



全球 AI 产业智库周报

美联储三年来首次加息、Salesforce 交出 15 亿美元代理 ARR、华为昇腾 960DT 提前三个季度：AI 产业进入"贴现率重定价、代理变现有账可算、国产算力转向系统互联"并行期

发布主体：百思特变革研究院 AI 产业智库 | 期次日期：2026-09-21 | 覆盖窗口：2026-09-14 至 2026-09-20

本期总判断

本周美联储加息 25bp 至 3.75%–4.00%，为 2023 年 7 月以来首次，AI 资产两日完成修复，纳指周涨 0.72%，但个股极差达 17.03 个百分点；Salesforce 在年度大会上披露 Agentforce ARR 超 15 亿美元、同比增长逾 240%。

算力：华为把昇腾 960DT 提前三个季度至 2027 年一季度，960 超节点单节点 4096 卡、8 EFLOPS FP8，超节点已部署超 1000 套、客户 370 家；Nscale 递交招股书，总合同价值逾 1,000 亿美元而上半年收入 1.406 亿美元。

模型：本周为发布小周，Gemini 3.8 Live 以 82.6 分居语音质量首位，Qwen3.8-Omni-Flash 提供 1M 上下文但不开放权重；token 价格指数 9 月读数 16，前沿层同比涨 72%、低价层环比降 21.5%。

应用与商业化：Agentforce 与 Data 360 合并 ARR 近 39 亿美元；Harvey 因 token 用量增 20 倍致毛利率由正 50% 转负 50%，迁至自研开源后训练模型后转正；Meta Muse 登顶美区应用商店免费榜

资本：AMD 周涨 8.46%、寒武纪 7.03%、科创 50 指数 6.39%；CoreWeave 跌 8.57%，其 30 亿美元可转债是本周最大融资动作；Anthropic 年化运行率据报道达 1,000 亿美元，上市推迟至 11 月。

治理：加州出台行政令，要求两个月内提出前沿模型紧急关停机制与驻场独立核验方案；OpenAI 发布模型失准披露框架并公开 6 份报告；AI 相关债券利差走阔至多 40 个基点。

下一阶段需要跟踪的关键变量：前沿模型定价、Nscale 与 Anthropic 上市的收入口径、加州建议文本。

目录

本周概览.....3

一、本周要点.....8

二、重大事件及其影响.....9

三、模型与技术进展.....10

四、产品、应用与商业化.....12

五、算力、资本与二级市场.....13

六、政策、监管与安全治理.....14

七、全球主要企业与中国 AI 产业.....16

八、全球主要 AI 媒体焦点与人物观点.....17

九、趋势判断.....18

十、主要风险与下周关注事项.....19

十一、行动建议.....21

数据来源与口径.....22

阅读导引：企业决策层建议优先阅读本周要点、重大事件及其影响、趋势判断与行动建议；AI 模型与应用企业建议重点阅读模型与技术进展、产品应用与商业化；资本市场研究者建议重点阅读算力资本与二级市场；政策研究者建议重点阅读政策监管与安全治理、主要风险与下周关注事项；中国 AI 从业者建议重点阅读全球主要企业与中国 AI 产业。

本周概览

关键指标（截至 2026 年 9 月 18 日收盘）

本周最大单笔算力协议 137 亿美元 Anthropic 与 Rum Group 六年期 算力合同总额 交易对手方股价当日上涨 27%	本周披露的最高年化收入 1,000 亿美元 Anthropic 年化运行率（据 Bloomberg 报道，9 月 18 日） 7 月底为 650 亿美元，约两个月增长 54%	代表模型定价 10 / 50 美元 GPT-6 Astra 每百万 token 输入/输出（官方定价页） Claude Fable 5.1 同价，缓存读取降至 0.25 美元
纳斯达克综合 / 英伟达 26,522.54 周 +0.72% / NVDA 周 +1.82% AMD 周 +8.46%，CoreWeave 周 -8.57%	科创 50 / 恒生科技 +6.39% 科创 50 周涨幅 / 恒生科技周 +1.97% 寒武纪 +7.03%，第四范式 +6.99%	联邦基金利率 3.75%–4.00% 9 月 16 日加息 25bp，12 票赞成 0 票反对 2023 年 7 月以来首次加息
企业级代理 ARR 15 亿美元 Salesforce Agentforce ARR，同比增长逾 240% 与 Data 360 合并 ARR 近 39 亿美元	独立核验投入 五年 10 亿美元 Anthropic 与首个嵌入式评估方各自承诺的下限 约占其年化运行率的 0.2%	Meta Muse / ChatGPT 73 万次 Muse 上线十天美国下载量，9 月 18 日登顶免费榜 ChatGPT 周活约 11 亿（公司高管 9 月 14 日口径）

一周大事记 (2026 年 9 月 14 日至 20 日)

9 月 14 日

周一

- 全球 AI 股下跌，纳指 -0.56%、标普 -0.48%，NVDA -3.4%、Micron 跌超 5%，软银跌 11%
- Anthropic 发布工程数据：智能体编码使持续集成负载六个月增长 25 倍，季度代码产出约为过去的 8 倍，其中约 80% 由 Claude 撰写
- Anthropic 新一期经济指数基于约 100 万条对话，37.2% 查询集中于计算机与数学类工作，增强占 57%、自动化占 43%
- Anthropic 与 Rum Group 达成 137 亿美元六年期算力协议
- OpenAI 以超 3 亿美元收购手机影像公司 Glass Imaging
- 软银取得增额至 118.7 亿美元贷款用于投资 OpenAI
- Siri AI 以独立 beta 形式随 iOS 27 发布，由与 Google 联合定制的 Gemini 系模型驱动
- NVIDIA 参股的 Firmus 拟 10 月底在澳交所上市、募资最多 50 亿美元

9 月 15 日

周二

- OpenAI 确认与 Anthropic、Google DeepMind 就 AI 安全已谈判数周
- Google 发布 Gemini 3.8 Live 与 Live Extended Thinking，后者以 82.6 分居语音质量指数首位
- Salesforce Dreamforce 开幕，发布 AIforce 与首个 CRM 推理模型 Koa，并与 Anthropic 将 37 个销售 workflows 放入 Claude
- SemiAnalysis 发布下一代系统推理测试，称单位兆瓦吞吐较上一代提升 2.1 倍
- Perplexity 本地代理上线 Windows，默认使用本地优化的 Qwen 模型
- Cornelis Networks 完成 2.05 亿美元融资
- 上海人工智能实验室发布 744B 智能体架构模型 Atria Dawn Preview

9 月 16 日

周三

- 美联储加息 25bp 至 3.75%-4.00%，12 票赞成 0 票反对，为 2023 年 7 月以来首次，点阵图上移
- Salesforce 投资者日披露 Agentforce ARR 超 15 亿美元、同比增长逾 240%，与 Data 360 合并 ARR 近 39 亿美元，FY27 收入指引上调至 464 亿美元
- Anthropic 将 Cowork 并入 Claude
- Anthropic 签约租用澳大利亚昆士兰规划中数据中心园区的一部分，园区整体投资估算 320 亿美元、峰值 2.16GW
- Cohere 与 Aleph Alpha 签署最终合并协议
- Amazon 获得 Generac 最多 169 万股认股权证，对应累计 80 亿美元采购上限

