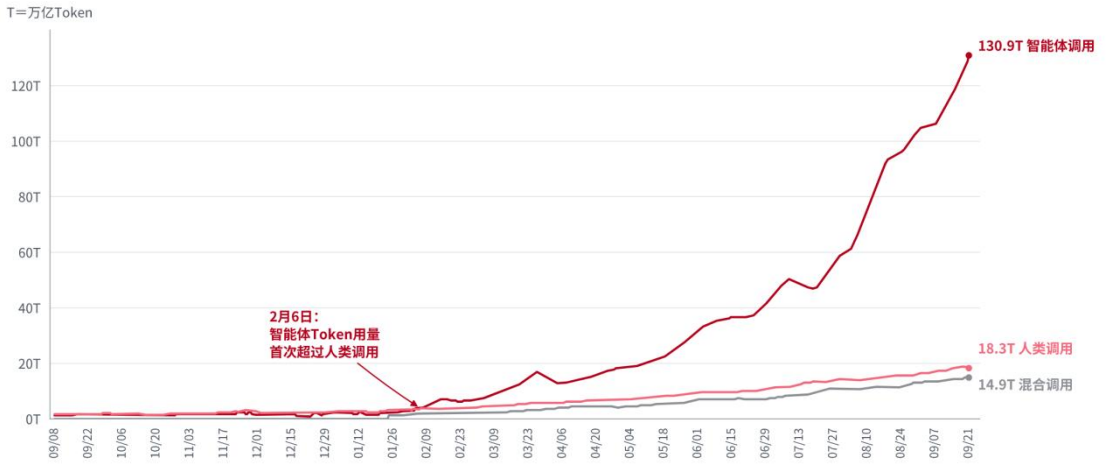
*数据截止 2026/8/28

来源：Openrouter，中泰证券研究所

Agent 连续执行提高模型使用强度

- Agent 已经成为 OpenRouter 等平台 Token 使用的重要组成部分。按 API 行为分类，Agent Token 用量于 2026 年 2 月 6 日首次超过人工类别；截至 9 月最后一周，周用量约为人工类别的 7 倍。**我们认为，这一变化表明模型调用的主体正在从单次真人交互，转向由软件持续驱动的多轮执行，调用量的增长因此更多来自任务深度与执行轮次的提升。**

图表 6：OpenRouter 上的 Agent 与人类 Token 调用量对比



*数据截止 2026/9/21

来源：Openrouter，中泰证券研究所

- **调用强度上升来自 Agent 围绕目标连续组织工作的方式。**真人交互通常由用户逐次提出需求，Agent 则需要在同一任务中读取上下文、调用工具、处理返回信息，并根据进展继续执行。任务历史随着工作推进而累积，同一用户需求可以对应多轮模型调用。**我们认为 Agent 在顺序执行一系列任务会**

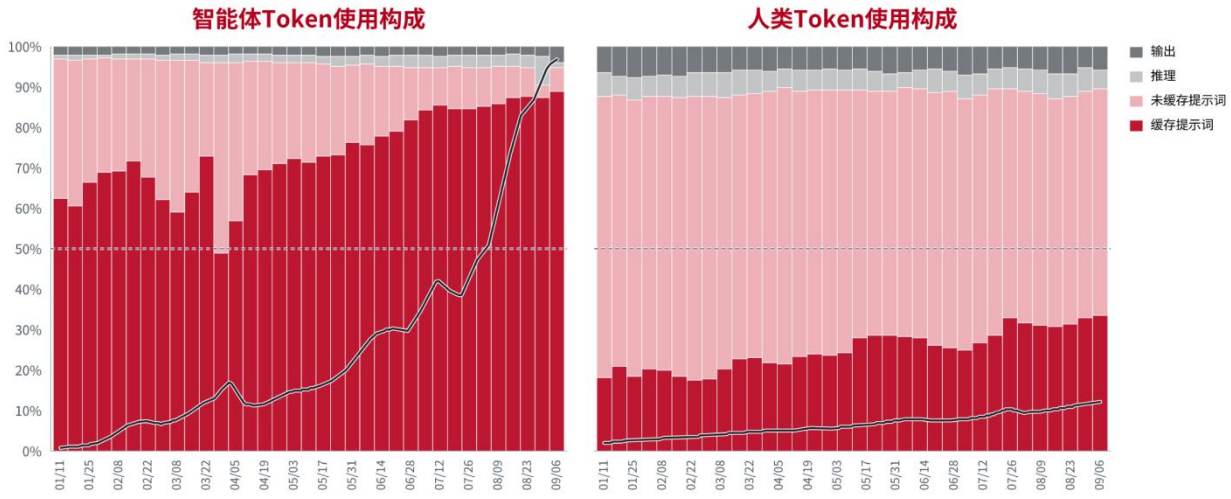
下载日志已记录，仅供内部参考

显著增加 Token 消耗，模型需求因而同时受到用户采用与任务执行深度的影响。

缓存降上下文复用的成本

- **Agent 的高 Token 用量与较低的有效调用单价可以同时存在。**2026 年 9 月，OpenRouter 上的 Agent Token 中，接近 90%属于缓存输入。Agent 在每一步都会重新发送系统指令、工具定义、已读文件和任务历史，其中大量内容模型此前已经处理过。提示缓存能够复用这部分计算结果，使持续执行过程中的重复上下文获得明显低于新增内容的收费价格。

