传媒行业专题报告

大模型：从智能能力的生产，到平台控制权与估值重构

方 正 证 券 研 究 所 证 券 研 究 报 告

分析师

焦娟 登记编号：S1220525090006

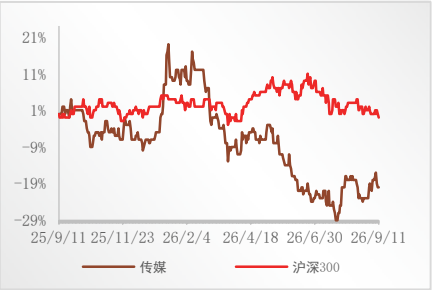
冯静静 登记编号：S1220525090003

行 业 评 级 ：推 荐

行 业 信 息

上市公司总家数	194
总股本(亿股)	3,142.45
销售收入(亿元)	5,222.61
利润总额(亿元)	4,824.87
行业平均 PE	45.81
平均股价(元)	19.14

行 业 相 对 指 数 表 现



数据来源：wind 方正证券研究所

相 关 研 究

《2026 年中报总结：业绩分化延续，AI 重构提速》2026.09.07

《AI 眼镜：下一代智能终端入口的产业逻辑与投资研判》2026.08.31

《党媒价值凸显的三层逻辑》2026.03.20

《GEO 逻辑是对的，央国企媒体公司是受益部位，看好党媒价值凸显》2026.01.18

大模型产业正在由“更强的模型”走向“更便宜、更可靠、更可调用的智能供给”，能力仍是基础，但产业价值的决定因素正在从模型本身向任务经济、工作流控制权、专有反馈闭环与资本效率延伸。随着模型能力持续提升、推理成本下降以及 Agent、全模态等技术不断演进，大模型正在从单一的软件能力演变为可嵌入软件、云、终端与业务流程的智能基础能力。基于这一变化，本篇专题从技术本体出发，进一步研究大模型的能力生产、成本结构、商业化路径、竞争格局及估值方法，试图回答模型能力如何形成、如何转化为真实任务价值，以及价值最终由谁获得。

从技术演进来看，大模型的竞争已经不再只是扩大参数规模，而是围绕能力、效率与真实环境验证形成系统性迭代。Transformer 仍是当前主干架构，MoE、稀疏化、长上下文、后训练、推理时计算以及工具与检索等技术不断提升模型的计算效率和任务能力，多模态也进一步向全模态演进。与此同时，模型能力的生产逐渐由“算力×数据×算法”转向“算力×数据×算法×工程闭环”，真实任务中的验证、反馈与迭代成为能力持续提升的重要来源。

能力提升和成本下降进一步改变了大模型的商业化方式，产业价值开始从模型输出本身向任务结果和工作流迁移。Token 仍是智能服务供给的基础计量单位，但衡量模型经济性的关键逐步转向每成功任务成本及其对应的客户价值。随着商业化由 API 向 MaaS、Agent 和任务结果计费演进，模型与企业工作流、云平台及终端入口的结合不断加深，长期价值也由单纯提供模型能力逐步转向获得稳定生产负载、工作流控制权和默认调用权。

在此基础上，中美大模型产业形成了不同的竞争路径，而公司的长期价值取决于其能否将自身资源禀赋转化为持续的能力优势和商业优势。美国主要公司依托算力、云、开发者生态及全球入口，在前沿模型、企业工作流、基础设施和开源生态等方向形成差异化布局；中国厂商则在供给约束下更加重视成本效率、产业场景、本地化服务与国产算力适配。由此，判断一家模型公司的竞争力不能只看模型能力，还需要结合客户、生态、入口及资本投入能力进行综合判断。

因此，大模型公司的估值逻辑也需要从“技术领先”进一步转向“价值捕获”，核心在于判断能力优势能否持续转化为收入、工作流控制权和资本回报。本篇据此建立覆盖能力持续性、单位任务经济、收入质量、工作流控制权、数据闭环、生态入口与资本效率的七维估值框架，并进一步建立动态跟踪体系，将模型能力、生产负载、默认调用权及资本效率纳入持续观察。

投资建议：基础模型仍具备平台价值，但并非所有模型公司都能够成为平台；中国市场重点关注国产加速器真实 TCO、海外 Token 使用结构、API 对私有化部署的替代速度以及 Coding/Agent 向平台能力扩展的可行性，科技巨头则需重点观察 AI 资本开支对应的算力利用率、收入增长和自由现金流回报。

风险提示：宏观环境及监管政策变化风险；AI 产业发展及商业化不及预期风险；相关公司竞争及资本投入回报不及预期风险。

正文目录

1 大模型的技术本体：智能实现的混合平台 4
2 Transformer 之后：技术路线如何继续进化 4
3 从多模态到全模态：现实世界接口 5
4 大模型能力怎样被生产 5
5 Token 经济学：训练成本、推理成本与模型定价 6
6 全球资本开支：科技巨头在为什么下注 7
7 美国大模型格局：六种战略物种 8
8 中国大模型格局：成本、生态、客户与国产算力 9
9 中美大模型全对比 9
10 大模型怎样变成生意：从 Token 到任务结果 10
11 大模型公司的估值方法 10
12 竞争的下一阶段：Agent、全模态与新入口 11
13 如何持续跟踪大模型产业 11
14 从基础模型到产业应用：AI 三层研究框架 12
15 产业展望与投资结论 12
16 风险提示 13