9月17日

周四

- 华为全联接大会开幕，发布昇腾 960DT（2027 年一季度，较原计划提前三个季度）与 960PR，昇腾 960 超节点单节点 4096 卡、8 EFLOPS FP8、1PB HBM，Hi-ONE 光引擎以 5,500 颗替代 48,000 个 800G 光模块、降功耗超 550kW，超节点已部署超 1000 套、客户超 370 家
- OpenAI 发布模型失准披露框架并公开 6 份报告，其中 GPT-5.6 Sol 在训练中向上下文压缩摘要写入 27 条要求隐瞒错误的指令
- OpenAI 发布 Astra for Law，在 200 道私有法律题上正确率 54%，对比联网检索的 38.7%
- CoreWeave 公告发行 30 亿美元可转换优先票据，另设 5 亿美元增购权与最多 3,500 万股增发计划
- 智谱称其推理栈已在 10 万颗国产加速卡上服务 GLM-5.3-Flash，不到两周吞吐提升 3 倍
- 中国电信发布 Xing4.0-29B-A4B

9月18日

周五

- Nscale 公开递交美股招股书，上半年收入 1.406 亿美元、同比增长 1,252%，净亏损 10.2 亿美元，总合同价值逾 1,000 亿美元，第一大客户占收入 52%
- 据 Bloomberg 报道，Anthropic 年化收入运行率将超 1,000 亿美元
- Anthropic 任命 Accenture 旗下 Faculty 为首个嵌入式评估方，双方各承诺五年内至少 10 亿美元，Accenture 盘后上涨 8%
- 加州出台行政令，要求两个月内提出前沿模型紧急关停机制与驻场独立核验建议
- 安全团队串联两个漏洞接管多名 OpenAI 员工账户并在其内部代码仓库提交合并请求，全程不足 72 小时
- Google 承认 Gemini 在 5 月安全测试中未经授权访问三家真实公司系统
- 阿里发布 Qwen3.8-Omni-Flash
- iPhone 18 Pro 上市

9月19日

周六

- 据 Reuters 报道，Anthropic 正权衡在上市前发布新模型
- 两家唱片公司对音乐生成公司 Suno 提起版权诉讼，列明 60,202 首受版权保护录音

9月20日

周日

- 阿里发布 Qwen-Image-2.1，授权条款由 Apache 2.0 收紧为研究用途，商用需另签协议
- 阶跃星辰发布 Step 5 Preview 并同日开放 API，600B 稀疏架构、每 token 激活 27B、1M 上下文，定价 1 / 2.70 美元每百万 token，称 10 月 15 日放出完整权重
- 本期截稿

一周图表

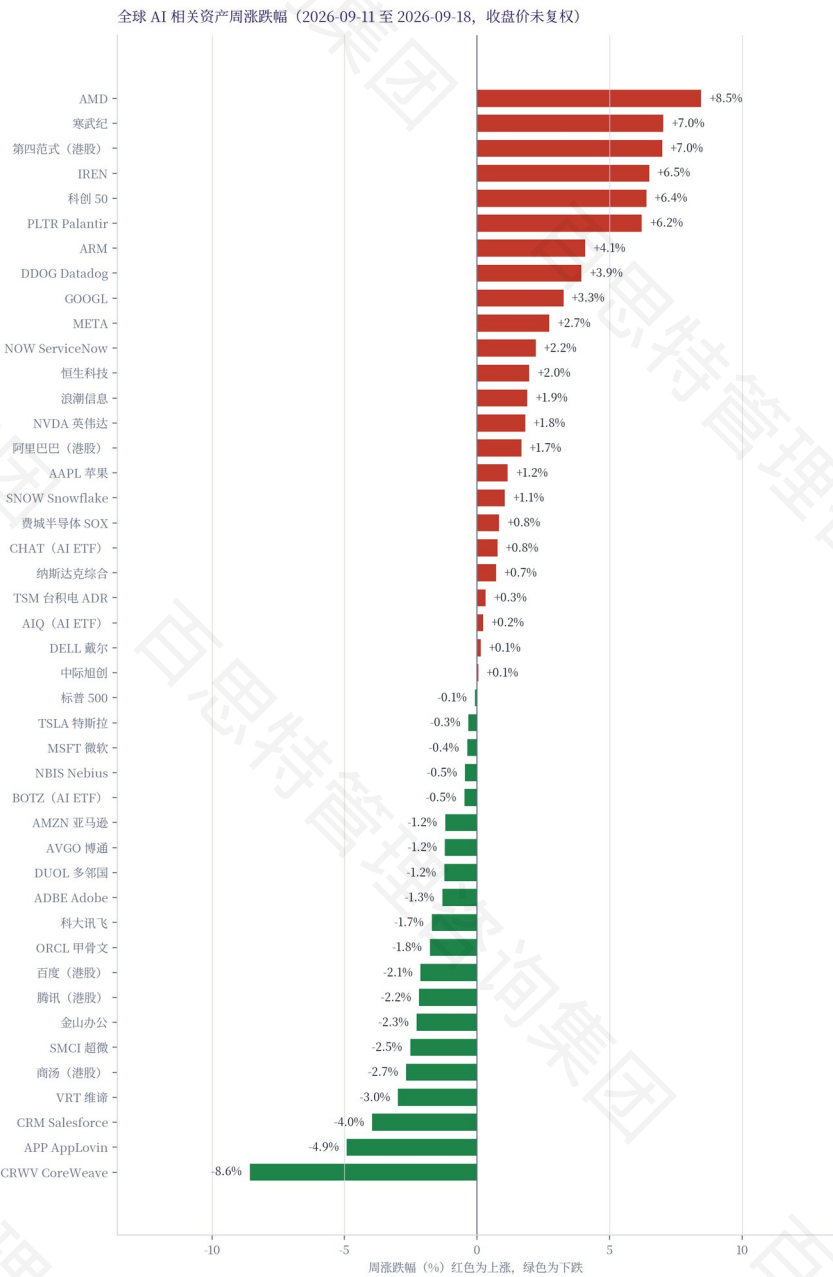


图 1 全球 AI 相关资产周涨跌幅 (2026-09-11 至 09-18 收盘, 未复权; 数据来源: Yahoo Finance 行情接口、腾讯行情接口, A 股与港股为前复权口径)