图表 7：OpenRouter 上的 Agent 与人类 Token 使用构成



*数据截止 2026/9/21

来源：Openrouter，中泰证券研究所

- **缓存 Token 之所以重要，是因为它们成本低廉。**在 Anthropic、OpenAI 和 Google 目前的前沿模型中，缓存读取的成本仅为全新输入令牌的十分之一，输出令牌的五十分之一。例如，Claude Sonnet 5 的收费标准为：每百万个输入令牌 2 美元，每百万个输出令牌 10 美元，每百万个缓存令牌 0.2 美元。最新版本进一步拉大了差距。Claude Opus 5.5 和 GPT-6.1 Sol 对缓存读取的定价为输入值的 1/20 和输出值的 1/100。而 Claude Fable 5.1 的定价则为输出值的 1/200。
- **缓存折扣已经对典型 Agent 工作负载产生明显影响。**OpenRouter 以 Sonnet 5 报价测算的一个示例显示：每百万 Agent Token 按 87%缓存输入、12%新增输入和 1%输出的结构计费，费用约 0.51 美元；若取消缓存折扣，同样结构的费用约为 2.08 美元。前者仅为后者的约四分之一。由此可见，缓存在多轮任务的经济性中已是核心组成部分。
- **多组数据共同表明，Agent 需求深化与成本优化正在并行推进。**平台 Agent 调用持续扩张，说明多轮执行已形成实际使用；缓存占比上升与缓存折扣说明重复上下文能够以更低价格复用；Coding Agent 的任务成本差异则表明执行配置仍有优化空间。这三类证据分别对应需求、成本机制和产品实现，共同构成 Agent 扩大应用范围的经济基础。

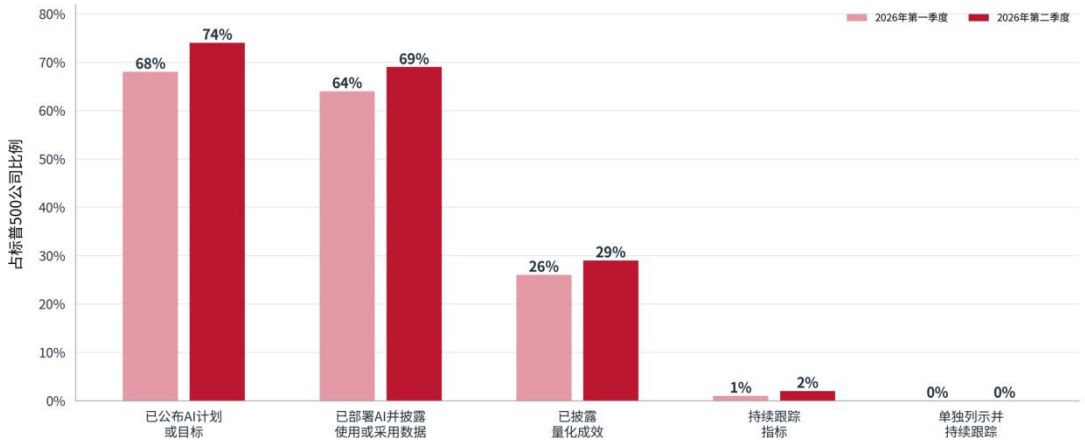
企业采用进入业务流程，经济价值开始形成采购支撑

企业 AI 部署与成效推动采购规模落地

- **AI 正在由企业技术配置进入实际业务运行。**Apollo 对 2026 年第二季度标普 500 公司财报季的统计显示，69%的企业提及已上线 AI 部署，较上一季度提高 5 个百分点；29%的企业提供量化结果。部署比例提升反映企业采用范围扩大，量化结果则说明部分企业已能够将 AI 使用与具体经营产出联系起来，企业应用同时进入扩面和效果兑现阶段。

