



# 计算机行业研究

买入（维持评级）

行业周报

证券研究报告

计算机组

分析师：刘高畅（执业 S1130525120005）

liugaochang@gjzq.com.cn

## Coding 以外的下一个应用是什么？

**模型迭代由单点能力提升转向研发、工具与生产反馈共同加速。**Astra 所体现的变化不只在评测分数，而在于模型开始更稳定地调用浏览器、代码和专业软件，并参与数据处理、自动评测与实验执行。代码测试、界面操作和结构化业务流程能够提供可验证结果，生产环境中的失败记录又可以反向暴露模型短板，使训练、评测和产品应用之间形成更紧密的反馈。当前仍不能将这一阶段等同于完整的递归式自我改进，Astra 更多是强化学习、工具执行和可验证任务环境结合的阶段性成果；但模型竞争已从单次回答质量扩展至长任务成功率、执行时延、推理效率和单位任务成本。随着研发侧与产品侧的边界逐步收窄，真实任务反馈可能成为后续模型训练和能力修正的重要输入，后续模型在推理、Computer Use、多模态和持续执行方面的进展值得关注。

**自定义 AI 与企业 AI 正在悄然渗透。**Muse 等产品表明，消费级 Agent 正在把个人数据、长期记忆、权限控制和 Computer Use 整合至同一入口；企业端则通过数据库、办公软件和业务系统连接形成专属工作流。用户能够自行组织数据、调用接口并设计任务流程，使应用形态更加个性化。市场未必出现移动互联网时期依靠统一界面覆盖所有用户的超级应用，但大量分散任务可能先在后台形成模型调用、Token 消耗和软件收入，再逐步体现为用户可感知的产品变化。教育、金融、法律和咨询行业已经由检索、摘要和内容生成，向资料核验、跨系统协同及受控执行延伸。未来一年，高频、规则相对明确、结果可复核且责任边界清晰的任务有望率先放量。判断应用景气度应更多跟踪生产任务量、付费客户扩张、功能增购、续费率、人工接管率和单位任务成本，而不是单一产品的下载排名。

**宏观扰动不改算力系统升级，产业研究应回到结构性增量。**能源价格、利率和地缘因素仍会影响市场风险偏好及云厂商建设节奏，前期油价快速冲高后的大幅波动也说明宏观博弈已经进入更复杂阶段。但短期价格信号与产业订单周期并不完全一致，研究上应区分资本开支节奏变化和技术规格升级。大模型正向更长上下文、更多工具调用、多模型协同和持续运行演进，推理侧的数据交换、功率密度与可靠性要求同步提高。算法效率提升可能降低单次调用成本，但 Agent 引入的规划、执行、重试和校验会增加单项任务的计算量，应用边界扩张仍可能推动推理总量增长。由此带来的增量不只体现在芯片数量，也体现在高速 PCB、光互连、电源转换和液冷系统的单位价值量。即使资本开支增速出现波动，只要网络速率、机柜功率和系统复杂度持续提升，关键部件仍具备相对独立的升级周期，产业判断可以更加关注产品代际、客户份额、认证进度、良率和交付能力。

### 相关标的

大模型&云：智谱、MiniMax、阿里巴巴、腾讯控股、金山云、百度集团、优刻得、首都在线、网宿科技、云赛智联、青云科技等。海外算力：东山精密、工业富联、胜宏科技、江海股份、中际旭创、中钨高新、蓝思科技、东阳光、火炬电子、三环集团、欧科亿、天孚通信、鼎泰高科、新易盛、领益智造、兆易创新、鹏鼎控股、唯科科技、天岳先进、大普微、源杰科技、麦格米特、景旺电子、英维克、京东方等；英特尔、SK 海力士、Lumentum、闪迪、铠侠、美光。国内算力：寒武纪、海光信息、长鑫科技、东阳光、利通电子、中芯国际、华虹半导体、华达科技、摩尔线程、天数智芯、沐曦股份、壁仞科技、协创数据、禾盛新材、华丰科技、杰华特、京基智农、中科曙光、浪潮信息、利扬芯片、胜蓝股份、华勤技术、国科微、中国长城、晶科科技、罗曼股份、盈峰环境、芯原股份、晶科科技、亿田智能、豫能控股、星环科技、鸿日达、盛视科技、神州数码、润泽科技、大位科技、润建股份、奥飞数据、瑞晟智能、科华数据、潍柴重机、欧陆通、杰创智能、奥尼电子。AI 应用：达梦数据、深信服、卓易信息、星环科技、广立微、绿盟科技、网宿科技、海天瑞声、合合信息、德才股份、美年健康、真爱美家、中控技术、迈富时、金蝶国际、康众医疗、用友网络、聚水潭、阜博集团、范式智能、汇量科技、金山办公、奔图科技、中国软件

### 风险提示

行业竞争加剧的风险；技术研发进度不及预期的风险；特定行业下游资本开支周期性波动的风险。



## 内容目录

|                                               |    |
|-----------------------------------------------|----|
| 一、自定义 AI 是这轮 AI 与互联网应用最大的差异.....              | 3  |
| 1.1 史上最快扩散但 AI 渗透仍在早期.....                    | 3  |
| 1.2 自定义 AI 改变爆款范式：Muse 展示消费级 Agent 平台雏形 ..... | 3  |
| 1.3 以 Coding 场景为样本：同一技术，千人千面 .....            | 4  |
| 二、模型迭代与 Computer Use 打开能力新边界.....             | 5  |
| 2.1 RSI 信号开始出现：Astra 是阶段性验证.....              | 5  |
| 2.2 Computer Use 把模型从回答系统推进到执行系统 .....        | 6  |
| 2.3 供给侧投入仍处高位：能力与成本同步改善.....                  | 6  |
| 三、细分行业潜能：下一个被“自定义化”的垂直领域.....                 | 7  |
| 3.1 教育行业：师生双侧渗透，AI 导师由单点辅助走向教学工作流.....        | 7  |
| 3.2 金融行业：从单点工具走向可验证任务闭环 .....                 | 7  |
| 3.3 法律与合规：权威内容和审计链支撑收入提速 .....                | 8  |
| 3.4 咨询行业：内部知识平台形成持续反馈 .....                   | 9  |
| 3.5 未来展望：垂直行业由试点验证走向任务量扩张 .....               | 10 |
| 四、算力结构性机会：回归产业与系统升级.....                      | 11 |
| 4.1 宏观扰动深化：回归产业并关注模型演进.....                   | 11 |
| 4.2 高速互连：PCB 与光通信受益系统复杂度提升 .....              | 11 |
| 4.3 电源与散热：功率密度成为数据中心硬约束.....                  | 11 |
| 五、相关标的.....                                   | 12 |
| 六、风险提示.....                                   | 12 |