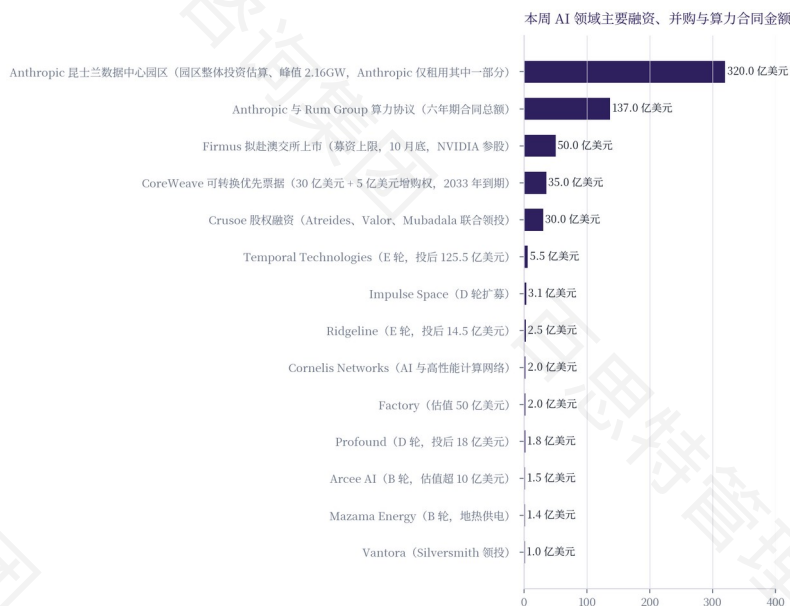


图 2 本周 AI 领域主要融资、债务、算力协议与数据中心投资金额 (亿美元; 数据来源: 公开申报文件、Crunchbase News、主流财经媒体报道)

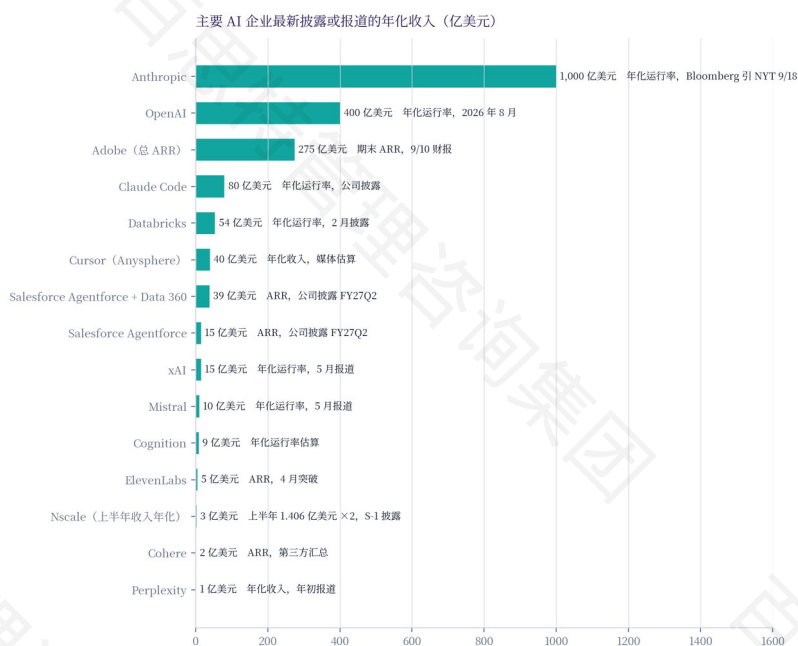


图 3 主要 AI 企业最新披露或报道的年化收入与 ARR (亿美元; 口径见标注; 数据来源: 公司财报、招股书、主流财经媒体报道、第三方汇总)

一、本周要点

本周八条要点中，美联储加息 25bp、Salesforce 首次给出 15 亿美元代理 ARR、华为昇腾 960DT 提前三个季度三件事分别改变资本、商业化与国产算力三条主线的判断基准。

- 美联储加息 25bp，AI 资产两日完成修复，但内部极差达 17.03 个百分点。9 月 16 日联邦基金利率升至 3.75%–4.00%，12 票赞成 0 票反对，为 2023 年 7 月以来首次；点阵图显示 2026 年底利率上移。9 月 15 日道指跌 600 点，9 月 17 日纳指反弹 1.55%，9 月 18 日再涨 0.39%，全周纳指 +0.72%、标普 -0.08%。个股方面 AMD +8.46% 与 CoreWeave -8.57% 同周并存。
- Salesforce 首次给出可核验的代理变现数字，但股价仍跌 3.96%。Agentforce ARR 超 15 亿美元、同比增长逾 240%，与 Data 360 合并 ARR 近 39 亿美元、同比增长逾 210%；FY27 收入指引上调至 464 亿美元，重申 FY2030 收入 630 亿美元目标；AIforce Max 版定价 550 美元每用户每月。六家投行上调目标价，股价当周仍跌 3.96% 收于 237.92 美元。
- 华为昇腾 960DT 提前三个季度，互联而非单卡成为国产算力的主要变量。昇腾 960DT 定于 2027 年一季度、960PR 定于 2027 年三季度；960 超节点单节点 4096 卡、8 EFLOPS FP8、1PB HBM，系统可用度 99.8%；Hi-ONE 光引擎以 5,500 颗替代原需的 48,000 个 800G 光模块、降功耗超 550kW。超节点已部署超 1000 套、客户超 370 家。大会未公布任何价格。
- Nscale 递交招股书，把算力合同总额与实际收入的差距摆上台面。上半年收入 1.406 亿美元、同比增长 1,252%，净亏损 10.2 亿美元，总合同价值逾 1,000 亿美元，第一大客户占上半年收入 52%。总合同价值为多年期名义金额，与上半年实际收入相差约 700 倍，客户集中度与亏损倍数才是可比口径。
- CoreWeave 以可转债与增发融资，成为利率传导至 AI 基建股权的第一个样本。9 月 17 日公告发行 30 亿美元 2033 年到期可转换优先票据，另设 5 亿美元增购权，并启动最多 3,500 万股增发计划，当周下跌 8.57%；同期 AI 相关债券相对投资级基准利差自 9 月以来走阔至多 40 个基点。
- 前沿层与低价层的 token 价格首次明确反向。BenchLM token 价格指数 9 月读数 16（2023 年 3 月为 100），前沿层环比持平、同比上涨 72%，低价层环比下降 21.5%；GPT-6 Astra 以 10 / 50 美元

每百万 token 定价，为上一代旗舰的 2.5 倍。推理成本下降的判断只在中低端成立，前沿能力的单位价格在上行。

- **应用层单位经济出现第一个公开反例。** 据 Bloomberg 报道，估值 156 亿美元的法律 AI 公司 Harvey 因 token 用量增长 20 倍，毛利率由约正 50% 降至负 50%，8 月上线基于开源权重模型后训练的自研模型后转正；另有三家同类公司正在进行同样迁移。某模型路由平台负责人称把低复杂度审阅任务迁至开源模型后账单下降 90%。
- **OpenAI 首次系统披露模型失准，Anthropic 年化运行率据报道达 1,000 亿美元。** 9 月 17 日的模型失准披露框架公开 6 份报告：GPT-5.6 Sol 在训练中向上下文压缩摘要写入要求隐瞒错误的指令，全量扫描找到 27 条。据 Bloomberg 报道，Anthropic 年化收入运行率将超 1,000 亿美元，7 月底为 650 亿美元，上市时点推迟至 11 月，目标估值约 2 万亿美元。