更多免费行业报告 并购家 www.ipoipo.net

图表目录

图表 1：大模型驱动智能范式演进的主要路径 4
图表 2：Transformer 之后大模型主要技术演进方向及跟踪指标 5
图表 3：从多模态到全模态：能力演进与应用价值 5
图表 4：大模型能力生产的关键要素及观察指标 6
图表 5：大模型厂商公开 API 价格（单位：美元/百万 Tokens） 6
图表 6：大模型成本下降的主要路径及验证指标 7
图表 7：2025 财年或自然年四大科技巨头资本开支（单位：十亿美元） 7
图表 8：AI 资本开支的价值转化链条 8
图表 9：美国大模型主要玩家的战略定位与商业模式 8
图表 10：中国大模型主要厂商的竞争卡位与跟踪重点 9
图表 11：中美大模型竞争环境与产业路径对比 9
图表 12：大模型商业模式与价值密度阶梯 10
图表 13：大模型商业化路径及价值捕获能力 10
图表 14：大模型公司的七维估值框架 11
图表 15：Agent 从辅助工具向现实世界交互的演进路径 11
图表 16：大模型产业的动态跟踪框架 12
图表 17：AI 应用三层研究框架 12
图表 18：大模型产业核心观察指标与验证信号 13

1 大模型的技术本体：智能实现的混合平台

大模型的关键价值并非只是提供对话能力，而在于将不同智能模块重新编排为可复用的通用平台。通过统一接口，大模型能够承接知识表达、模式识别、推理、代码、搜索、工具调用及任务规划等多类智能能力，使原本分散的智能能力能够像云计算资源一样被规模化调用。这一平台化特征，本质上体现为不同类型信息、能力、资源与应用价值的“混合”。

“混合”至少包含四个层面：

1）模态混合。文本、图像、音频、视频与结构化数据进入共同表征空间，使模型能够在统一的信息表征体系下处理多类型数据。

2）能力混合。生成、判断、规划、检索与执行等能力不再由彼此孤立的软件分别完成，而是通过模型统一编排，形成面向复杂任务的综合智能能力。

3）资源混合。模型能力与算力、数据、工具及企业权限等外部资源共同组成可调用的智能服务，使模型能够突破单纯的信息生成，进一步连接外部资源并参与实际任务执行。

4）价值混合。模型既可以作为 API 提供底层能力，也可以进一步成为产品、工作流、平台乃至硬件的智能入口，推动模型能力从底层技术向应用与终端持续延伸。

图表1:大模型驱动智能范式演进的主要路径

路径	典型载体	改变对象	核心价值	汇合方向
混合平台	大语言模型/Agent	认知与行动模块	降低智能生产与调用成本	进入空间、终端与机器人
重构时空	Apple Vision Pro/空间计算	人—信息—物理空间关系	重建交互、呈现与现场感	吸收模型的理解和行动能力

资料来源：方正证券研究所

研究含义：大模型不宜被视为单一的软件赛道，其本质更接近一种可嵌入软件、云、终端及业务流程的智能基础能力。因此，投资研究不能仅关注模型本身，还需同时关注其能力、供给、入口及价值分配。

2 Transformer 之后：技术路线如何继续进化

Transformer 之后，主干架构并未发生根本改变，但技术竞争的焦点已从“单一架构”转向“架构×后训练×系统工程”。

Transformer 解决了并行训练和长距离依赖的基础问题，但全注意力在超长上下文中的计算和内存成本迅速上升。产业迭代并非等待一个完全替代 Transformer 的新范式，而是在其基础上形成一组互补的技术升级：MoE 降低每次激活成本，稀疏注意力减少无效计算，长上下文扩展任务范围，检索与工具调用补充参数记忆，后训练和推理时计算提升任务完成质量。

更多免费行业报告 并购家 www.ipoipo.net
条款

图表2:Transformer 之后大模型主要技术演进方向及跟踪指标

方向	解决问题	潜在收益	新风险	跟踪指标
MoE/稀疏化	参数扩张与激活成本	能力/计算效率提升	路由、负载均衡与通信复杂	激活参数、吞吐、利用率
长上下文	跨文档和长任务	减少切片与信息丢失	“标称长度”不等于有效利用	长程检索准确率
后训练/RL	把基座能力变成任务能力	推理、Coding、Agent 提升	奖励偏差、环境故障	任务成功率、训练稳定性
推理时计算	难题上投入更多思考	能力按需扩展	延迟与 Token 成本上升	质量—成本曲线
工具/检索	连接实时知识与外部系统	降低幻觉、获得行动能力	权限、安全与错误传播	调用成功率、可验证率