## 图表目录

|                                                    |   |
|----------------------------------------------------|---|
| 图表 1：ChatGPT 周活跃用户数：18 个月内从 4 亿增至 10 亿 .....       | 3 |
| 图表 3：Muse 运行在配备浏览器的虚拟机上，聊天具有对话感.....               | 4 |
| 图表 4：AI Coding 渗透率：工具层已接近饱和，但用法高度碎片化.....          | 5 |
| 图表 5：Astra 在科研、复杂终端任务等方面表现突出 .....                 | 6 |
| 图表 6：Anthropic 于 2026 年 5 月发布 10 类金融 Agent 模板..... | 8 |
| 图表 7：2025 年法律、政府相关工作 AI 使用情况.....                  | 9 |
| 图表 8：麦肯锡内部平台 Lilli 使用数据 .....                      | 9 |

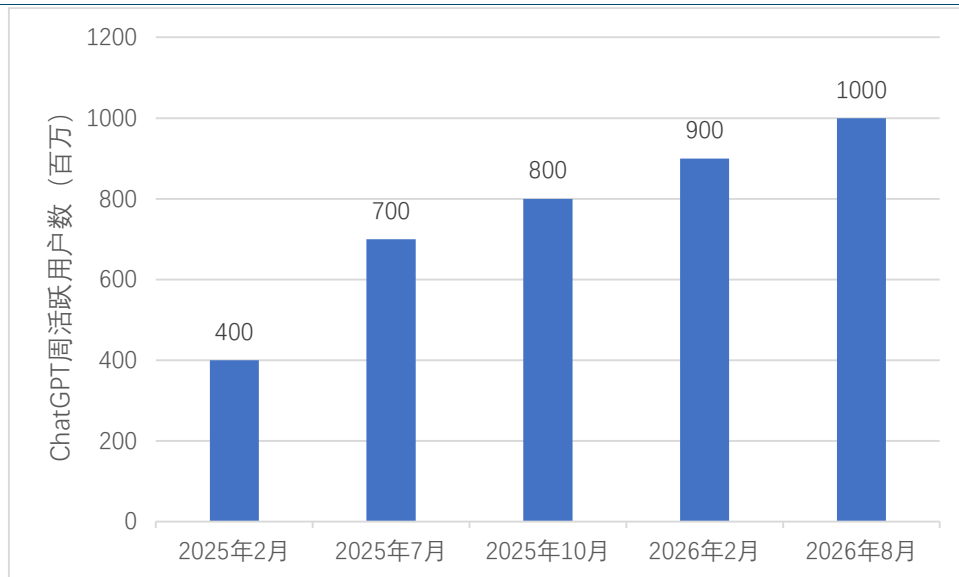


## 一、自定义 AI 是这轮 AI 与互联网应用最大的差异

### 1.1 史上最快扩散但 AI 渗透仍在早期

生成式 AI 在用户规模上的扩散速度史无前例。截至 2026 年 8 月，ChatGPT 全球周活跃用户已攀升至 10 亿，使用过生成式 AI 工具的全球人口累计超过 12 亿。在美国，18–64 岁人群中生成式 AI 三年普及率超过五成，是个人电脑发布三年后（19.7%）的近三倍、互联网商业化三年后（30.1%）的近两倍，体现出本轮技术扩散的高弹性。

图表1：ChatGPT 周活跃用户数：18 个月内从 4 亿增至 10 亿



来源：OpenAI, Instant Press, 国金证券研究所

在全球视角下，微软 AI 经济研究所数据显示，2026 年一季度全球 15 至 64 岁劳动年龄人口中，生成式 AI 工具使用率达到 17.8%，较 2025 年下半年提升 1.5%。同期，阿联酋和新加坡的 AI 使用率分别达到 70.1%和 63.4%，挪威、爱尔兰和法国分别为 48.6%、48.4%和 47.8%，美国为 31.7%。以阿联酋和新加坡已经验证的渗透水平作为参照，即便是美国等发达经济体，生成式 AI 仍有较大的提升空间；对目前使用率仅为 15.4%的全球南方而言，AI 尚未形成系统性应用，未来渗透率理论上仍有接近四倍的提升空间。

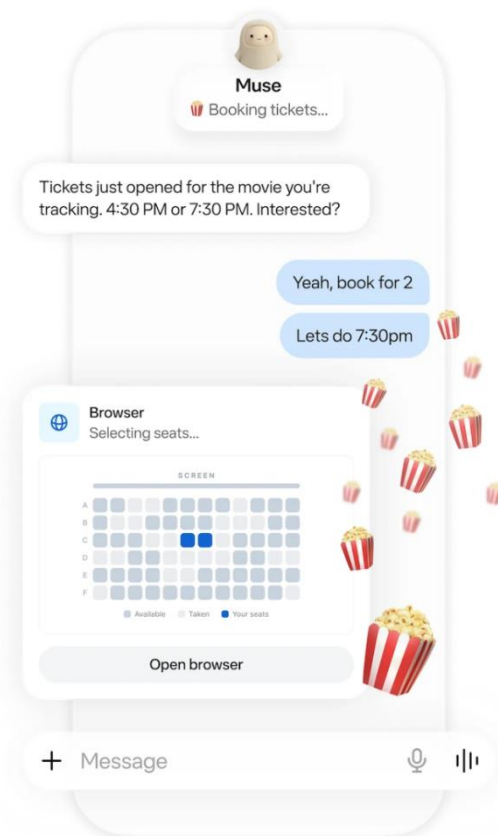
### 1.2 自定义 AI 改变爆款范式：Muse 展示消费级 Agent 平台雏形