下载日志已记录，仅供内部参考

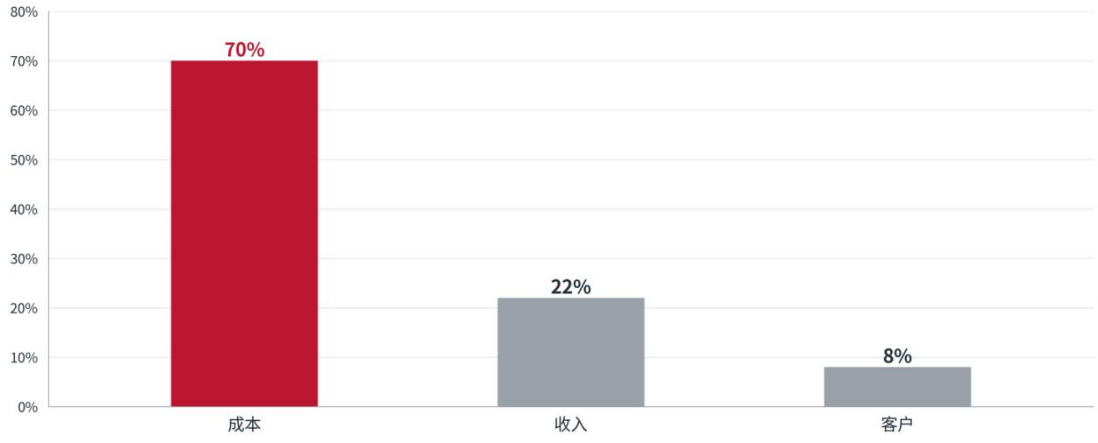
图表 8：标普 500 公司企业 AI 部署情况



来源：Apollo，中泰证券研究所

- **现阶段，企业更常通过提高效率和降低成本，衡量 AI 的使用价值。** 根据 Apollo 数据显示，已量化 AI 成效中，成本相关内容占 70%，收入相关内容占 22%。这与 AI 改善已有工作处理方式的用途相联系：当企业把 AI 用于现有流程时，可以围绕完成工作所需的时间、资源和费用评估效果。收入相关成效也已出现，表明企业同时开始报告 AI 在业务增长方面的作用。

图表 9：AI 的成效目前主要体现在降本



来源：Apollo，中泰证券研究所

- **节省工作时间，已经成为企业扩大 AI 使用的具体依据。**NHS England 的试验显示，向 50.5 万名临床及其他工作人员提供 Microsoft 365 Copilot 的访问权限后，员工平均每天节省 43 分钟。每天节省的时间，使客户能够直观衡量 AI 对日常工作的作用；向更多员工推广，则是扩大使用范围的实际安排。这一案例把产品功能、员工使用效果和客户部署行动联系起来，展示了 AI 如何通过改善已有工作获得进一步采用。
- **企业愿意为 AI 正式付费，已经体现在合同和订阅规模上。**ServiceNow 在 2026 年第二季度公布，ServiceNow AI 年度合同价值超过 10 亿美元，Agent 部署在九个月内增加九倍；微软 Microsoft 365 Copilot 付费席位超过 3000 万，季度净新增席位较上一季度翻倍以上。年度合同价值反映 AI 产品合同按一年计算的规模，付费席位则指客户购买了多少个付费使用名额。两项指标分别从合同金额和用户采购数量说明，AI 已成为企业软件采购中的实际付费项目。
- **行业统计、客户案例和软件公司的经营数据，分别说明了 AI 商业化的不同进展。**标普 500 公司的部署与成效披露，说明越来越多企业开始使用 AI 并衡量效果；NHS England 的试验，让节省工作时间这一收益变得具体；

下载日志已记录，仅供内部参考

ServiceNow 的合同规模和 Copilot 的付费席位，则说明企业已经为相关产品支付费用。这几类证据相互补充，使 AI 应用的价值有了更完整的观察路径：先看企业是否使用，再看工作中取得了什么效果，最后看是否形成实际采购。已有资料表明，这些环节都已出现可观察的进展。

收费单位更贴近工作交付

- **AI 产品的收费方式，正在与客户购买的具体内容对应起来：**
  - 直接调用模型时，客户按输入、输出等 Token 用量付费；
  - 购买办公软件时，客户可以按付费使用名额订阅；
  - 使用业务 Agent 时，客户还可以按执行动作或合同约定的处理结果付费。
- 这些方式分别对应模型调用、软件使用和业务处理三类需求。对客户而言，价格开始更清楚地对应到所使用的功能或得到的服务；对厂商而言，也可以通过提供完整产品和业务服务形成收入。
- **按动作收费，让客户能够根据 Agent 执行的操作计算费用。** 例如，Salesforce Agentforce 每 10 万 Flex Credits 报价 500 美元，标准动作消耗 20 Credits，折合每动作 0.10 美元。Credits 是平台的计费额度，一个动作可以是查询、更新记录或执行流程中的某项操作。客户因此可以根据任务需要哪些操作，理解费用如何产生；一个完整任务包含的动作越多，其费用构成也随之变化。这种方式把模型调用组织成客户能够识别的软件功能。
- **结果收费进一步将价格与合同约定的处理结果对应。**例如，Fin 外接客服系统的计划为每月 49 美元，包含 50 次解决，额外可计费结果每次 0.99 美元。其结果定义既包含解决客户问题，也包含按配置程序完成的移交；解决情况区分用户确认与回答后未再请求帮助。同一会话按条款最多收取一次结果费用。这种方式将多步模型处理封装为客户可以采购的服务结果。

图表 10：AI 大模型收费模式

服务层次	已落地的收费方式	对应的客户采购内容
模型推理	按输入、输出、缓存等 Token 分别计费	推理资源与模型能力的调用
办公软件	按付费席位 + 使用量计费	工作工具的使用权及实际消耗
业务流程	标准业务流程折合 0.10 美元/次（Salesforce Agentforce）	查询、更新与流程执行
客服服务	计划内额度 + 额外 0.99 美元结果费（Fin）	合同约定的问题解决或程序移交