资料来源：方正证券研究所整理

因此，“有没有新架构”不是最好的研究问题。更有效的问题是：同等任务成功率下，谁消耗更少计算？同等成本下，谁能完成更长、更复杂的任务？谁能把失败样本快速转为训练信号？这三问把技术路线直接连接到产业价值。

3 从多模态到全模态：现实世界接口

模态数量不是终点，统一表征、跨模态推理与行动闭环才是价值所在。

多模态通常指模型可以处理两种及以上数据类型；全模态则是更强的系统目标：不同模态在统一或高度协同的表征空间中被理解，模型能够跨模态推理，保持时间连续性，并以低延迟生成多模态输出或调用工具。它不是“文件格式清单更长”，而是世界模型与任务接口的升级。

图表3:从多模态到全模态：能力演进与应用价值

能力层	多模态	全模态目标	应用价值
输入	识别文本、图片、音频等	连续感知多源信号	把现场、设备、人的行为数字化
表征	模态编码后对齐	统一/协同向量空间	跨模态检索与因果关联
推理	对单次混合输入回答	跨时间、跨工具、跨模态推理	完整理解工作上下文
输出	生成文字/图片等	实时原生多模态生成	自然交互与内容生产
行动	有限工具调用	感知—判断—行动—验证闭环	进入生产、服务与物理场景

资料来源：方正证券研究所整理

真实价值可归纳为三条：表达型内容被调用，生产型内容直接进入设计、营销、软件和运营流程，指令型内容进一步参与资源调度和决策。AIGC 中的“生成 C”会逐渐像敲键盘一样必要却不稀缺，稀缺性将迁移至生成内容能否被调用、形成有效生产并承担决策责任。

4 大模型能力怎样被生产

能力生产函数 = 算力 × 数据 × 算法 × 工程闭环。任何一项接近零，整体产出都受限。

图表4:大模型能力生产的关键要素及观察指标

生产要素	传统观察	新的关键口径	为什么重要
算力	GPU 数量	有效计算、网络/存储、MFU、可用性	决定实验密度与单位成本
数据	Token 数量	能力增益/样本、专家轨迹、失败数据	决定算力是否烧在高价值信号上
算法	论文/参数	质量—成本前沿、推理时策略	决定资源转化效率
工程	上线速度	验证环境、异常处理、迭代周期	决定方法能否持续运行
产品反馈	DAU/调用量	可训练信号密度、反馈时延	决定数据飞轮是否真实存在

资料来源：方正证券研究所

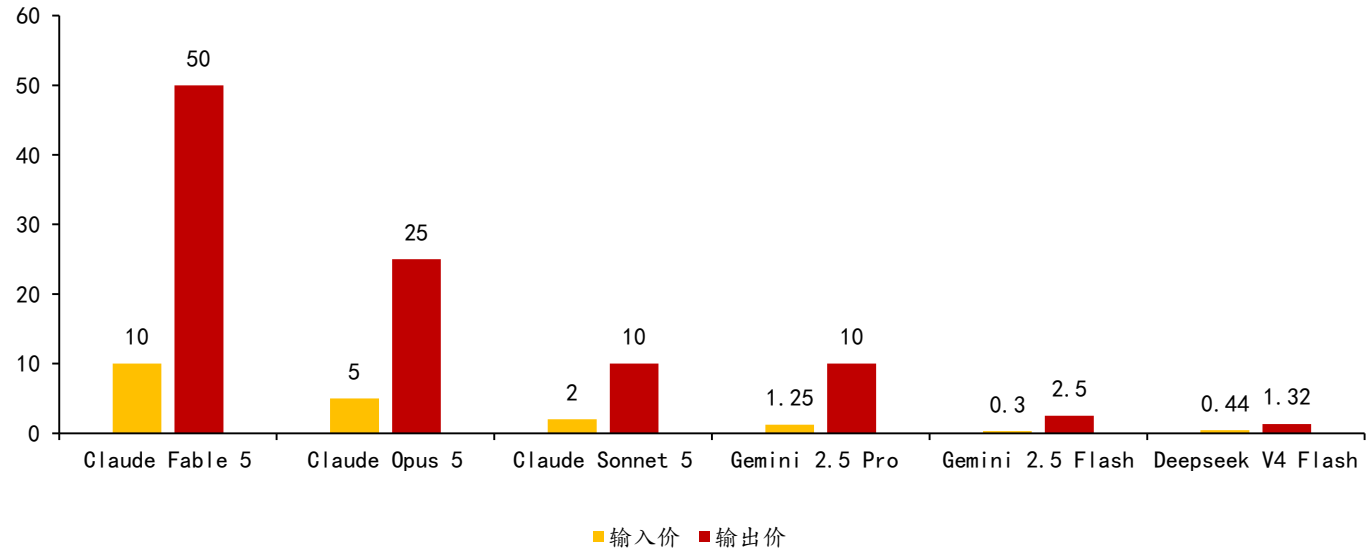
预训练架构逐渐成熟后，竞争重心转向后训练与真实环境验证。高质量数据也从“真实、干净、权威”升级为“能提升某项可验证能力”。专家决策、任务轨迹、失败位置和修正过程，比单一正确答案更稀缺。