生成式 AI 的核心价值取决于用户的任务、数据、权限和工作流程。同一底座模型在不同个人与企业中承担的功能并不相同：个人端需要连接邮箱、日历、浏览器、支付与内容平台，企业端则需要接入数据库、客服系统、内部知识库和审批流程。产品价值由通用回答转向持续执行，入口形态也由单一应用转向可组合的 Agent 平台。

Meta 于 2026 年 9 月 8 日推出 Muse。Muse 稳定运行在专属安全虚拟机中，配备浏览器并连接用户授权的日常服务。它能够完成电话预订、填写表单、处理客户服务等多项日常而琐碎的任务，且能够在用户关闭应用后继续运行，在任务有进展或需要批准时主动通知用户。该产品将长期记忆、个人数据、权限控制和 Computer Use 整合在同一消费级界面，采用 3D 卡通形象 and 对话式交互，体现了“自定义 AI”从开发者工具向大众产品迁移的方向。



图2: Muse 运行在配备浏览器的虚拟机上, 聊天具有对话感



来源: Muse, 国金证券研究所

因此, AI 应用的商业化不宜仅用“是否出现统一爆款”衡量。个人数据和工作流具有天然差异, Agent 越深入真实任务, 产品越可能呈现千人千面。市场感知上会缺少移动互联网时期的集中爆款, 但后台的调用频次、Token 消耗、数据处理量和按任务计费收入仍在快速增长。Salesforce 二季度 Data 360 摄取 104 万亿条记录, 同比增长 355%, Agentforce 与 Slack 累计完成 70 亿个 Agentic Work Units, 说明个性化任务并不妨碍总需求形成规模。

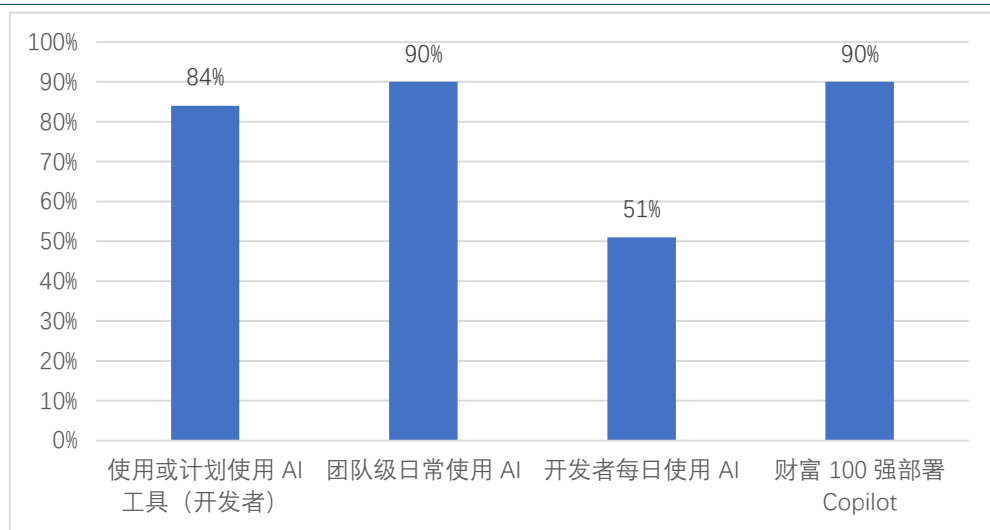
从产品结构看, 自定义 AI 至少包含四个相互依赖的层次: 底座模型负责理解与推理, 数据层提供个人或企业上下文, 连接器与权限系统决定可访问范围, 执行环境负责把结果写回真实业务。模型能力提升会扩大可完成任务的边界, 但稳定性取决于身份认证、授权撤销、状态保存、异常恢复和人工确认。平台厂商可通过统一模型与连接器获得规模效应, 垂直软件则依靠行业数据、流程模板和交付责任形成差异化定价。

### 1.3 以 Coding 场景为样本: 同一技术, 千人千面

开发者群体是当前 AI 渗透最深的用户群之一, 90% 的开发者日常工作中经常使用至少一种 AI 工具, 74% 使用专门的 AI 编程工具。但每个人的用法仍差别很大, 没有任何单一工具统治市场, 59% 的开发者并行使用三种以上 AI 编程工具, 开发者往往根据代码补全、调试、代码库理解和自动化执行等任务组合使用 Github Copilot、Cursor、Claude Code 等不同工具。



图3: AI Coding 渗透率: 工具层已接近饱和, 但用法高度碎片化



来源: Stack Overflow 2025, Google DORA 2025, JetBrains, Microsoft, 国金证券研究所

尽管工具席位渗透率已处于高位, 开发者对 AI 输出质量的信任仍然有限。调查显示, 仅 29% 的开发者信任 AI 输出的准确性, 46% 的开发者表示不信任; 66% 的开发者将“AI 解决方案几乎正确但并不完全正确”视为主要困扰, 45% 的开发者认为调试 AI 生成代码耗时更多, 这说明当前 AI 主要解决了代码生成、示例编写和简单重构等局部任务, 开发者仍需承担代码审查、测试验证和系统集成工作。从技术演进看, Coding 场景的下一阶段重点在于 Agent 化。开发者需要的并非单纯的代码补全, 而是能够理解整个代码库, 调用终端、文件系统和测试工具, 拆解任务并持续执行的完整开发系统。

因此, AI coding 渗透远未到位: 当前渗透的是“工具席位”, 尚未渗透的是“可信的工作流”, 这一结论也适用于教育、金融和法律等行业。

## 二、模型迭代与 Computer Use 打开能力新边界

### 2.1 RSI 信号开始出现: Astra 是阶段性验证

RSI (递归式自我改进) 的产业意义, 在于 AI 开始参与数据生成、评测、代码开发、实验执行和结果分析, 使研发循环中的部分环节缩短。RSI 的判断需要区分“模型参与研发”与“模型自主完成研发”。当前更明确的变化是, 代码执行、自动评测、数据合成和实验分析正在被纳入模型训练流程, 使研究人员可以更快发现失败模式并补充训练信号; 模型仍需依赖人类设定目标、选择评测和控制算力预算。产业上应关注研发周期是否缩短、单位实验成本是否下降, 以及新能力从研究版本扩散到主线模型的时间。