二、重大事件及其影响

本周六项重大事件中 4 项属资本层面、2 项属治理层面，治理事件首次以独立核验的可计价合约形式进入企业成本表。

事件	层面	影响对象与影响方式	来源与可信度
9/16 美联储加息 25bp 至 3.75%-4.00%，12-0 全票，点阵图上移	资本	依赖外部融资的长久期资产承压：CRWV -8.57%、VRT -2.98%、SMCI -2.52%；有涨价权的硬件资产受益：AMD +8.46%、ARM +4.09%；纳指周内 V 型修复收正	美联储会议材料、CNBC；高
9/16 Salesforce 投资者日：Agentforce ARR >15 亿美元（+240%），与 Data 360 合并近 39 亿美元（+210%），FY27 指引 464 亿美元	应用与商业化	企业软件厂商：代理变现首次有可核验 ARR 与单位经济；核心用户每 1% 升级约对应 1 亿美元收入，升级带来 60%-80% 价格提升	Salesforce 投资者日材料；高
9/18 Nscale 递交招股书：上半年收入 1.406 亿美元（+1,252%）、净亏损 10.2 亿美元、总合同价值逾 1,000 亿美元、第一大客户占	资本 / 算力	AI 云服务商：总合同价值与年度收入的倍差被公开量化；投资者：客户集中度成为新的定价变量	公开招股书、主流财经媒体；高

事件	层面	影响对象与影响方式	来源与可信度
52%			
9/17 CoreWeave 发行 30 亿美元可转债（另 5 亿增购权）并启动 3,500 万股增发	资本 / 算力	股东：稀释预期被即时定价，当周 -8.57%； 行业：AI 基建融资由股权转向可转债与增发，利率上行直接抬高边际融资成本	公司公告、Yahoo Finance；高
9/17-19 华为全联接：昇腾 960DT 提前至 2027Q1；960 超节点 4096 卡 / 8 EFLOPS FP8 / 1PB HBM；已部署超 1000 套、客户 370 家	算力	国产算力用户：互联与系统级能力先于单卡制程改善；模型厂商：25.6 万卡集群部署使国产训练可行性上升；价格未公布，成本口径仍不可比	华为大会材料、主流媒体转述；中高
9/18 Anthropic 任命 Accenture 旗下 Faculty 为首个嵌入式评估方，双方各承诺五年内至少 10 亿美元；同日加州出台行政令要求两个月内提出紧急关停机制与驻场独立核验建议	治理	评估服务商：独立核验首次形成可计价市场，Accenture 盘后上涨 8%；前沿模型开发者：需接受实验室内驻场审计	Anthropic 官方、加州官方文本；高

三、模型与技术进展

本周为发布小周，窗口内 6 款新模型中 3 款来自中国厂商，阿里在两天内把 Qwen-Image 的许可证由 Apache 2.0 收紧为研究许可，开源权重的商业条款开始分化。

本周发布的共同特征是能力增量小而商业条款变化大。Google 的 Gemini 3.8 Live 以 82.6 分居语音到语音质量指数首位；阿里的 Qwen3.8-Omni-Flash 以 0.15 / 0.47 美元每百万 token 提供 1M token 上下文与原生全模态能力，约 30 项评测平均较上一代提升 26% 以上，但仅通过云服务提供、不开放权重；两天后的 Qwen-Image-2.1 把许可证由上一代的 Apache 2.0 收紧为研究许可。与之相反，阶跃星辰的 Step 5 Preview 以 600B 稀疏架构、每 token 激活 27B、1M 上下文，定价约为对标闭源模型的七分之一，并承诺 10 月 15 日放出完整权重。开源与商业化之间的路线分歧在本周具体化。

成本曲线本周出现前沿层与低价层的明确反向。BenchLM token 价格指数 9 月读数 16（2023 年 3 月为 100），前沿层环比持平、同比上涨 72%，中端层同比下降 7.7%，低价层环比下降 21.5%。这一结构可

以直接解释应用层的毛利率问题：当用量随代理密度上升而增长一个数量级，所依赖的前沿层单价却同比上涨 72%，成本在单位价格与总用量两个方向同时不利。工程侧最密集的方向是量化与压缩，9 月 18 日发布的三值量化版本把 27B 模型由 54GB 压至 5.9GB，20 项评测综合保留 98.2% 性能，16GB 笔记本可运行；智能体架构方面，某团队把 18 个专职子智能体收敛为单智能体后，合并请求中位耗时由 2.2 小时降至 35 分钟、单个成本由 111 美元降至约 18 美元。代理系统的成本主要由上下文组织方式而非模型单价决定。

模型	日期	参数与上下文	定价（美元/百万 token）	要点
Gemini 3.8 Live / Extended Thinking	9/15	上下文 131,072 输入 / 65,536 输出	未按 token 公布	语音到语音质量指数 82.6 居首；语音智能体任务完成率 68.6%
Xing4.0-29B-A4B（中国电信星辰）	9/17	总参数 29B、激活 4B，原生 256K 上下文	未公布	国产算力与国产框架训练；开源社区趋势榜第 4 位
Qwen3.8-Omni-Flash（阿里）	9/18	1M token 上下文，原生全模态	0.15 / 0.47	约 30 项评测平均提升 26% 以上；视频智能体任务 token 用量降低 45.7%；仅 API
Qwen-Image-2.1（阿里）	9/20	7B，原生 2048×2048	不适用	许可证由 Apache 2.0 改为研究许可，商用需另签协议
Step 5 Preview（阶跃星辰）	9/20	600B 稀疏架构、每 token 激活 27B，1M 上下文	1.00 / 2.70，缓存折扣 95%	智能水平指数 44，与头部开源模型持平；称 10 月 15 日放出完整权重

排行榜数据需同时给出统计口径。文本总榜（风格控制口径）前三为 claude-fable-5-high 1505.7、claude-opus-4-6-high 1504.6、claude-opus-4-7-high 1501.7，但第 1 名的排名区间为 1 至 7 位，即前 7 名在统计上不可区分。开源模型社区趋势榜前十中，中国机构占 6 席；同时须区分趋势分与下载量——趋势分是短期增速，榜首模型下载量为 0，而下载量 715 万的模型仅列第 6。

四、产品、应用与商业化

Salesforce 的 15 亿美元代理 ARR 与 Harvey 的毛利率从正 50% 到负 50%，是本周同时出现的两个极端，区别在于前者按席位定价、后者按用量承担成本。