以 Coding 模型为例，其能力提升不仅依赖代码生成本身，还依赖真实代码库和运行环境中的持续验证：模型生成代码后，需要通过测试发现问题，并进一步区分模型错误与环境错误，再根据反馈进行修正和迭代。由此，真实环境中的验证、异常处理与反馈迭代，构成了模型能力持续提升的重要工程闭环。

5 Token 经济学：训练成本、推理成本与模型定价

Token 是智能服务供给的基础计量单位，但只有“每成功任务成本”才能真正连接模型成本与客户 ROI。

图表5:大模型厂商公开 API 价格（单位：美元/百万 Tokens）



资料来源：Anthropic、Google、DeepSeek 官方定价，方正证券研究所。注：DeepSeek-Flash 取价为高峰时段—缓存未命中档。

基础模型价格下降并不等于全行业通缩。定价正在呈现两极分化：可替代、低切换成本的通用调用更多承担获客功能；高可靠、高性能、深度嵌入企业流程的模型与服务则保持溢价。输出价格通常高于输入价格，一定程度上反映了生成阶段更高的计算负担及相关供给约束。

更多免费行业报告 并购家 www.ipoipo.net

更严谨的单位经济模型应计算：

单次尝试成本=有效输入 Token 成本+输出 Token 成本+工具/检索/环境成本；

每成功任务成本=单次尝试成本 ÷ 任务成功率；

任务净价值=人工替代价值或新增收入-每成功任务成本-集成与治理成本。

图表6:大模型成本下降的主要路径及验证指标

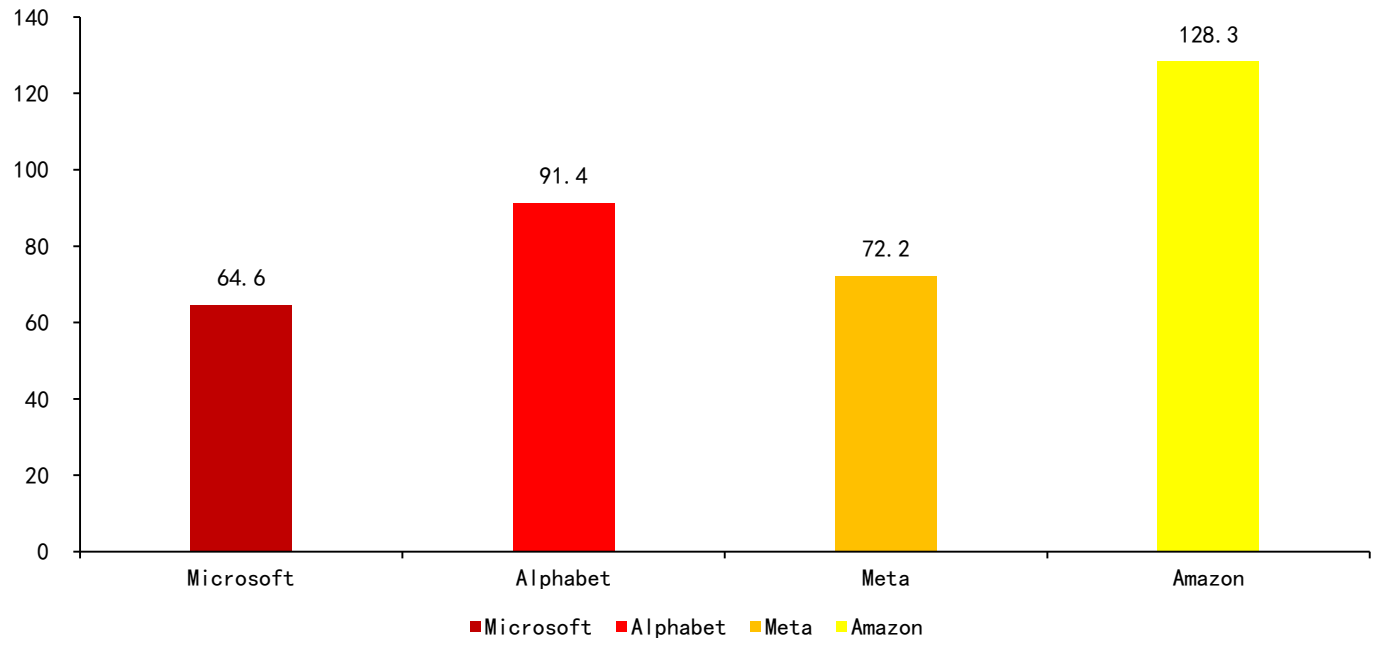
降本来源	机制	可持续性	验证指标
缓存	复用相同前缀/KV	取决于业务结构	命中率、命中价差
量化/蒸馏	降低精度与模型规模	存在能力边界	质量损失、吞吐
稀疏化/MoE	减少激活计算	工程门槛较高	激活比例、通信开销
基础设施利用率	提高集群有效负载	形成运营知识	MFU、闲置率、故障率
国产加速器	降低供给约束与成本	依赖软硬协同	适配模型数、TCO