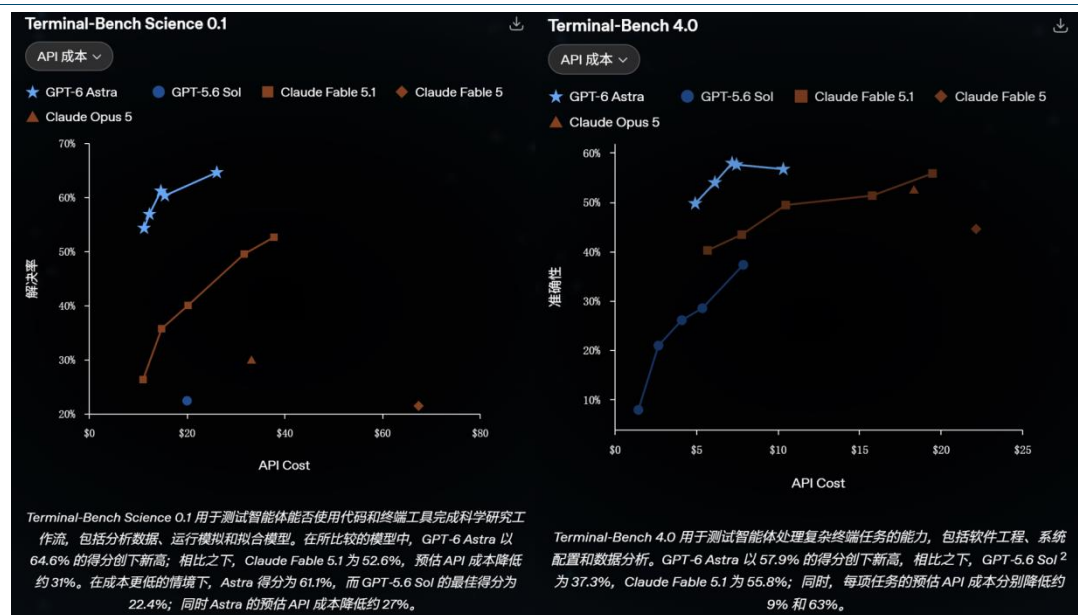
OpenAI 公开材料显示, GPT-6 Astra 综合了预训练、强化学习和对齐研究, 且当前进步速度可能延续到递归式自我改进。从可观测能力看, Astra 在 OSWorld 2.0 上取得 72.6%, 高于 GPT-5.6 Sol 的 65.7%, 并在模拟中以约 40 分钟完成单任务, 较 Sol 约 75 分钟缩短 47%。开发者文档还新增异步工具调用、中途指令调整和跨软件多步骤执行。据此判断, Astra 更像强化学习、可验证任务环境和工具执行共同作用的阶段性载体, 而非模型迭代的终点, Astra 到后续主线模型 (如 GPT-6 SOL, O/A) 的效率扩散仍值得持续跟踪。

模型迭代正在形成三类相互连接的反馈。第一类来自可验证任务环境: 代码测试、浏览器操作和结构化业务流程能够自动判断结果是否正确, 为强化学习提供高质量信号。第二类来自研发过程: 模型参与数据清洗、评测设计、代码实现和实验分析, 缩短研究人员的试错周期。第三类来自生产环境: 真实用户的任务失败、人工修改和权限中断暴露模型短板, 再回流到评测与产品设计。Astra 在 Computer Use 和长任务上的提升, 说明这三类反馈已开始反映到产品能力, 但尚未构成完全自主的模型研发闭环。

这一变化使得模型压缩输出、缓存和推理优化会降低单位 Token 价格, 另一方面, Agent 任务包含界面状态、工具结果、重试、校验和多 Agent 协作, 单次任务的计算量通常高于一次问答。只要单位任务成本下降速度快于需求增长, 应用范围就会扩张, 总 Token 与全产业收入仍可能增长。



图表4: Astra 在科研、复杂终端任务等方面表现突出



来源：OpenAI，国金证券研究所

## 2.2 Computer Use 把模型从回答系统推进到执行系统

Computer Use 推动模型由信息生成工具向任务执行系统演进。模型不仅需要理解用户指令，还要识别页面状态，调用鼠标、键盘、浏览器或代码工具，并在不同软件之间连续操作，模型竞争因此由回答质量逐步延伸至任务完成率、执行时延、权限边界和异常处理能力。

在企业场景中，Agent 的实际价值高度依赖执行深度，即其能够调用的上下文、数据权限和工具链范围。输入上下文长度每增加 1%，输出质量和长度平均提升约 0.38%。对于需要连续追踪、多源汇总和反复核验的任务，模型如果只能访问公开网页，通常仍停留在信息辅助阶段；接入企业数据库、业务系统和付费数据后，才可能理解历史状态、识别任务约束，并将分析结果写回真实业务流程。身份认证、最小权限、授权撤销和操作留痕也因此成为 Agent 落地的重要基础设施。具备行业数据、系统入口和工具编排能力的软件厂商，能够在同一工作流中积累用户修改、任务失败和人工复核数据，逐步形成“执行—反馈—优化—扩大使用”的业务循环。

Computer Use 的产品化通常包括感知、规划、执行和校验四个环节。模型先识别界面与任务状态，再拆解步骤并选择工具；执行过程中需要处理登录验证、弹窗、加载延迟、页面变化和权限中断，完成后还要核对数据是否被准确写入目标系统。任务链越长，单步错误越容易累积，因此生产环境的评价指标需要从回答准确率扩展到端到端成功率、平均重试次数、人工接管率、审计完整性和单位任务成本。异步工具调用和中途指令调整可以缩短等待时间，也允许用户在长任务中及时纠偏。