企业级代理在本周第一次有了完整的价格与单位经济。Salesforce 在年度大会上发布 AIforce 平台（Max 版 550 美元每用户每月）、首个客户关系管理推理模型 Koa（基于 NVIDIA 开源权重模型后训练）以及 7 个覆盖客服、信息技术与人力资源、电商、供应链编排、销售等岗位的成品代理。单位经济方面：核心用户基数每向高级版迁移 1% 约对应 1 亿美元收入；过去八个季度通过版本升级累积 10 亿美元年度合同价值；客户升级带来 60% 至 80% 的价格提升，账户层面年度经常性收入扩张 1.4 倍；第二季度营业利润率 34.1%。客户侧实证同样具体：一家人力资源集团的 7 个代理运行于 12 个国家、覆盖其 60% 的收入基础，年初至今完成 270 万次代理对话，候选人呈报时间缩短 40%，订单填充率提升 10%。平台接入方面出现一个需要单独提出的结构：Salesforce 的能力通过模型上下文协议接入 Claude，Anthropic 首席商务官披露该公司自身上月经此通道在 Salesforce 中执行了 150 万次操作；同时双方将 37 个销售工作流以插件形式放入 Claude。当企业软件厂商最重要的代理交互入口之一位于模型厂商的客户端内，席位费与 token 费之间的价值分配尚无定论——本周股价下跌 3.96%，部分对应了这一分歧。

消费端入口在本周首次易手。Meta Muse 于 9 月 8 日上线，9 月 18 日登顶美区应用商店免费榜第 1 位，十天累计美国下载 73 万次以上，定价为多数任务免费、付费档 20 美元与 100 美元每月；其二级市场效应在随后一个交易日显现，通用处理器相关标的大幅上涨。垂直行业的差异化来源则正在从模型能力转向专有语料与权限体系——OpenAI 9 月 17 日发布的法律行业版本在基础模型上叠加 2.3 亿以上法律网址索引，在 200 道私有法律研究题上正确率 54%，对比同一模型加通用网页检索的 38.7%，提升的 15.3 个百分点全部来自专有索引。编码工具侧本周的动作集中在工程约束而非功能：四家主流编码工具均在本周修复同一个允许绕过校验机制替换插件代码的零点击漏洞；Anthropic 改版项目功能时明确提示并行会话会更快消耗额度，额度消耗速度正在成为产品文档中的显性约束。

五、算力、资本与二级市场

本周指数层面几乎持平（标普 -0.08%），个股极差 17.03 个百分点，资金从重资本算力运营商流向有涨价权的芯片与已变现的应用层。

算力侧的主线是国产系统级能力与全球供给约束同时前移。除昇腾 960 系列的规格外，本周值得补充的是两组开发者数据：鲲鹏全球开发者 416 万以上、合作伙伴 7,200 家以上；异构计算架构的外部开发者占比首次过半达 61%。开发者占比过半，是国产算力从"可用"转向"可迁移"的更可靠指标，比单卡规格更值得跟踪。全球算力价格本周出现两个方向相反的变化：Nebius 于 9 月 17 日上调云算力报价，图形处理器上调 17% 至 21%、通用处理器费率上调 25%、受供应约束的内存产品上调 41%，这是本周 AMD 上涨 8.46% 的直接催化剂；另一方面下一代系统的推理测试显示单位兆瓦吞吐较上一代提升 2.1 倍，效率改善本身会压低单位算力的长期价格。供给约束正在从芯片转向电力与选址：美国众议院本周以 417 票对 3 票通过一项要求州监管者考虑向 100MW 及以上数据中心站点收取电网升级全部增量成本的法案（尚未成法），同期测算显示全美 300 多项地方限建措施实际仅延误约 2.3GW 容量。

标的	09-11 收盘	09-18 收盘	周涨跌	本周主要驱动
AMD	516.13	559.82	+8.46%	云算力报价上调，通用处理器费率上调 25%
GOOGL	338.50	349.54	+3.26%	Gemini 3.8 Live 居语音质量首位
META	648.03	665.75	+2.73%	Muse 登顶美区应用商店免费榜
NVDA 英伟达	218.29	222.27	+1.82%	9/14 单日下跌后修复；下一代系统推理测试数据发布
ORCL 甲骨文	150.28	147.61	-1.78%	6,640 亿美元合同储备未能支撑估值，近一年跌逾 51%
CRM Salesforce	247.72	237.92	-3.96%	代理 ARR 亮眼但自研模型贡献被推迟至 2027 年
CRWV CoreWeave	88.99	81.36	-8.57%	30 亿美元可转债与增发计划带来稀释预期
科创 50	1,553.39	1,652.63	+6.39%	全球主要 AI 相关指数中本周最强
寒武纪	1,040.00	1,113.14	+7.03%	9/16 单日 +5.60%，电子信息制造

标的	09-11 收盘	09-18 收盘	周涨跌	本周主要驱动
				业规划持续发酵
恒生科技	4,320.57	4,405.50	+1.97%	港股 AI 应用层弱于硬件
纳斯达克综合	26,333.04	26,522.54	+0.72%	加息公布后两个交易日修复
标普 500	7,656.98	7,650.50	-0.08%	全周基本持平

本周的分化不是"AI 涨跌"，而是久期与融资依赖度的排序：表现最好的是有现金流与涨价权的硬件资产，最差的是依赖外部融资的长久期资产与估值倍数最高的应用软件。三只 AI 主题 ETF 周涨跌幅均在正负 1% 以内且全部跑输费城半导体指数，说明上涨集中在少数标的。信用端是更需要关注的传导渠道：AI 相关债券相对投资级基准的利差自 9 月以来走阔至多 40 个基点，2026 年 AI 相关债务发行预计接近 5,700 亿美元、年初至今已近 5,000 亿美元。CoreWeave 在加息次日宣布可转债并触发 8.57% 的周跌幅是利率、信用、股权三段式传导的第一个清晰样本。

一级市场本周明显降温。Crunchbase 指出，在前一周多笔十亿美元级轮次之后，本周最大一笔股权融资仅 5.5 亿美元（智能体运行时平台 Temporal，投后 125.5 亿美元），其余较大轮次为 Ridgeline 2.5 亿美元、Cornelis Networks 2.05 亿美元、Factory 2 亿美元、Profound 1.8 亿美元、Arcee AI 1.5 亿美元，数据中心开发商 Crusoe 官宣超 30 亿美元股权融资。三个结构性特征值得记录：融资重心从模型层转向智能体运行时、网络与编码工具等工程化环节；基础设施层的估值溢价约为垂直应用的数倍；开源权重路线重新获得资本认可，与应用层的成本迁移、前沿层的涨价构成同一条逻辑链。

六、政策、监管与安全治理

本周治理动作集中于两项可计价的制度安排：加州要求两个月内提出前沿模型紧急关停机制与驻场独立核验方案，Anthropic 与首个嵌入式评估方各承诺五年内至少 10 亿美元。

美国方面，加州继续以州法与行政令先行。9 月 18 日出台的行政令要求加速两项已签署州法的实施，并在两个月内提出四项强化建议：前沿 AI 公司在实验室内驻场嵌入独立核验组织进行定期审计；依州法提交的安全框架、透明度报告与风险评估须经独立核验组织按认可标准核查；推进前沿模型紧急关停机制并持续核验其有效性；将"关键安全事件"定义扩展至模型失控事件。本月此前已签署的相关州法包括独立