来源：Salesforce，Fin，中泰证券研究所

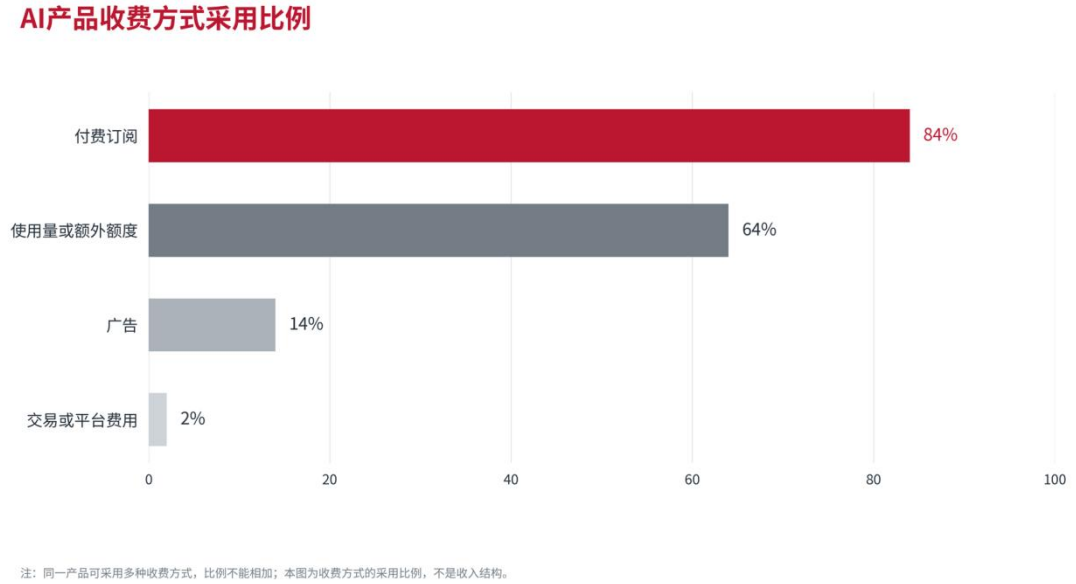
个人付费持续扩大，专业工具获得更强需求

付费用户规模扩张，存量用户支出持续提升

- **消费 AI 目前仍以用户直接付费（订阅与 token 额度）为主。** 根据 a16z 数据统计 44 款 AI 原生产品，其中 84%提供付费订阅，64%采用使用量或额外额度收费，14%采用广告，2%获得交易或平台费用。同一产品可以采用多种方式，因此这些比例不能相加。数据说明，现阶段多数产品仍依靠订阅和额度收费获得收入。

下载日志已记录，仅供内部参考

图表 11：AI 消费支出增长集中于最高支出群体



来源：a16z，中泰证券研究所

- **消费 AI 已经形成持续扩大的付费需求。** 根据 YipitData 数据显示，13%的付费用户会在购买一款 AI 产品的同时，再购买另一款 AI 工具。消费额最高的 1%用户贡献了全部可观测消费者 AI 支出的 19.5%，已高于消费额最低的 50%用户合计占比（16.6%）。截至 2026 年 8 月，排名前 1%的用户平均每月通过信用卡支出 903 美元用于人工智能（若计入工作用途，实际金额可能更高），且过去 18 个月累计增长 80%。相比之下，支出中位数用户每月仅花费 25 美元，且随时间推移几乎没有增长，两者差距十分明显。

图表 12：不同 AI 客户群体的支出指数

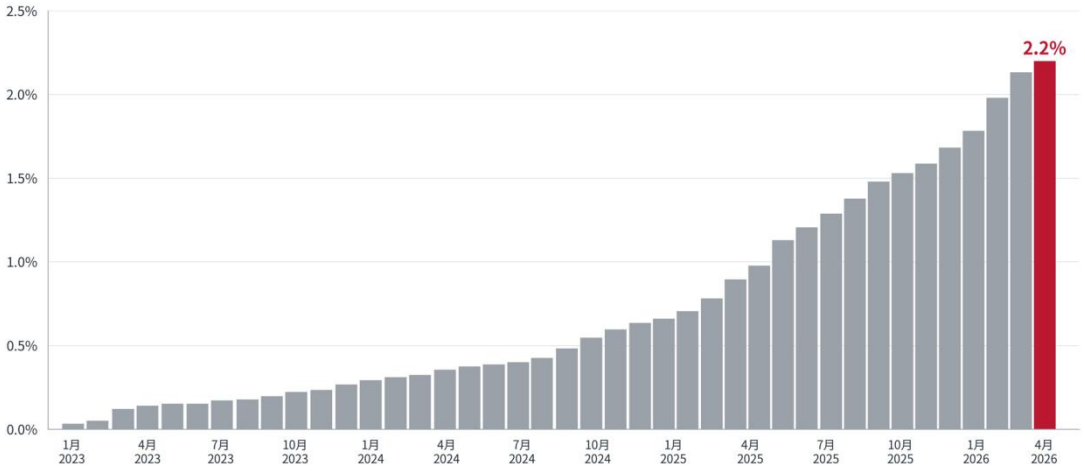


来源：YipitData，中泰证券研究所

- **家庭订购 AI 服务仍有非常大的商业空间。**PNC Research 数据显示，2026 年 5 月，约 2.2%的美国家庭有 AI 订阅付款记录；在付费家庭中，平均每月 AI 支出从 2024 年 5 月约 22 美元，提高至 2026 年 5 月约 31 美元，增幅约四成。家庭订阅 AI 服务仍有非常大的市场空间，AI 产品已经获得更多家庭月度预算，消费付费正在体现出金额上的增长。