## 2.3 供给侧投入仍处高位：能力与成本同步改善

从供给端看，全球 AI 产业正在进入显著的重资产周期。2026 年全球头部云厂商继续提高 AI 基础设施投入。微软 2026 财年第四季度资本开支为 410 亿美元，其中约三分之二用于 CPU、GPU 等短寿命资产；Meta 预计 2026 年资本开支约 1,300 亿至 1,450 亿美元；2026 年 6 月 Alphabet 通过发行 A 类股、C 类股及强制可转换优先股的组合方式完成融资，净募资 496 亿美元，重点投向扩大 AI 基础设施与全球算力的资本开支。

模型效率也在改善。GPT-5.6 Sol 官方定价为每百万输入 Token 4 美元、输出 Token 20 美元，较上一代旗舰的输入和输出价格分别降低 20% 和 33%。Astra 虽然单 Token 价格更高，但在评测中依靠更少输出 Token 和更短任务时间降低单任务成本。完成同样的任务，GPT-5.6 Sol 花费 75 分钟而 Astra 仅花费 40 分钟，估算的单任务成本比前代低约 57%。能力提升与单位任务成本下降并行，将扩大可经济部署的任务边界。

继而，供给侧的关键变量从单纯的 GPU 数量扩展到整机柜功率、PCB 层数、液冷和供电效率等方向。即便未来资本开支增速从高位回落，单位系统价值量上升仍可能支撑部分环节的收入增长。



### 三、细分行业潜能：下一个被“自定义化”的垂直领域

#### 3.1 教育行业：师生双侧渗透，AI 导师由单点辅助走向教学工作流

教育行业，师生两端的渗透正在同步推进。截至 2026 年 6 月，全球约 92% 的学生和 88% 的教师已在教学环节中使用 AI 工具。国内方面，2025 年发布的中小学教师人工智能应用调研报告显示，约 75%—81% 的教师有 AI 使用经历，每日高频使用者比例达 26.2%，AI 已覆盖备课、授课、学练、考试、评价、管理等六个核心教学环节，当前主要任务仍集中在教案撰写、课件制作和资料检索等效率提升型场景。

在具体产品层面，海外与国内均已出现具备代表性的“AI 导师”与自适应学习平台。Khan Academy 推出的 Khanmigo 在 2024—2025 学年服务约 200 万学生、教师和家长，累计交互超过 1.08 亿次，工作日期间日均交互量约 26.9 万次；在接入学生既往练习数据后，下一道练习题的正确率平均提升 6.1 个百分点，展示了基于学习轨迹的个性化辅导效果。国内《2026 年中国 GenAI+教育行业发展报告》测算，2025 年中国校内教育信息化经费约 5,515 亿元，GenAI 产品在高校与中小学采购项目中的覆盖率约 20%—35%，在成人与 K12 C 端教育市场中的覆盖率约 10%—20%，GenAI+教育产品与服务总规模约 3,442 亿元，预计到 2028 年仍将以约 37% 的年复合增速扩张。

#### 3.2 金融行业：从单点工具走向可验证任务闭环

金融行业具备数据基础较好、业务流程标准化和付费意愿较强等特点，是 AI 由通用问答进入专业工作流的重要场景。2025 年中国金融行业生成式 AI 市场规模约 17.4 亿元，同比增长 90.4%。目前，AI 应用已覆盖智能风控、贷前贷后评估、反洗钱监测、投研内容生产和客户服务等环节，应用重点也开始由信息检索和内容生成转向可交付、可复核的业务任务。

高责任行业的落地通常遵循由辅助到执行的渐进路径。Morgan Stanley 面向理财顾问推出的 AI Assistant 已被 98% 的顾问团队采用，后续推出的 Debrief 可在客户授权下记录会议、提炼行动项、生成待审核邮件并写入 Salesforce。该模式保留人工审核和客户授权，同时将模型接入客户管理与业务记录系统，能够压缩资料检索、会议整理和信息录入时间，为进一步进入分析与执行环节建立基础。

金融 Agent 正在形成更加标准化的任务模板。Anthropic 于 2026 年 5 月发布 10 类金融 Agent 模板，覆盖投行材料制作、KYC 文件筛查、估值审阅和月末结账，并通过 Excel、PowerPoint、Word 及数据连接器承接多步骤任务。金融流程通常具有明确输入、固定规则和可核验输出，估值结果可以与财务数据交叉验证，KYC 判断可以依据合规条款复核，月末结账也可通过账务勾稽检查。因此，金融行业有条件形成类似 Coding 的反馈循环，任务过程被记录，错误能够定位，人工修改可以转化为新的评测与优化信号。

国内垂直金融 AI 产品也在由单点工具向平台化发展。Wind Alice 依托长期积累的金融数据，连接多类 MCP 工具与智能体，可完成可比公司筛选、组合构建和情景分析，并输出附带数据来源的可编辑 PPT 和 Excel 底稿。Alpha 派从会议纪要切入，逐步扩展至公司一页纸、业绩点评、调研大纲和主题选股等场景；同花顺则通过 iFinD MCP 和技能包，将金融数据、研究方法、业务流程封装为可调用模块。产品形态的变化说明，竞争重点正从单一模型能力转向数据质量、工具覆盖、任务编排和结果可追溯性。

未来，金融 AI 的主要增量可能由席位数量转向任务调用量。随着 Agent 进入投研、风控和运营流程，同一用户每天可以发起多项任务，带动 Token 消耗、数据查询和软件调用同步增长。对于金融软件厂商而言，真正的竞争壁垒仍在于数据授权、合规审计、业务系统入口以及对交付结果承担责任的能力。


**图表5: Anthropic 于 2026 年 5 月发布 10 类金融 Agent 模板**

| Function                 | Agent                              | What it does                                                                            |
|--------------------------|------------------------------------|-----------------------------------------------------------------------------------------|
| Coverage & advisory      | <a href="#">Pitch Agent</a>        | Comps, precedents, LBO → branded pitch deck, end to end                                 |
|                          | <a href="#">Meeting Prep Agent</a> | Briefing pack before every client meeting                                               |
| Research & modeling      | <a href="#">Market Researcher</a>  | Sector or theme → industry overview, competitive landscape, peer comps, ideas shortlist |
|                          | <a href="#">Earnings Reviewer</a>  | Earnings call + filings → model update → note draft                                     |
|                          | <a href="#">Model Builder</a>      | DCF, LBO, 3-statement, comps — live in Excel                                            |
|                          | <a href="#">Valuation Reviewer</a> | Ingests GP packages, runs valuation template, stages LP reporting                       |
| Fund admin & finance ops | <a href="#">GL Reconciler</a>      | Finds breaks, traces root cause, routes for sign-off                                    |
|                          | <a href="#">Month-End Closer</a>   | Accruals, roll-forwards, variance commentary                                            |
|                          | <a href="#">Statement Auditor</a>  | Audits LP statements before distribution                                                |
| Operations & onboarding  | <a href="#">KYC Screener</a>       | Parses onboarding docs, runs the rules engine, flags gaps                               |