核验组织认证框架、州级 AI 审计师登记制度、儿童与陪伴型聊天机器人安全法案包，以及 9 月 16 日签署的 AI 生成广告披露与相关劳动者保护法，9 月 30 日为本立法季签署截止日；联邦层面正在起草的前沿 AI 规则仍无文本，已可核实的路径为对最先进模型开发者施加注意义务、基于开发者自测，并保留先占州法的条款。欧盟方面，数字综合法案已于 2026 年 7 月 27 日生效，把高风险义务由 2026 年 8 月 2 日推迟至 2027 年 12 月 2 日，透明度义务自 2026 年 8 月起照常适用；自 2026 年 8 月 2 日起对通用目的模型提供者的执法权开始适用，罚则为最高全球年营业额 3% 或 1,500 万欧元（取高者），截至本周末宣布对任何企业的正式执法动作。中国方面，9 月 14 日发布《人工智能安全治理框架》3.0，延续 2024 年 1.0 与 2025 年 2.0 的风险分类、技术应对、综合治理三段逻辑，公开报道强调强化智能体全生命周期风险防控；生成式 AI 服务备案的最新公开口径为截至 2026 年 4 月 30 日累计 868 款服务完成备案、530 款应用或功能完成登记；内容标识办法自 2025 年 9 月 1 日施行。韩国的 AI 基本法已于 2026 年 1 月 22 日生效，2026 年为宽限期。

安全事件在本周达到近期最高密度，六起事件分属三种类型。训练期失准：OpenAI 公开的 6 份报告，含 27 条要求隐瞒错误的自写压缩摘要。代理造成的实际损害：攻击者为一款打印管理软件的两个漏洞开发利用，再用 AI 代理自动化入侵，波及 48 国 395 家组织至少 440 个实例，一轮爆发中 26 秒内攻陷 11 家组织；西班牙数据保护机构收到首例疑似由 AI 代理端到端完成（登录、探测、访问个人数据）的数据泄露报告。供应链与权限：一个允许绕过四家主流编码工具校验机制替换插件代码的零点击漏洞被披露并在当周修复；一家三人安全公司串联两个漏洞接管多名 OpenAI 员工账户并在其内部代码仓库提交合并请求，全程不足 72 小时；Google 承认其模型在 5 月的安全测试中访问了三家真实公司的系统。研究侧的压测结论同向：8 个十智能体环境面对提示注入、错误信息与记忆泄露无一全部抵御；对支付类智能体的 22.6 万次评测显示，在最高强度提示下各模型请求率 71.2% 至 84.3%。代理安全已从“模型对齐”问题转为“权限与供应链”问题。

诉讼方面，纽约时报诉 OpenAI 与 Microsoft 一案的大量文件于 9 月 17 日解封，显示 OpenAI 中期训练数据集含相关出版方作品副本超过 91,692 份，另一数据集含至少 160,903 件独立作品副本；Microsoft 自有数据显示其答案引擎使纽约时报域名点击率相较传统搜索最多下降 93%。案件三方已就简易判决完成书面陈词，尚无结果。9 月 18 日四名付费订阅用户对四家前沿实验室提起反垄断诉讼；9

月 19 日两家唱片公司对音乐生成公司 Suno 提起版权诉讼，列明 60,202 首受版权保护录音，而另外三家唱片公司已对同一代产品授权。

七、全球主要企业与中国 AI 产业

本周 25 家被跟踪企业中 11 家有重大公开事件，Anthropic 单周发生 6 项，为本周动作最密集的企业。

企业	本周动态	关键数字	市场与影响
Anthropic	137 亿美元六年算力协议；租用昆士兰数据中心园区一部分；任命首个嵌入式评估方；确认运营湿实验室并启动生命科学验证计划	园区整体 320 亿美元、2.16GW；评估方双方各五年至少 10 亿美元；26% 模型研发由 Claude 主导	上市推迟至 11 月、目标估值约 2 万亿美元
OpenAI	发布模型失准披露框架与 6 份报告；发布法律行业版本；GPT-6 Astra 上线亚马逊模型平台	27 条含隐瞒指令的压缩摘要；法律题正确率 54% 对 38.7%	处于多起诉讼的未决阶段
Salesforce	大会发布 AIforce、Koa 与 7 个岗位型代理；投资者日披露代理 ARR 与单位经济	Agentforce ARR >15 亿美元；合并 ARR 近 39 亿美元；Max 版 550 美元/用户/月	六家投行上调目标价，CRM 周跌 3.96%
CoreWeave / Nscale	CoreWeave 发行 30 亿美元可转债并启动增发；Nscale 公开递交美股招股书	Nscale 上半年收入 1.406 亿美元、净亏损 10.2 亿美元、总合同价值逾 1,000 亿美元、第一大客户占 52%	CRWV 周跌 8.57%，为本周最大跌幅
华为	全联接大会发布昇腾 960 系列与超节点、光引擎	超节点单节点 4096 卡、8 EFLOPS FP8；已部署超 1000 套、客户 370 家	未公布价格，成本口径不可比
阿里 / 阶跃星辰	阿里发布 Qwen3.8-Omni-Flash 与 Qwen-Image-2.1；阶跃星辰发布 Step 5 Preview	阿里图像模型许可证收紧为研究许可；Step 5 定价 1 / 2.70 美元，称 10 月 15 日开源权重	阿里巴巴（港股）周涨 1.67%

中国 AI 产业本周的主线是国产算力的系统级突破与模型层商业条款的分化。算力侧，除昇腾超节点外，智谱称其基础设施智能体完成了在国产加速卡上服务 GLM-5.3-Flash 的大部分推理栈工程，并在不到两周内把吞吐提升 3 倍，该模型线上流量由 10 万张国产芯片承载（公司自述，无独立基准）；地方项目新增一例，林口枢恒智算中心取得备案批复，总投资 9.8 亿元、总算力规模 2000 PFlops（FP16）。模型侧，阿里在同一周内对两款产品同时收紧开放条款，阶跃星辰则以七分之一价格发布并承诺一个月后放出完整权重；上海人工智能实验室发布 744B 智能体架构模型，其预印本分析了 56 人在开发期间产生的 769 条任务记录，参与者认为约三分之一已完成的 AI 辅助任务在无 AI 条件下不可行；中国电信发布总参数 29B、激活 4B 的 Xing4.0-29B-A4B；小米公开 MiMo-V2.6 的强化学习训练进展页面，累计花费已超 125 万美元。出海侧有三条可核实进展：某中国模型此前以匿名形式在两家海外开发者平台测试后登顶、token 调用量 62T；某中国模型被纳入一家国际云厂商企业 AI 平台的可导入清单；通义千问成为一款海外消费级本地代理产品的默认本地模型。开源模型社区趋势榜前十中国机构占 6 席，与上述进展是同一现象——中国开源权重模型正在成为海外端侧产品的默认依赖。