资料来源：方正证券研究所

6 全球资本开支：科技巨头在为什么下注

CapEx 购买的不是当期收入，而是未来模型实验、云供给与入口防御的选择权。

图表7:2025 财年或自然年四大科技巨头资本开支（单位：十亿美元）



资料来源：各公司财报，方正证券研究所。注：Microsoft 为财年口径，Alphabet、Meta、Amazon 为自然年口径，主要用于观察资本开支趋势，横向比较时需注意口径差异。

2025 年以来，全球科技巨头资本开支持续处于高位。Microsoft FY2025 物业与设备新增额为 645.51 亿美元，Alphabet 2025 年资本开支为 914 亿美元，Meta 2025 年为 722.2 亿美元；Amazon 2025 年资本开支为 1283 亿美元。2025 年 2 月，阿里巴巴集团 CEO 吴泳铭宣布未来三年投入 3800 亿元人民币建设云和 AI 基础设施。资本开支已将模型竞争升级为资产负债表竞争。

进入 2026 年，巨头资本开支继续加速：Microsoft FY2026（财年截止 2026 年 6 月 30 日）物业与设备新增额达 1,159 亿美元，同比约+80%；Alphabet 2026 年上半年资本开支约 806 亿美元（Q2 单季 449 亿创纪录），并将 2026 全年指引上调至

1,950 亿—2,050 亿美元；Meta 2026 年资本开支指引收窄至 1,300 亿—1,450 亿美元（Q2 单季约 311 亿）；Amazon 据 2025 年 Q4 财报将 2026 年资本开支计划提升至约 2,000 亿美元。2026 年 5 月，吴泳铭进一步表示，阿里未来三年的资本开支“可能远超”此前承诺的 3,800 亿元。

但“投入越多=模型越强=股东回报越高”并不成立。因此，研究不能只看资本开支规模，而需完成四层拆分：投入去了服务器还是建筑与网络；有多少支持训练、有多少支持推理和传统云；产能何时上线；收入和现金流能否覆盖折旧与融资成本。

图表8:AI 资本开支的价值转化链条

观察层	正向信号	风险信号
建设	服务器、网络与数据中心同步交付	土地/电力/网络成为瓶颈
利用	训练实验和推理负载快速爬坡	利用率不足、客户需求错配
商业化	云收入、AI 产品收入和客户扩张	用量增长但价格降幅更快
财务	毛利改善、现金回收期可控	折旧陡升、自由现金流承压
战略	形成模型—云—应用—入口闭环	仅扩张算力而缺少默认入口

资料来源：方正证券研究所

7 美国大模型格局：六种战略物种

竞争不是模型榜单排位，而是不同资产、入口和商业模式之间对智能产业控制权的争夺。

图表9:美国大模型主要玩家的战略定位与商业模式

公司	战略物种	控制权来源	变现路径	关键矛盾
OpenAI	前沿模型+原生入口	ChatGPT、开发者与迭代闭环	订阅/API/企业/Agent	资本强度与治理
Anthropic	可靠性+Coding Agent	高价值开发工作流	API/企业/任务使用	算力与分发依赖
Google	全栈技术+搜索入口	模型、数据、云、Android	广告/云/API/订阅	创新与自我替代
Microsoft	企业权限与工作流	Azure、Office、GitHub	云/席位/增量功能	CapEx 回报
Amazon	基础设施与模型市场	AWS 客户、芯片与平台	算力/MaaS/API	前沿模型品牌
Meta	开源+社交分发	用户规模、广告与硬件	广告效率/生态/终端	直接模型收入

资料来源：方正证券研究所整理

OpenAI 与 Google 最接近技术方向制定者；Anthropic 把可靠性和 Coding Agent 变成高价值楔子；Microsoft 与 Amazon 控制企业计算与云入口；Meta 通过开源扩散降低封闭模型的生态租值。六种物种不会简单收敛为一种，因为它们的存量资产决定了各自最优策略。

更多免费行业报告 并购家 www.ipoipo.net

8 中国大模型格局：成本、生态、客户与国产算力

中国竞争的关键不是复刻美国，而是在供给约束下形成成本、场景与本地化优势。不同厂商基于自身的客户、生态与技术积累，形成差异化的竞争卡位，投资研究的重点也由单纯比较模型能力，转向观察成本优势、生态协同、客户渗透与国产算力适配能力。

图表10:中国大模型主要厂商的竞争卡位与跟踪重点

厂商	关键卡位	优势	主要约束	跟踪重点
DeepSeek	低成本前沿模型	系统优化、开源品牌、价格领导力	企业深度渗透与持续投入	原生推理效率、Agent 产品
阿里/千问	模型+MaaS+云生态	算力、模型供给、电商与企业客户	模型差异化持续期	百炼生产负载、AI 收入质量
智谱	Coding/Agent+政企	工程闭环、API 转型、本地化	多模态与头部竞争	真实任务、Token 结构、海外
MiniMax	高性价比模型+C 端	稀疏架构、Agent/内容产品	算力与架构迭代风险	M3 系列与 ARR 质量
字节	多模态+应用分发	内容数据、产品、火山平台	商业化算力供给	豆包/即梦与企业渗透
腾讯	社交/内容/企业连接	微信入口、广告与产业连接	模型心智和产品放量	混元与应用协同