来源：Anthropic, Github, 国金证券研究所

### 3.3 法律与合规：权威内容和审计链支撑收入提速

法律与合规具有高价值、文本密集、流程相对标准化和结果可复核等特征，是 AI 在非编程领域较早形成专业工作流的行业之一。2026 年调研显示，约 41% 的律师事务所和 47% 的企业法务部门已经在工作使用生成式 AI，较 2025 年的 28% 和 23% 明显提升。目前，应用主要集中在法律检索、文件摘要、合同审查、简报起草和法律文书初稿等环节，AI 正在由外围的信息辅助工具逐步进入法律服务链条的核心流程。

法律任务也具备形成数据反馈循环的条件。检索结果、合同条款和文书初稿均可由律师逐项复核，人工修改能够沉淀为新的评测样本；判例、法规和专业内容库则可以用于校验引用及结论。试验显示，RAG 法律工具与推理模型在六类法律任务中的五类显著提高生产率，提升幅度约 50%—130%，并改善多类法律文书质量，其中 RAG 工具没有带来更多幻觉。随着模型具备更长上下文和 Computer Use 能力，产品可进一步覆盖案件材料归档、跨文档比对、合同条款标记、风险提示及审批流转。

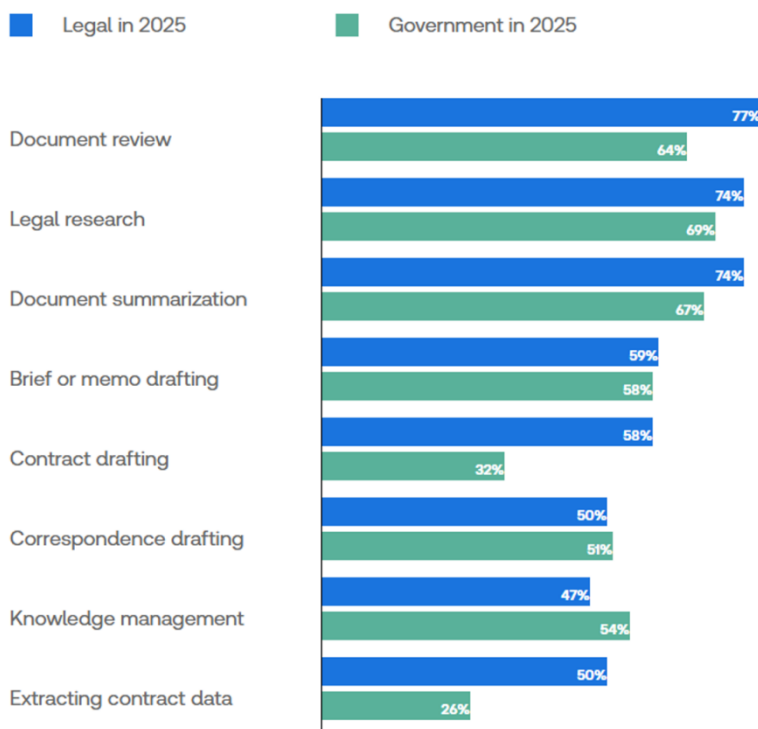
法律 AI 产品的竞争重点已从通用模型能力转向权威内容、工作流和可验证输出。LexisNexis 的 Lexis+ AI 依托长期积累的判例、法规和法律内容，提供对话式检索、法律写作、摘要与文档分析，并将生成结果与可追溯的资料库绑定，便于律师核查引用。Thomson Reuters 推出的新一代 CoCounsel Legal，则强调复杂法律任务、多轮推理和深度研究能力，并逐步与 Westlaw 等专业内容和业务工具整合。Harvey AI 主要面向律所和企业法务团队，通过连接客户内部文档与法律数据，将能力扩展至尽调、合同分析和多步骤法律工作流。2026 年 3 月，Harvey 完成 2 亿美元融资，投后估值约 110 亿美元，反映资本市场对法律垂直 Agent 商业空间的关注。

收入端已经出现一定验证。Thomson Reuters 披露，2026 年第二季度法律专业业务有机收入增长 10%，经常性收入增长 9%，主要由 Westlaw、CoCounsel 及相关产品推动；公司“Big 3”业务有机收入同样增长 10%。这一表现说明，法律 AI 正在由试验功能转化为订阅、续费和增



购，并开始对成熟应用软件公司的收入增速形成贡献。

图6：2025 年法律、政府相关工作 AI 使用情况



来源：Thomson Reuters，国金证券研究所

法律行业的壁垒不只在模型。权威内容、客户历史文档、工作底稿、权限体系、引用追溯和职业责任共同构成产品护城河。模型升级会提高可自动化任务比例，但数据与流程资产决定客户能否安全地扩大部署。

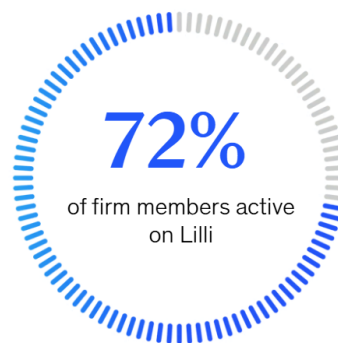
### 3.4 咨询行业：内部知识平台形成持续反馈

咨询行业以信息搜集、知识检索、数据分析和材料制作为主要生产环节，工作成果又大量沉淀为行业数据库、项目底稿、访谈纪要和方法论模板，因此具备较好的 AI 应用基础。相比通用搜索，接入内部知识库的模型能够在历史项目中定位相关案例，综合多份资料并生成带有来源的初步分析，帮助顾问减少重复性的信息整理工作，将更多时间投入问题定义、客户沟通和方案判断。