下载日志已记录，仅供内部参考

图表 13：美国家庭订阅 AI 服务占比



来源：PNC Research，中泰证券研究所

- **消费者 AI 市场的增长，既体现在付费人数扩大，也体现在平均支出增加。**YipitData 观察到个人订阅占比上升，PNC 则观察到付费家庭的平均月支出提高。两项研究分别统计个人和家庭，可以各自说明付费范围和支出的变化。结合来看，消费 AI 已经拥有直接付费的用户基础，并在不同样本中呈现增长。这种付费需求的形成条件：当 AI 更有用、用户更容易负担其费用时，用户采用相关产品的动力也会增强。

通用助手扩大使用，专业工具服务具体工作

- **通用助手已经拥有大规模访问，专业产品则覆盖更多具体用途。**根据 AICPB 的 2026 年 9 月网站数据，ChatGPT、Gemini 和 Claude 月访问量分别为 62.1 亿次、27.6 亿次和 10.5 亿次；榜单同时覆盖代码、笔记、设计、音乐和客服等产品。访问次数反映网页上的使用活动，表明头部通用产品已经形成较大使用规模，也有多类工具服务于细分需求。个人订阅和专业工具支出，AI 应用已经同时出现广泛使用与直接付费，通用助手和专业工具各自形成了相应市场。

下载日志已记录，仅供内部参考

图表 14：全球 AI 网站访问量榜单前列产品

全球排名	产品名 AI产品榜	网站(web)分类 AICPB	09月上榜网站 Web访问量	09月上榜网站 环比变化
1	ChatGPT	AI ChatBots	6.21B	6.54%
2	New Bing	AI Search Engine	3.63B	-6.72%
3	Gemini	AI ChatBots	2.76B	2.28%
4	Claude | Anthropic	AI ChatBots	1.05B	7.85%
5	DeepSeek	AI ChatBots	458.01M	2.18%
6	Github Copilot	AI Code assistant	383.62M	-2.07%
7	Shop	E-COMMERCE	344.93M	-2.25%
8	Gemini Notebook	AI Notebook	214M	21.44%
9	Grok	AI ChatBots	210.92M	-3.22%
10	Notion AI	AI Notebook	181.79M	1.49%
11	Character AI	AI Character Generat	171.09M	-6.59%
12	豆包 | 字节	AI ChatBots	154.23M	-4.91%
13	JanitorAI	AI Character Generat	143.79M	-7.78%
14	纳米AI | 360	AI Agent	123.28M	3.17%
15	Google AI Studio	Model Inference (Clou	115.67M	-3.26%
16	Salesforce AI	AI Customer Support	114.11M	2.48%
17	Deepl	AI Translate Tools	99.26M	-3.55%
18	Perplexity AI	AI Search Engine	94.45M	-4.91%
19	Jambot	AI Design Tool	83.81M	1.90%
20	Suno	AI Music Generator	83.31M	5.13%
21	SpicyChat AI	AI Character Generat	75.51M	2.31%
22	Q-Chat	AI Tools for Educatior	75.15M	73.22%
23	Candy ai	AI Character Generat	64.58M	-2.59%
24	copilot | 微软	AI ChatBots	64.45M	-13.67%
25	zendesk	AI Customer Support	62.16M	-4.72%
26	Grammarly	AI Writer Generator	61.47M	22.94%
27	文心 | 百度	AI ChatBots	59.84M	5.36%
28	duolingo	AI Tools for Educatior	58.41M	-1.73%
29	polybuzz	AI Character Generat	55.65M	-13.34%
30	Hostinger	AI Code assistant	55.41M	2.08%

**数据截止 2026 年 9 月*

来源：AICPB，中泰证券研究所

下载日志已记录，仅供内部参考

- **国内 AI 应用发展通用问答和办公等专业用途。**根据 AICPB 的 2026 年 9 月网站数据中，DeepSeek 月访问量约 4.58 亿次，豆包约 1.54 亿次，WorkBuddy 约 1774 万次；WorkBuddy 当月网站访问环比增长 21.42%。腾讯同期开放 WorkBuddy 平台，首批引入超百家生态伙伴。这两类信息分别说明产品的网页使用活动和可接入能力的扩展：通用产品获得了较大访问规模，办公 Agent 也在通过平台合作增加功能，为用户处理具体工作提供更多选择。