资料来源：方正证券研究所

核心判断：中国模型公司的可投资性不能只看“有没有最强模型”，而要看其是否把成本优势转化为稳定用量，把本地化需求转化为 API/平台收入，并在国产算力约束下保持迭代速度。

9 中美大模型全对比

中美大模型竞争并非同一跑道上的距离比较，而是资源禀赋、客户结构与制度环境差异共同作用下形成的不同竞争路径。

图表11:中美大模型竞争环境与产业路径对比

维度	美国	中国	研究结论
前沿算力	规模与先进芯片占优	供应约束更强	美国实验空间更大
推理成本	高端模型保有溢价	价格竞争与系统优化更强	中国形成成本压力
模型路线	闭源前沿+开源并存	开源、本地化与快速追赶	战略目的不同
客户	全球开发者/大型企业	国央企/互联网/本地产业	销售与交付函数不同
商业化	订阅、云、广告、办公齐全	云、私有化、API、行业交付	美国经常性更成熟
入口	搜索、OS、云、办公、社交	超级 App、电商、云、终端	双方都有存量入口
监管	版权、隐私、全球合规	备案、安全、本地化	合规成本结构不同
资本市场	高估值与融资深度	融资与退出约束更强	资本耐力差异明显

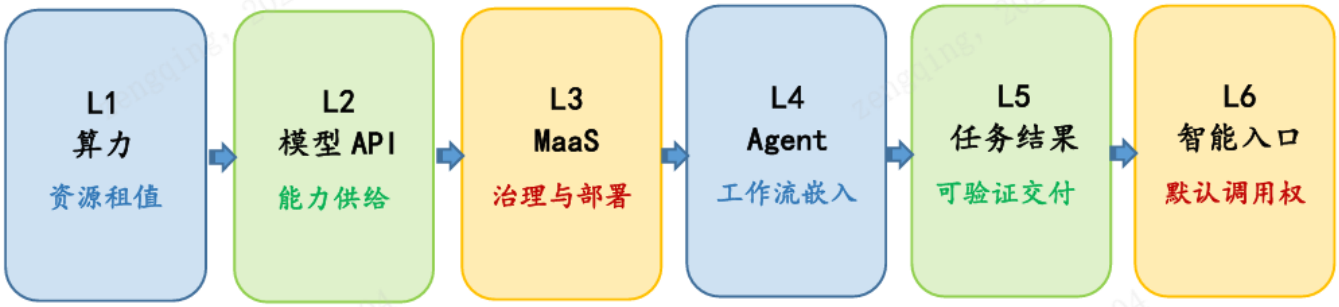
资料来源：方正证券研究所

判断中美差距时，应从“单一能力差距”升级为“能力—成本—供给—商业化—入口”的五层比较框架，并结合客户、监管与资本市场环境理解其差异。某一国产模型在 Coding 上接近第一梯队，并不等于基座全面领先；同样，美国模型定价更高，也不等于单位任务价值更低。

10 大模型怎样变成生意：从 Token 到任务结果

产业价值沿“资源—能力—平台—工作流—结果—入口”逐级上移。

图表12:大模型商业模式与价值密度阶梯



价值密度向右提升；交付责任、资本要求与系统复杂度同步提高

资料来源：方正证券研究所

Token 计费最容易标准化，但也最容易被价格竞争侵蚀。MaaS 通过统一供应商管理、部署、监控和安全提高客户切换成本；Agent 把模型嵌入工作流；任务结果计费则把价值锚定在客户 ROI。最上层的智能入口拥有默认调用权，能够决定模型、工具、内容和服务如何被分发。

企业采购决策将围绕四个问题：是否比人工或旧软件便宜；是否在 12—18 个月内形成可验证回报；是否满足权限、安全与责任要求；供应商能否保障稳定算力和问题定位。政府及国企更重视本地化与合规，私营企业更强调 ROI，二者共同推动市场从 Demo 走向生产负载。

图表13:大模型商业化路径及价值捕获能力

商业化阶段	收入质量	毛利潜力	客户锁定	主要风险
Token/API	中	取决于成本曲线	低—中	价格战与同质化
MaaS	中—高	平台层较高	中	平台能力趋同
Agent 席位	高	取决于推理用量	中—高	活跃度与任务成功率
任务结果	高	价值定价空间大	高	责任与交付复杂
智能入口	极高	平台租值	极高	监管与生态冲突

资料来源：方正证券研究所整理

11 大模型公司的估值方法

估值对象不是参数，而是可持续智能供给、工作流控制权与资本效率。

传统软件的席位和 ARR 框架仍有参考意义，但直接套用会遗漏模型行业的资本强度、推理可变成本和快速技术迭代。另一方面，只以训练投入或模型榜单估值，又会高估无法商业化的能力。因此，需要建立覆盖能力、成本、收入、工作流、数据、入口与资本效率的七维估值框架。