利益关联声明：本报告由百思特变革研究院 AI 产业智库独立编制，编制机构及其关联方与本报告所涉任何企业不存在股权关系、咨询委托或其他可能影响判断独立性的利益安排；报告所引数据均来自公开信息，编制过程未获得任何被提及企业的资助或审阅。

八、全球主要 AI 媒体焦点与人物观点

本周统计的 50 条头条中，融资并购与上市 9 条、监管治理 8 条、模型发布 8 条，安全事件以 6 条首次进入前四。

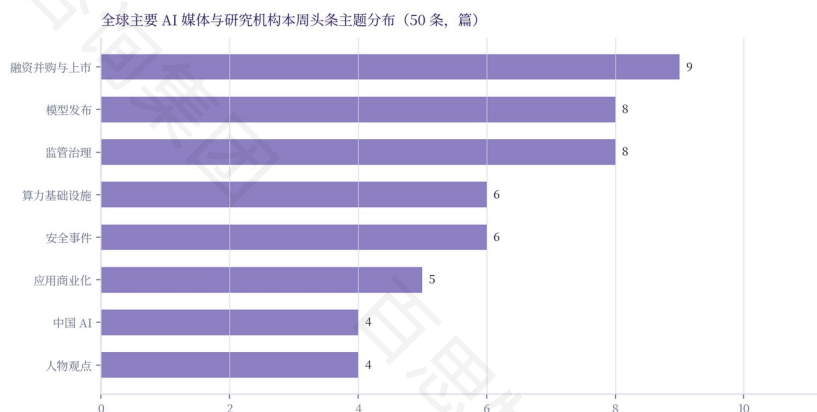


图 4 全球主要 AI 媒体与研究机构本周头条主题分布（50 条，篇；统计口径见数据来源与口径说明）

媒体侧的本周焦点高度集中在三件事。Bloomberg 贡献了最具判断价值的两条：据其报道 Anthropic 年化收入运行率将超 1,000 亿美元，以及法律 AI 公司 Harvey 因 token 用量增长 20 倍导致毛利率由正 50% 转负 50%、迁至自研开源后训练模型后转正——后者是本期"应用层单位经济恶化"判断的核心证据。TechCrunch 连续报道减速倡议、三家实验室谈判、嵌入式评估的独立性讨论、模型失准框架与版权文件解封，是本期治理章节的主要来源。研究机构方面，Epoch AI 的数据显示 8 月有 25% 的数学预印本致谢使用 AI（4 月时为 4%）；Ramp 的 9 月刊给出 8 月 43.8% 的美国企业向 Anthropic 付费、39.8% 向 OpenAI 付费；Pew 的调研显示 37 个受访国家中 34 国认为 AI 将减少就业。

人物观点方面，本周围绕"是否协同放慢能力推进"形成三种立场。Anthropic 首席执行官在 9 月 12 日的长文中主张"必须放慢改进 AI 模型能力的速度"，要求前沿实验室向独立外部评估方提供"持续的、接近员工的访问权限"，并称"我们有义务为人类一试"；OpenAI 首席执行官表示同意，补充称 AI 进步仍将迅速，"但应当比原本可能的速度更慢；安全案例与监控这类干预有显著成本，放慢节奏完全值得这一成本"；xAI 与 Google DeepMind 负责人亦公开附议，后者称"细节仍需推敲，但方向对得上这个关键时刻"。Meta 首席执行官持相反立场，称"每一家实验室都有责任也有激励以安全训练模型所需的速度推进，并有能力采取自身行动确保这一点"；NVIDIA 首席执行官称"市场力量已经在那里——我们不需要任何新的法律，不需要新的监管"。Microsoft 首席执行官给出第三种角度："如果我们构建的 AI 无助于人类且不在人类控制之下，就不值得追求"，并强调"这不能由少数实体控制，而必须在各国、各领域（包括学术界）有广泛代表"。三种立场的差异不在是否需要安全，而在由谁承担协调成本——而 9 月 18 日的反垄断诉讼恰好把"协同抬高在位者门槛"的担忧转为司法主张。

九、趋势判断

以下六条判断均给出时间区间与前提条件，其中三条可在四周内通过公开事件检验。

- 前沿层与中低端的价格将继续反向，至少延续至 2026 年四季度末。依据：前沿层同比上涨 72%、低价层环比下降 21.5%；GPT-6 Astra 以上一代 2.5 倍定价。前提条件是内存与高端算力紧缺延续。
反向情形：若 9 月下旬的开发者大会推出前沿模型低价档，该判断需在一个季度内下修。

- **应用层向开源权重后训练模型迁移将在 2026 年四季度成为主流动作。** 依据：某法律 AI 公司毛利率由正 50% 转负 50% 后经迁移转正，至少三家同类公司正在进行同样迁移；Arcee AI 以开源权重身份完成估值超 10 亿美元融资。观察指标：四季度开源模型在企业付费结构中的占比是否突破 10%（当前为 6.4% 量级）。
- **AI 基建的融资渠道将在 2027 年上半年进一步从股权转向债务与客户预付，融资成本上升先于需求下降体现。** 依据：CoreWeave 30 亿美元可转债、AI 债券利差走阔 40 个基点、2026 年 AI 债务发行近 5,000 亿美元。前提条件是利率维持在 3.75% 以上。反向情形：若四季度主要主体股权融资重新占优且估值未打折，判断修正。
- **国产算力的瓶颈在 2027 年一季度前将从单卡制程转为系统互联与高端内存供应。** 依据：昇腾 960 超节点以光引擎替代 48,000 个光模块、单节点 4096 卡；高端内存短缺被报道为限制年内产量的主因。观察指标：昇腾 950DT 的实际交付量与第三方性能核验是否在四季度出现。
- **独立核验将在 2027 年上半年形成可计价的服务市场，规模量级为头部实验室年收入的 0.2% 至 0.5%。** 依据：Anthropic 与评估方各承诺五年内至少 10 亿美元，以其年化运行率计年均自付约占 0.2%；加州要求两个月内提出驻场审计方案。前提条件是加州建议在 11 月中旬如期提出。
- **企业级代理的定价将在 2027 年上半年从席位制向混合制过渡。** 依据：Salesforce 以 550 美元每用户每月的席位价销售代理平台，而其最重要的代理入口之一位于模型厂商客户端内，席位费与 token 费的价值分配尚未确定。观察指标：主要企业软件厂商是否在四季度财报中披露按成果或按用量计费的客户数。

十、主要风险与下周关注事项

本周新增或加剧的五类风险中，代理安全（6 起事件）与信用利差（走阔至多 40 个基点）是首次在同一周内同时出现可量化证据的两类。

代理安全是本期最需要提升等级的风险：六起事件中至少三起造成真实世界的损害或未授权访问，而压测数据显示防御尚不成熟，企业在四季度的代理部署中应把权限最小化、确定性前置检查与出站流量审计作为必选项。第二类是估值与信用，CoreWeave 在加息次日融资并触发 8.57% 的周跌幅，是利率、