图表 15：中国 AI 网站访问量榜单前列产品

国内排名	产品名 AI产品榜	网站(web)分类 AICPB	09月上榜网站 Web访问量	09月上榜网站 环比变化
1	DeepSeek	AI ChatBots	458.01M	2.18%
2	豆包 | 字节	AI ChatBots	154.23M	-4.91%
3	纳米AI | 360	AI Agent	123.28M	3.17%
4	文心 | 百度	AI ChatBots	59.84M	5.36%
5	Kimi | 月之暗面	AI ChatBots	38.62M	-25.89%
6	千问 | 阿里	AI ChatBots	31.05M	-11.53%
7	DeepSeek API	Model Inference (Clou	30.89M	-15.57%
8	百度文库	AI Writer Generator	20.5M	-15.12%
9	Workbuddy cn | 腾讯	AI Agent	17.74M	21.42%
10	百度搭子 | 百度	AI Agent	12.75M	2.46%

**数据截止 2026 年 9 月*

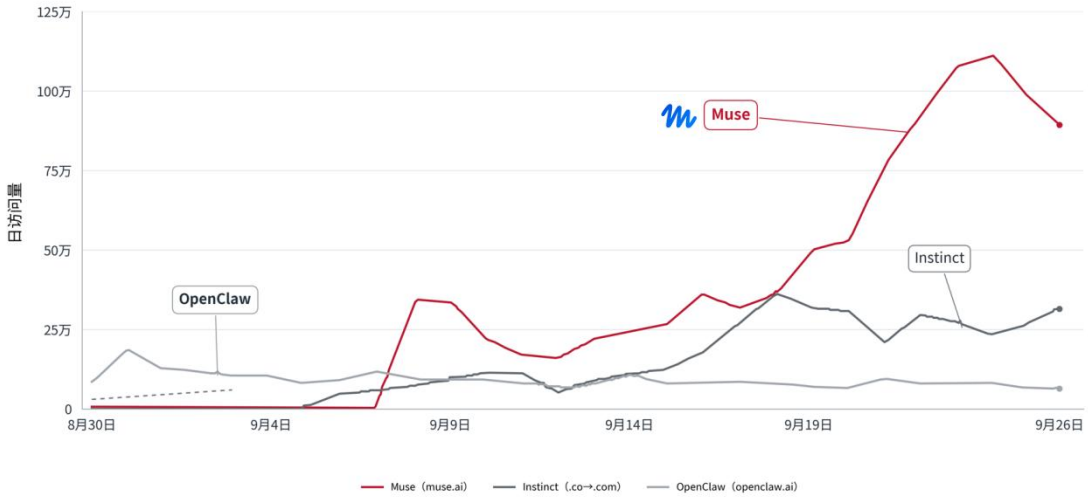
来源：AICPB，中泰证券研究所

Agent 走向长期协作，记忆与主动性拓宽应用边界

个人 Agent 访问量扩张，需求方向明确

- **个人 Agent 已获得较强的初始市场关注。**Similarweb 数据显示，2026 年 9 月 Muse 注册相关页面的日访问量一度达到百万次，Instinct 等产品也形成独立的访问规模。这一统计仅覆盖网页渠道。个人助手正在成为通用模型之外的一类新产品入口，其市场启动已经拥有可观察的需求信号。

图表 16：不同个人 Agent 访问量趋势



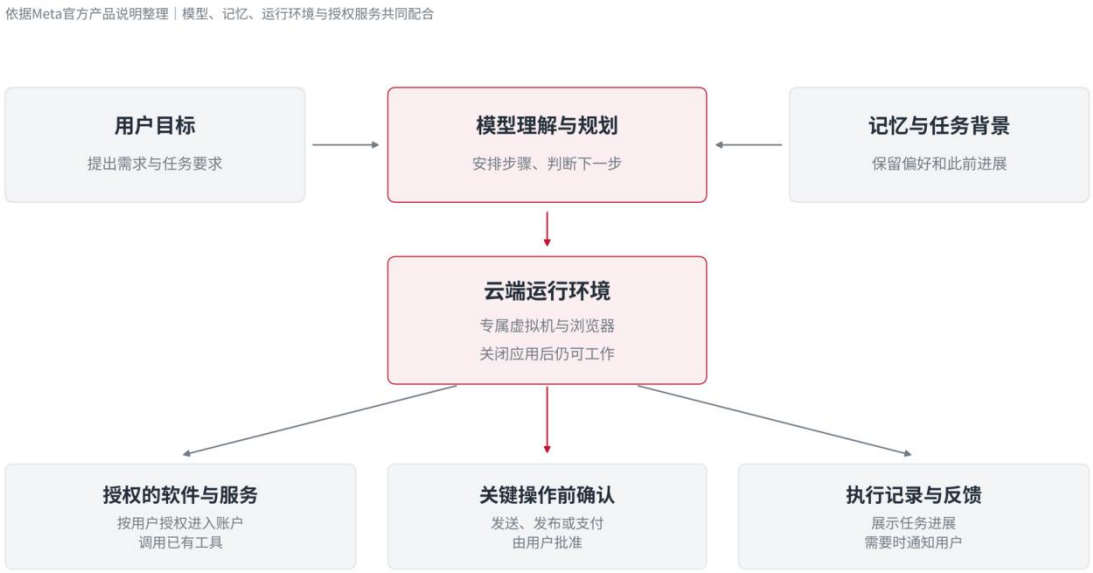
来源：Similarweb，中泰证券研究所

- **个人 Agent 能够持续做事，需要模型、运行环境和服务账户共同配合。**Muse 运行在带有浏览器的专属虚拟机中，可以理解为拥有一台用于执行任务的云端电脑。它结合用户授权的账户、记忆和任务背景开展工作，并能在用户关