McKinsey 内部平台 Lilli 已经提供了较明确的使用数据。Lilli 的活跃覆盖率达到 72%，员工在知识检索与综合环节节省时间最高约 30%，每月回答超过 50 万条提示。其应用范围不仅包括资料查询，还延伸至专家识别、访谈准备、行业分析和项目材料起草。这说明咨询行业对 AI 的需求并非停留在个人试用阶段，而是可以在统一知识平台上形成较高频率的组织级调用。

图7：麦肯锡内部平台 Lilli 使用数据

- quickly getting up to speed on unfamiliar topics
- accessing and learning McKinsey frameworks and methodologies
- cleaning and analyzing data
- researching industry trends
- finding internal subject-matter experts
- creating drafts of meeting agendas and project work plans
- developing presentations in McKinsey's style and format
- pushing team thinking toward more-creative solutions



来源：McKinsey，国金证券研究所



### 3.5 未来展望：垂直行业由试点验证走向任务量扩张

移动互联网时期的爆款通常依靠统一界面和标准化需求快速积累用户，而 Agent 的价值更多来自个人或企业独有的数据、权限和业务流程。相同的底座模型可以被配置为研究助理、合同审阅工具、客户服务 Agent、财务结账助手或个人生活助理。应用形态因此会更加分散，单一产品的用户规模和市场声量未必能够复制移动互联网时代的超级应用，但不同场景仍共同消耗模型调用、云计算、数据库、连接器和安全治理资源。

这种商业结构更接近“平台加长尾应用”。模型和云平台厂商通过底座能力、数据层及工作流入口获得规模收益，垂直软件厂商则依靠专业数据、行业模板、客户关系和交付责任形成差异化，并通过席位订阅、任务量、Token 消耗、功能增购或结果付费实现收入。

未来一年，AI 应用的主要增量可能来自现有行业软件中的任务量扩张，而非单一新入口的集中爆发。应用软件公司收入增速抬头也是证明。随着模型可靠性、Computer Use 和系统连接能力提高，教育、金融、法律和咨询等行业将由检索、摘要和内容生成进一步延伸至资料整理、跨系统核验、流程协同和受控执行。率先进入生产环境的仍将是高频、规则相对明确、结果能够复核且责任边界清晰的任务。单个应用未必形成全民级爆款，但大量垂直任务同时增长，仍可推动 Token 消耗、数据查询和软件收入持续扩大。



## 四、算力结构性机会：回归产业与系统升级

### 4.1 宏观扰动深化：回归产业并关注模型演进

能源、利率与地缘因素的博弈已经进入更复杂阶段，短期价格信号对市场风险偏好的影响明显增强。EIA 数据显示，布伦特油价在 2026 年 3 月快速突破 100 美元，并在 3 月底至 4 月初完成上一轮集中冲高，随后出现较大幅度回落。宏观变量可能迅速改变估值和交易节奏，但其持续时间与产业订单周期并不完全一致。

因此，算力研究需要区分宏观扰动、资本开支节奏和技术升级三类变量。前两者会影响项目开工及收入确认时点，真正决定中期产业价值量的仍是模型能力、推理负载、网络速率和机柜功率。市场波动加大时，更应跟踪客户订单、供应商认证、产能利用率 and 产品结构，而不是仅凭资本开支同比增速判断景气方向。

大模型本身也在快速变化。竞争重点正由参数规模和单项评测，转向推理效率、Computer Use、长任务执行、多模态输入以及不同模型之间的协同调度。算法效率提升可能降低单次调用成本，但 Agent 会引入更长上下文、更多工具调用、重试和校验步骤，单位任务所需计算量未必同步下降。应用边界扩张与成本下降共同作用，仍可能推高推理总量，并持续提高对互连、供电和散热系统的要求。

### 4.2 高速互连：PCB 与光通信受益系统复杂度提升

PCB 的结构性机会来自计算与交换系统复杂度提升，而不只是服务器数量增加。更高速率的信号传输要求更低损耗材料、更严格的阻抗控制、更高层数和更精细的布线，GPU 板、交换机板、背板及封装基板的单机价值量随之提高。即使整机出货增速波动，只要网络代际和板级设计继续升级，高阶 PCB 收入仍可由规格提升和份额集中获得支撑。

光通信则受益于集群规模、传输距离和端口速率共同提升。机柜内部的短距离连接仍会并存多种铜连接方案，但跨机柜和交换层对带宽、距离与能耗的要求上升，会推动更多链路向光连接迁移。光模块、光引擎、激光器和无源器件同时受到端口数量与单端口速率提升驱动，产业重点将由单一速率升级扩展到光连接比例、良率、功耗和每比特成本。

公司经营数据已经体现这一趋势。TTM Technologies 披露，其 2026 年第二季度数据中心与网络业务收入同比增长 91%，占总收入 40%；Coherent 2026 财年第四季度收入达 20.9 亿美元，同比增长 34%，并将 AI 数据中心由铜向光迁移视为多年机会。两类产品都具有较长认证周期和较高制造门槛，核心供应商有望在产品升级中获得份额与价值量提升，但仍需关注新产能爬坡、客户集中、产品降价和良率波动。

### 4.3 电源与散热：功率密度成为数据中心硬约束

电源与散热的需求核心是单机柜功率和部署复杂度提升。高密度 GPU 集群需要在有限空间内完成更高效率的电力转换、配电和热量转移，传统风冷与低压配电方案的空间和能效约束更加突出。UPS、母线、电源模块、高压直流、冷板液冷、CDU 及热管理控制系统的配置价值，取决于机柜功率、冗余要求和系统复杂度，对服务器台数或单季资本开支增速的依赖相对较低。

这一环节也从单项设备采购转向系统方案交付。电力接入、备用电源、能效管理、液冷介质和监控软件之间的耦合增强，客户更重视从设计、部署到运维的完整能力。Vertiv 披露，2026 年第二季度收入同比增长 24% 并上调全年主要指标指引，同时指出 AI 与通用计算需求增强、基础设施部署更加复杂和密集。具备系统设计、全球交付和售后服务能力的厂商，更容易提高单项目价值量并延长客户合作周期。