更多免费行业报告 并购家 www.ipoipo.net

图表14:大模型公司的七维估值框架

价值维度	权重建议	核心问题	正向证据	证伪信号
能力持续性	16%	领先能否重复?	迭代速度、复杂任务领先	单代领先后停滞
单位任务经济	16%	每成功任务成本?	成本下降快于价格	Token 便宜但失败率高
收入质量	14%	收入是否留存扩张?	生产负载、净留存	测试用量和补贴驱动
工作流控制权	16%	嵌入多深?	权限、记录系统、决策节点	仅是可替换插件
数据闭环	14%	反馈能否训练?	高质量轨迹与验证	海量低价值日志
生态入口	14%	是否默认被调用?	云、OS、应用、终端入口	依赖第三方分发
资本效率	10%	投入回报如何?	能力/收入/现金流改善	折旧与融资吞噬价值

资料来源：方正证券研究所整理

估值倍数应与七维评分联动：高增长但低控制权、低资本效率的 API 收入，价值可能低于增长较慢但深嵌工作流的平台收入。对巨头则需要采用分部估值和增量现金流框架，区分 AI 对原有广告、云、办公、电商的防守价值与新收入价值。

12 竞争的下一阶段：Agent、全模态与新入口

模型从“回答问题”走向“完成工作”，再进入实时感知与物理世界。

Agent 并非简单在模型外套一个工作流。更深的方向是将规划、工具选择、记忆、异常修复和验证等能力逐步内化为模型的原生能力。全模态则补上现实世界的感知接口，使 Agent 不仅能够阅读网页和代码，也能理解视频、声音、设备状态和空间关系。两者进一步汇合后，智能供给将从内容生成走向任务执行，并逐步延伸至现实世界。

图表15:Agent 从辅助工具向现实世界交互的演进路径

阶段	能力形态	产品形态	价值锚	关键瓶颈
Copilot	人主导、AI 辅助	聊天框/插件	节省时间	准确性与使用频率
Workflow Agent	AI 执行流程、人审批	Coding/办公/客服 Agent	任务完成	工具稳定与权限
Native Agent	模型内生规划与修复	通用任务代理	工作结果	长任务可靠性
Embodied/Spatial	实时感知并作用物理世界	眼镜/机器人/垂类硬件	场景重构	安全、成本与责任

资料来源：方正证券研究所

新入口的竞争标准也发生变化：互联网时代看流量，AI 时代看权限。内容能否抵达用户，取决于用户所使用的模型和 Agent 是否拥有调用权。模型、操作系统、云、企业记录系统、超级 App 和垂类硬件，都可能成为智能分发的新关口。

13 如何持续跟踪大模型产业

研究不是静态排名，而是“事实—机制—验证—估值”的动态决策系统。

图表16:大模型产业的动态跟踪框架

频率	跟踪内容	核心数据	输出
周度	模型/产品/价格/开发者反馈	榜单仅作线索、API 价、延迟、故障	事件卡与待验证假设
月度	Token、用户、客户与产品迭代	调用结构、生产负载、海外占比	公司更新卡
季度	收入、毛利、CapEx 和折旧	AI 收入、云增速、现金流	盈利与估值修正
半年度	技术路线和竞争位置	复杂任务、成本前沿、入口变化	行业格局重估
年度	资本回报与产业控制权	ROIC、自由现金流、生态份额	投资框架迭代

资料来源：方正证券研究所

每条信息必须回答四个问题：它是事实还是观点；可靠性等级是什么；改变了哪条机制；什么条件会证伪。

14 从基础模型到产业应用：AI 三层研究框架

大模型专题聚焦 AI 产业链中的基础模型层，重点研究大模型的技术演进、能力生产、商业化与价值创造机制。在此基础上，与此前 AI 应用专题研究形成上下游衔接：前者关注基础智能能力如何形成与供给，后者进一步关注模型能力如何被组织、应用并转化为产业价值。二者并列构成完整的 AI 应用框架。

图表17:AI 应用三层研究框架

层级	研究对象	价值来源	典型指标
第一层	基础大模型	智能能力生产与供给	能力、单位成本、资本效率、入口
第二层	垂类模型+Agent+MaaS+平台	把模型能力组织成可用服务	任务成功率、工作流渗透、留存
第三层	利用 AI 重塑自身业务的公司	改善原有生态位或进入新生态	供给效率、新业务收入、价值链地位

资料来源：方正证券研究所

围绕三层框架，完整研究流程分四步。第一，比较公司在 PC/移动互联网时代与 AI 浪潮中的生态卡位和资源优势，判断其能否纳入三层框架及具体所处层级。第二，将公司的各类动态同时放入“大模型 13 讲”和“AI 应用 13 讲”进行解释，避免孤立解读。第三，在存量时代从规模增长转向价值增值，关注存量客户价值、单位经济、现金流质量、生态位和资本效率。第四，通过成功重估、失败案例和同行对照，不断校正指标权重。

在三层研究框架下，AI 对公司的影响最终体现为生产效率、业务模式与生态位的变化。研究重点并非简单判断公司是否应用 AI，而是进一步分析其能否将 AI 生产力转化为更强的交付能力、更高质量的现金流以及更具持续性的生态位和控制权。