下载日志已记录，仅供内部参考

闭应用后继续运行。模型负责理解目标和安排下一步，浏览器用于操作网页，授权账户用于进入相关服务，记忆则帮助保留此前的任务信息。这些功能组合起来，使助手可以围绕同一个目标持续处理多步工作。

图表 17：个人助手如何把用户目标变成持续执行



来源：Meta，中泰证券研究所

- **连接已有经营软件，使个人助手能够参与更具体的商业工作。**Muse 面向小企业提供 Shopify、QuickBooks、Canva、Stripe、Notion 和 Slack 等连接器，并接入 Facebook 及 Instagram 商业账户。连接器是让助手访问其他软件的工具，通过这些连接，店铺、财务、内容和营销等工作可以与用户原有的软件衔接。a16z 消费支出样本同时显示，自动化、开发和创作工具获得了高支出用户较多购买。两类资料分别反映产品功能与付费需求，说明专业工作已经成为个人 AI 助手可以服务的重要用途

模型能力与产品系统推动 Agent 走向长期协作

- **个人 Agent 开始承担需要持续跟进的事务。** 助手能够保存任务进度，在收到新消息或到达约定时间后继续处理，用户离开电脑后也可以保持跟进。这让用户能够把一件需要多次协调的事情交给 AI，减少反复交代和催促的工作。

下载日志已记录，仅供内部参考

图表 18：个人助手定时心跳与事件唤醒图解机制



来源：腾讯研究院，中泰证券研究所

- **记忆、主动性和权限管理，共同支撑个人助手办事。**记忆让助手了解用户偏好、保留此前进展；主动性让它从邮件、日历等信息中发现需要处理的事项，例如提醒即将截止的报名；权限管理则明确它可以做什么，并在付款、发送消息等关键环节安排用户确认。这些能力结合，使助手能够理解个人需求并持续推进。

图表 19：个人助手记忆系统图解机制



来源：腾讯研究院，中泰证券研究所

- **模型能力决定如何判断，产品系统负责把判断落实为行动。**系统保存记录、唤醒任务并连接软件，模型判断哪些信息值得记住、下一步使用什么工具，以及何时需要询问用户。模型与产品共同改进，才能让助手更稳定地完成复杂事务，也让更强的模型能力转化为用户可以使用的功能。

下载日志已记录，仅供内部参考

- **个人助手与办公 Agent 可以共享这些基础能力。**服务个人时，助手关注日程、偏好和生活安排；服务团队时，则关注项目进度、职责和协作记录。两类产品都需要记住事情进展、主动跟进，并在授权范围内执行，只是处理的信息和工作目标不同。

图表 20：个人助手与办公 Agent 的共同基础



来源：腾讯研究院，中泰证券研究所

- **商业化可以从已有的办公付费需求切入。**将个人助手与办公 AI 结合销售，以工作提效承接用户付费，再通过日常事务协助增加持续使用的价值。同时，按需启动运行环境、复用缓存和整理记忆，可以降低长期服务的成本，为这类产品持续运营提供支持。
- 根据 a16z 研究，长期运行 Agent 的机会，在于跨越数周或数月持续推进客户事务，例如保险理赔、抵押贷款办理及复杂企业销售。其描述的执行方式包括按需保存和检索背景、由事件或截止日期唤醒、协调不同参与者，并持续整理资料和推动后续步骤。**这为个人助手的记忆、主动性与持续执行提供了明确的业务解释：这些能力可以共同服务于一项需要长期协调的目标，使 AI 的应用范围延伸至完整事务的推进。**
- 当 Agent 参与更完整的客户流程，其对最终产出及收入增长的贡献更容易被观察和归因，供应商因而可能按工作成果或结果定价。**我们认为，在交付标准和贡献归属能够清晰界定的业务中，持续执行能力有望拓宽 Agent 的收费基础，使产品价值与客户获得的业务成果建立更直接的联系。**

图表 21：企业 AI Agent 的演进

	AI助手	智能体	长时运行智能体
	负责单项任务 “发送这封跟进邮件”	负责一次交互 “处理这位客户的退款来电”	负责最终结果 “最大化这位客户的回收收益”
时间跨度	秒级	分钟级	数周至数月
任务范围	一条消息	一次通话	全流程
预算来源	软件支出	人力支出	按成果分成