产业层面仍需关注并网资源、建设审批、供应链交期和液冷方案切换，但这些约束本身也会推动电力与热管理投入前置。研究上应重点跟踪在手订单、交付周期、机柜功率、液冷渗透率、大功率产品占比及毛利率变化。整体而言，PCB、光互连、电源与散热的共同逻辑是系统规格持续升级，其景气度虽会受到建设节奏扰动，但相较单纯依赖资本开支总量，更具结构性和可持续性。



## 五、相关标的

大模型&云：智谱、MiniMax、阿里巴巴、腾讯控股、金山云、百度集团、优刻得、首都在线、网宿科技、云赛智联、青云科技等。

海外算力：东山精密、工业富联、胜宏科技、江海股份、中际旭创、中钨高新、蓝思科技、东阳光、火炬电子、三环集团、欧科亿、天孚通信、鼎泰高科、新易盛、领益智造、兆易创新、鹏鼎控股、唯科科技、天岳先进、大普微、源杰科技、麦格米特、景旺电子、英维克、京东方等；英特尔、SK 海力士、Lumentum、闪迪、铠侠、美光。

国内算力：寒武纪、海光信息、长鑫科技、东阳光、利通电子、中芯国际、华虹半导体、华达科技、摩尔线程、天数智芯、沐曦股份、壁仞科技、协创数据、禾盛新材、华丰科技、杰华特、京基智农、中科曙光、浪潮信息、利扬芯片、胜蓝股份、华勤技术、国科微、中国长城、晶科科技、罗曼股份、盈峰环境、芯原股份、晶科科技、亿田智能、豫能控股、星环科技、鸿日达、盛视科技、神州数码、润泽科技、大位科技、润建股份、奥飞数据、瑞晟智能、科华数据、潍柴重机、欧陆通、杰创智能、奥尼电子。

AI 应用：达梦数据、深信服、卓易信息、星环科技、广立微、绿盟科技、网宿科技、海天瑞声、合合信息、德才股份、美年健康、真爱美家、中控技术、迈富时、金蝶国际、康众医疗、用友网络、聚水潭、阜博集团、范式智能、汇量科技、金山办公、奔图科技、中国软件。

## 六、风险提示

行业竞争加剧的风险：

云厂商、模型厂商及芯片企业降价或推出替代方案，可能压缩份额与利润率。

技术研发进度不及预期的风险：

模型能力、Agent 稳定性、芯片迭代及液冷机架交付延期。

特定行业下游资本开支周期性波动的风险：

互联网云厂商、运营商、数据中心及金融、政府客户延后或削减 AI 服务器与算力基础设施投资。



行业投资评级的说明：

买入：预期未来 3—6 个月内该行业上涨幅度超过大盘在 15%以上；

增持：预期未来 3—6 个月内该行业上涨幅度超过大盘在 5%—15%；

中性：预期未来 3—6 个月内该行业变动幅度相对大盘在 -5%—5%；

减持：预期未来 3—6 个月内该行业下跌幅度超过大盘在 5%以上。



## 特别声明：

国金证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

任何形式的复制、转发、转载、引用、修改、仿制、刊发，或以任何侵犯本公司版权的其他方式使用。经过书面授权的引用、刊发，需注明出处为“国金证券股份有限公司”，且不得对本报告进行任何有悖原意的删节和修改。

本报告的产生基于国金证券及其研究人员认为可信的公开资料或实地调研资料，但国金证券及其研究人员对这些信息的准确性和完整性不作任何保证。本报告反映撰写研究人员的不同设想、见解及分析方法，故本报告所载观点可能与其他类似研究报告的观点及市场实际情况不一致，国金证券不对使用本报告所包含的材料产生的任何直接或间接损失或与此有关的其他任何损失承担任何责任。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，在不作事先通知的情况下，可能会随时调整，亦可因使用不同假设和标准、采用不同观点和分析方法而与国金证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。

本报告仅为参考之用，在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告提及的任何证券或金融工具均可能含有重大的风险，可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可能会受汇率影响而波动。过往的业绩并不能代表未来的表现。

客户应当考虑到国金证券存在可能影响本报告客观性的利益冲突，而不应视本报告为作出投资决策的唯一因素。证券研究报告是用于服务具备专业知识的投资者和投资顾问的专业产品，使用时必须经专业人士进行解读。国金证券建议获取报告人员应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。报告本身、报告中的信息或所表达意见也不构成投资、法律、会计或税务的最终操作建议，国金证券不就报告中的内容对最终操作建议做出任何担保，在任何时候均不构成对任何人的个人推荐。

在法律允许的情况下，国金证券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多种金融服务。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。国金证券并不因收件人收到本报告而视其为国金证券的客户。本报告对于收件人而言属高度机密，只有符合条件的收件人才能使用。根据《证券期货投资者适当性管理办法》，本报告仅供国金证券股份有限公司客户中风险评级高于 C3 级（含 C3 级）的投资者使用；本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具，本报告的收件人须保持自身的独立判断。使用国金证券研究报告进行投资，遭受任何损失，国金证券不承担相关法律责任。

若国金证券以外的任何机构或个人发送本报告，则由该机构或个人为此发送行为承担全部责任。本报告不构成国金证券向发送本报告机构或个人的收件人提供投资建议，国金证券不为此承担任何责任。

此报告仅限于中国境内使用。国金证券版权所有，保留一切权利。

### 上海

电话：021-80234211

邮箱：researchsh@gjzq.com.cn

邮编：201204

地址：上海浦东新区芳甸路 1088 号

紫竹国际大厦 5 楼

### 北京

电话：010-85950438

邮箱：researchbj@gjzq.com.cn

邮编：100005

地址：北京市东城区建国门内大街 26 号

新闻大厦 8 层南侧

### 深圳

电话：0755-86695353

邮箱：researchsz@gjzq.com.cn

邮编：518000

地址：深圳市福田区金田路 2028 号皇岗商务中心

18 楼 1806



【小程序】  
国金证券研究服务



【公众号】  
国金证券研究