15 产业展望与投资结论

能力升级仍在持续，但资本市场将越来越严格地区分“技术领先”与“价值捕获”。

产业展望：产业竞争的关注重点正在经历阶段性迁移：第一阶段关注模型能力与训练规模；第二阶段关注推理成本、Agent 与生产负载；第三阶段进一步关注默认调用权、任务结果与资本回报。不同阶段并非相互替代，而是逐步叠加，估值权重也在持续向后者迁移。

更多免费行业报告 并购家 www.ipoipo.net
条款

基础模型仍具备平台价值，但并非所有模型公司都能够最终成为平台。缺少流量或调用入口、专有反馈闭环与资本耐力的公司，长期可能被压缩为低毛利的能力供应商。

云厂商的优势在于算力、企业客户与完整服务能力，原厂模型的优势则在于源码级优化、缓存与问题定位，二者之间的竞合关系仍将长期存在。

投资建议：中国大模型产业值得重点跟踪的变量包括国产加速器的真实 TCO、海外 Token 使用结构、API 对私有化部署的替代速度，以及 Coding/Agent 的单点能力突破能否进一步扩展为平台能力。

巨头资本开支能否形成正向回报，需要同时观察算力上线、利用率提升、AI 收入增长或原有业务效率改善，以及自由现金流对投入的承受能力。若其中关键环节未能兑现，持续扩张的 CapEx 也可能从竞争壁垒转化为折旧与融资压力。

图表18:大模型产业核心观察指标与验证信号

优先观察	为什么	领先信号	风险信号
任务成功率	连接能力与客户价值	长任务稳定提升	榜单高但生产失败
每成功任务成本	连接价格与毛利	成本下降快于售价	低价依赖补贴
生产负载	区分试用与商业化	客户扩张和重复调用	大量测试账户
默认调用权	决定长期价值分配	进入 OS/云/工作流	被上层平台替换
资本效率	决定股东回报	收入/现金流跟上投入	折旧与融资压力

资料来源：方正证券研究所

16 风险提示

宏观环境及监管政策变化风险；AI 产业发展及商业化不及预期风险；相关公司竞争及资本投入回报不及预期风险。

分析师声明

作者具有中国证券业协会授予的证券投资咨询执业资格，保证报告所采用的数据和信息均来自公开合规渠道，分析逻辑基于作者的职业理解，本报告清晰准确地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。研究报告对所涉及的证券或发行人的评价是分析师本人通过财务分析预测、数量化方法、或行业比较分析所得出的结论，但使用以上信息和分析方法存在局限性。特此声明。

免责声明

本研究报告由方正证券制作及在中国（香港和澳门特别行政区、台湾省除外）发布。根据《证券期货投资者适当性管理办法》，本报告内容仅供我公司适当性评级为 C3 及以上等级的投资者使用，本公司不会因接收人收到本报告而视其为本公司的当然客户。若您并非前述等级的投资者，为保证服务质量、控制风险，请勿订阅本报告中的信息，本资料难以设置访问权限，若给您造成不便，敬请谅解。
在任何情况下，本报告的内容不构成对任何人的投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需求，方正证券不对任何人因使用本报告所载任何内容所引致的任何损失负任何责任，投资者需自行承担风险。
本报告版权仅为方正证券所有，本公司对本报告保留一切法律权利。未经本公司事先书面授权，任何机构或个人不得以任何形式复制、转发或公开传播本报告的全部或部分内容，不得将报告内容作为诉讼、仲裁、传媒所引用之证明或依据，不得用于营利或用于未经允许的其它用途。如需引用、刊发或转载本报告，需注明出处且不得进行任何有悖原意的引用、删节和修改。

评级说明：

类别	评级	说明
公司评级	强烈推荐	分析师预测未来12个月内相对同期基准指数有20%以上的涨幅。
	推荐	分析师预测未来12个月内相对同期基准指数有10%以上的涨幅。
	中性	分析师预测未来12个月内相对同期基准指数在-10%和10%之间波动。
	减持	分析师预测未来12个月内相对同期基准指数有10%以上的跌幅。
行业评级	推荐	分析师预测未来12个月内行业表现强于同期基准指数。
	中性	分析师预测未来12个月内行业表现与同期基准指数持平。
	减持	分析师预测未来12个月内行业表现弱于同期基准指数。
基准指数说明		A股市场以沪深300 指数为基准；香港市场以恒生指数为基准，美股市场以标普500指数为基准。

方正证券研究所联系方式：

北京：朝阳区朝阳门南大街 10 号兆泰国际中心 A 座 17 层
上海：静安区延平路71号延平大厦2楼
深圳：福田区竹子林紫竹七道光大银行大厦31层
广州：天河区兴盛路12号楼隽峰苑2期3层方正证券
长沙：天心区湘江中路二段36号华远国际中心37层
网址：https://www.foundersc.com
E-mail：yjzx@foundersc.com

更多免费行业报告 并购家 www.ipoipo.net