来源：a16z，中泰证券研究所

下载日志已记录，仅供内部参考

- **我们认为，个人助手向长期协作发展，有望同时拓宽模型需求与应用空间。**当 AI 能够记住用户背景、保存任务进度并持续跟进，一次使用就可以延伸为围绕同一目标的多次服务，用户也更有动力将日常事务交给助手处理。随着承接的事务增多，模型需要完成更多信息理解、行动判断和工具调用，带动持续的模型使用需求；产品则可以通过节省用户时间、提高任务完成效率，形成订阅和续费的基础。因此，模型能力、记忆与执行系统、软件生态的协同，将成为个人助手提升使用价值和推动商业化的重要因素。

投资建议

投资建议与标的

- **模型调用增长与个人 Agent 需求扩展，有望共同推动大模型产业商业化迈向纵深。我们认为，**随着底层模型能力提升、调用成本与缓存复用机制持续改善，AI 已经能够参与更多需要持续跟进和多步执行的实际工作。无论是在企业端通过节省工作时间转化为千万级付费席位与按动作/按结果收费合同，还是在个人端依托专属运行环境、事件唤醒、长期记忆库与商业连接器走向跨周期长期协作，Agent 对记忆、主动跟进和工具调用的高频需求，都有望进一步扩大底层模型的 Token 使用量，同时使应用产品的商业价值更多体现为“能否持续完成任务、为用户节省时间和提高交付效率”。
- 建议关注模型侧的**智谱、MiniMax**，重点跟踪模型迭代、调用量增长及商业收入兑现；个人 Agent 生态侧建议关注**腾讯、阿里巴巴**，重点观察腾讯在社交与办公场景中的产品整合，以及阿里巴巴围绕购物、出行等生活服务的执行能力。我们认为，模型能力与用户入口、业务服务的结合，有望拓宽个人 Agent 的应用范围，并为持续使用和付费转化提供支持。
- **垂类 Agent 与终端个人助手延伸标的**：建议关注**迈富时、国脉文化**。跟踪**迈富时**在企业级 Agent 相关产品与商业化场景的落地进展，以及**国脉文化**依托云智手机切入个人 AI 助手入口的协同潜力。

风险提示

行业政策与监管风险、技术研发不及预期风险、市场竞争加剧风险、盈利不及预期风险、测算偏差和数据更新不及时的风险等

- 全球 AI 行业监管政策持续演变，数据安全、隐私保护、算法合规等方面的监管要求可能趋严，若公司未能及时适应相关政策变化，可能影响业务开展。
- 大模型技术迭代速度快，若公司研发投入不足、核心技术人员流失或技术路线判断失误，可能导致产品竞争力下降，错失市场机会。
- 全球 AI 大模型市场参与者增多，包括科技巨头与新兴企业，市场竞争日趋激烈，可能导致公司用户流失、市场份额下降或盈利能力承压。
- 行业目前属于新兴阶段，若用户付费转化不及预期、成本控制效果不佳或业务拓展缓慢，可能导致盈利受影响。
- 本次测算依赖既定假设与存量数据，结果或存在偏差；底层数据存在更新滞后风险，基本面变化可能导致结论失效。

下载日志已记录，仅供内部参考

投资评级说明

	评级	说明
股票评级	买入	预期未来 6~12 个月内相对同期基准指数涨幅在 15%以上
	增持	预期未来 6~12 个月内相对同期基准指数涨幅在 5%~15%之间
	持有	预期未来 6~12 个月内相对同期基准指数涨幅在-10%~+5%之间
	减持	预期未来 6~12 个月内相对同期基准指数跌幅在 10%以上
行业评级	增持	预期未来 6~12 个月内对同期基准指数涨幅在 10%以上
	中性	预期未来 6~12 个月内对同期基准指数涨幅在-10%~+10%之间
	减持	预期未来 6~12 个月内对同期基准指数跌幅在 10%以上
备注：评级标准为报告发布日后的 6~12 个月内公司股价（或行业指数）相对同期基准指数的相对市场表现。其中 A 股市场以沪深 300 指数为基准；新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以摩根士丹利中国指数为基准，美股市场以标普 500 指数或纳斯达克综合指数为基准（另有说明的除外）。		

重要声明

中泰证券股份有限公司（以下简称“本公司”）具有中国证券监督管理委员会许可的证券投资咨询业务资格。本报告仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为客户。

本报告基于本公司及其研究人员认为可信的公开资料或实地调研资料，反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响。本公司力求但不保证这些信息的准确性和完整性，且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，可能会随时调整。本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。本报告所载的资料、工具、意见、信息及推测只提供给客户作参考之用，不构成任何投资、法律、会计或税务的最终操作建议，本公司不就报告中的内容对最终操作建议做出任何担保。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。

市场有风险，投资需谨慎。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。

投资者应注意，在法律允许的情况下，本公司及其本公司的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。本公司及其本公司的关联机构或个人可能在本报告公开发布之前已经使用或了解其中的信息。

本报告版权归“中泰证券股份有限公司”所有。事先未经本公司书面授权，任何机构和个人，不得对本报告进行任何形式的翻版、发布、复制、转载、刊登、篡改，且不得对本报告进行有悖原意的删节或修改。