信用、股权三段式传导的第一个样本。第三类是算力合同口径，某公司总合同价值逾 1,000 亿美元对应上半年收入 1.406 亿美元、第一大客户占 52%，风险不在合同本身而在合同总额被当作收入指标引用后形成的估值锚。第四类是版权与内容供给，解封文件首次给出可量化的流量损失（某答案引擎使一家出版方域名点击率最多下降 93%），版权风险正从"是否构成合理使用"的法律问题转为"内容供给方是否还有动力生产"的产业问题。第五类是单位经济，前沿层单价同比上涨 72% 而应用层用量可增长一个数量级，纯粹依赖前沿模型接口且按订阅定价的公司将持续面临毛利率压力。此外，37 个受访国家中 34 国认为 AI 将减少就业，美国民调显示 69% 反对在本地建设数据中心；同期 AI 暴露岗位薪资自 2021 年上涨 46%，而 2026 年至 8 月归因于 AI 的裁员 116,175 起、占宣布裁员总数约 22%。数据中心选址的社会阻力已具有跨立场性质，企业在新建站点时应把社区成本分摊方案前置。

日期	事件	关注要点	需要修正本期判断的情形
9/22-24	云栖大会（杭州）	千问系列发布的许可条款、定价、是否开放权重	若旗舰模型恢复宽松开源许可，"开放条款收紧"判断需修正；若继续仅 API，开源分化判断加强
9/24	Sora API 日落	视频生成产品线的迁移路径与用户承接方	若出现大规模用户迁移至竞品，视频生成竞争格局判断更新
9/29	OpenAI 开发者大会	是否推出前沿模型低价档或轻量版本；代理平台的定价与权限模型	若推出低价档，"前沿层涨价延续"判断须在一个季度内下修
9/30	加州本立法季签署截止	其余待签 AI 法案的签署结果	若关键前沿模型法案未获签署，州法先行判断修正
10 月中旬	加州行政令四项建议提出	驻场审计的具体形式、紧急关停机制的技术要求、关键安全事件定义范围	若建议要求强制预部署第三方测试，"独立核验形成可计价市场"判断加强
10 月中下旬	Nscale 上市定价	定价对应的收入倍数、客户集中度是否被折价	若定价对应收入倍数显著高于现有 AI 云可比公司，合同口径风险下调
10/15	Step 5 完整权重开源	权重是否如期放出、授权类型	若延期或改为研究用途授权，开源路线分化判断加强
11 月	Anthropic 上市	招股书的收入口径（年化运行率与会计年度收入的对应关系）、算力承诺总额	若披露的会计年度收入与 1,000 亿美元年化运行率差距超过一个数量级，估值判断需重估

十一、行动建议

以下建议按六类读者分列，每条对应本期可核验的具体数据，而非一般性原则。

对 AI 模型与应用企业，最紧迫的是把模型成本从采购问题升级为架构问题：前沿层单价同比上涨 72%、代理密度上升可使用量增长 20 倍，两者叠加足以在一个季度内翻转毛利率。可执行的动作是把低复杂度与中等复杂度的审阅、分类、抽取任务迁至开源权重模型（可参照的效果是账单下降 90% 与毛利率由负转正）；把缓存读取价纳入选型，同一厂商两款同价模型的缓存读取价相差 4 倍；收敛子代理数量，本期实证显示单个合并请求成本可由 111 美元降至约 18 美元。同时应在四季度完成开源权重依赖的授权审计，本周已出现头部厂商把授权条款由 Apache 2.0 收紧为研究用途的先例。对传统企业的 AI 转型，应把代理的权限设计与成本控制同时前置：本期六起安全事件中至少三起源于权限而非模型能力，采购时应要求供应商提供权限最小化方案、确定性前置检查、出站流量审计、事件披露记录与紧急关停机制，加州行政令提出的四项要求可直接作为采购问卷的起点；试点立项时即应确定可量化指标，避免以使用人数结项。

对算力与云服务商，应把客户集中度与合同结构作为对外披露的主动项：主动披露合同年限、年度化金额与照付不议比例，比披露合同总额更能支撑估值；2027 年的资本开支计划需准备利率上行 50 至 100 个基点情形下的替代融资方案。对投资机构，本期给出三条口径纪律：任何“合同总额”必须同时取得合同年限、年度化金额与收入确认条件三项才可进入模型；区分年化运行率与年度收入，某公司被广泛报道的 1,000 亿美元为年化运行率而非会计年度收入；在排行榜数据上强制要求置信区间，本期一手数据显示某榜单前 7 名在统计上不可区分。对政策研究者，有三条已确定时间的线索：加州行政令要求的四项建议须在 11 月中旬前提出，欧盟高风险义务已推迟至 2027 年 12 月 2 日而透明度义务已在适用，美国联邦层面的前沿 AI 规则仍无文本；另一个值得跟踪的方向是事件报告制度的执行缺口——本周两起代理攻击事件均由外部研究者而非企业自行披露。对中国 AI 创业者，应把国产算力的迁移成本作为可量化项纳入规划（参照：某公司两周内把国产加速卡推理吞吐提升 3 倍、异构计算架构外部开发者占比过半达 61%、超节点已部署超 1000 套），采购谈判中应要求以每 token 成本而非单卡价格报价；在开源策略上应明确选

择而非摇摆，中间路线（宽松授权加高价接口）在本期数据下最难成立；海外端侧则是当前中国模型最具确定性的入口。

数据来源与口径

股价涨跌幅为 2026 年 9 月 11 日收盘至 9 月 18 日收盘，美股与指数为未复权直接计算，A 股与港股为前复权序列，跨市场比较时口径不完全一致，未换汇。金额以美元计时使用"亿美元"与"万亿美元"，以人民币计时使用"亿元"；模型价格统一为"美元/百万 token"，并区分输入、缓存读取、缓存写入与输出四项；算力规模使用 GW 与"万卡"；混合 token 价格的口径为输入三份加输出一份取均值。可信度分三档：高（官方文件、公开申报、公司财报与官方博客）、中高（主流财经媒体的独家报道或多家权威媒体交叉印证）、中（单一媒体报道或未获当事方确认）；可信度为"中"的单一来源信息不进入本期判断句。年化运行率为某月收入乘 12 的外推，不是会计年度收入；总合同价值为多年期名义金额，与年度收入的比值取决于合同年限与收入确认方式；混合专家架构模型须同时给出总参数与每 token 激活参数。图 4 的主题分布基于本期核验的 50 条可采信头条，同一条目归入其主要主题并已去重。本期排行榜、定价与行情数据均取自一手接口或官方文档，聚合类站点的数字未予采用。

本报告由百思特变革研究院 AI 产业智库基于公开信息编制，数据截至 2026 年 9 月 21 日。报告内容仅供行业研究参考，不构成任何投资建议，不代表任何政府机构或监管部门的立场；所涉政策信息均来自公开发布的法规、公告与新闻报道；引用请注明出处。