

AI产业链剖析，阵营分化谁能笑到最后

报告日期：2026-06-01

研究部

姓名：李承儒

SFC：BLN914

电话：0755-21519182

Email: licr@gyzq.com.hk

写在前面

- 本报告围绕AI产业展开分析，研究视角聚焦互联网软件领域，主要针对算力、应用服务、平台产品、商业模式、生态体系等软件及服务相关内容进行解读。未涉及芯片、终端设备、算力硬件、配套元器件等硬件产业链相关分析，敬请读者知悉。
- 本篇报告我们先简单回顾之行业发展进程。其中重点分析了通用大模型、智算中心、AI Agent细分领域的竞争格局以及发展前景。最后对于细分领域的相关标的进行分析和梳理。
- 感谢您对国元国际长期以来的支持，我们认为AI所引领的全球科技革命浪潮仍在持续，当前各行各业在AI科技领域的坚定投入，以及市场愿意给予AI创新企业更高的估值溢价，都为真正具备核心技术和商业化能力的公司提供发展土壤，因此当前是甄选优质资产、进行战略性布局的良好窗口期。

一、板块行情回顾

二、大模型赛道资本狂飙，商业化预期高涨

三、智算中心产业爆发式增长，算力租赁迎来风口

四、AI应用故事还未结束，Agent技术重塑业务价值

五、行业展望与战略建议

六、风险提示

复盘：AI叙事的变化走到哪个阶段？

近三年情绪主导下的算力超前投入，AIcapex叙事贯穿美股始终，但算力需求预期主导

➤ 2023年语言模型规模化，AI模型训练主导算力投资需求

- Chatgpt周活跃用户快速破1亿，GPT功能实现飞跃，使得语言模型引领时代升级的概念深入人心。

- 训练主导的算力投资：GPU相关企业业绩大增、云计算厂商CAPEX恢复投入。

➤ 2024年多模态模型训练使得推理需求崛起，AI融合和Agent成为发展方向。

- OpenAISora发布；Meta在升级AI广告系统后收入增长超预期；企业级服务agent开始落地。

- 市场认可推理需求趋势，伴随着算力需求大幅提升，引起模型训练算力长期需求的担心

➤ 2025年Token经济。OpenAI、Genimi、Anthropic、Deepseek普及，全球开始大量调用Token数据

- 年初Deepseek横空出世并首次在全球范围内开源，5月以来Token数据暴增持续验证推理需求。企业服务Agent落地缓慢。

- 9月OpenAI循环投资引发AI泡沫担忧，Gemini3.0将发布，不同模型阵营的算力和云厂走势分化。

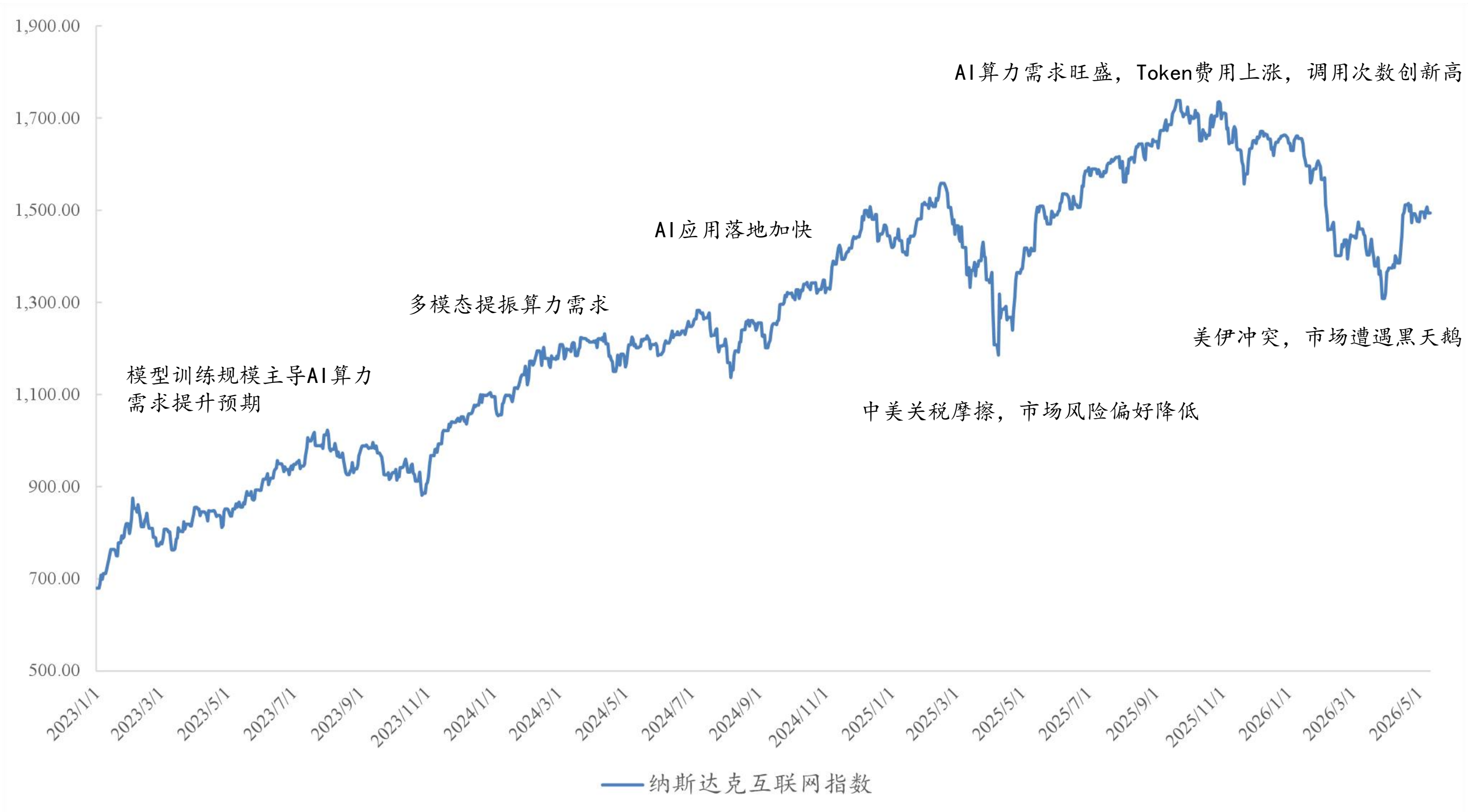
➤ 2026年投资追求回报

- Anthropic上调收入增长预期，Anthropic预计2027年现金流转正，估值逼近1万亿美元，超过OpenAI。

当下中美AI发展的路线出现分歧。中国更强调算力效率优先，更多依赖算法优化、开源生态和工程化提高效率，美国更强调“物质学科范式”，以技术自由主义为主导。尽管目前仍缺算力，但国产替代算力在投资视角可能更为占优，而大模型追赶也是重要主线。二是，经济结构不同决定中美AI冲击的传导路径存在差异。美国由于服务业占比高、白领人工成本昂贵，需要面临短期AI带来的替代性冲击和通缩压力。

复盘：AI叙事的变化走到哪个阶段？

图：海外互联网指数波动与AI技术落地息息相关



市场趋势与资本动态：AI概念企业上市潮与价值重估

- 资本化进入新阶段，香港成为全球AI创新集资中心。2025年至2026年初，港股市场迎来了一波AI企业的密集上市潮，包括滴普科技、明略科技、智谱AI、MiniMax等在内的多家重磅企业相继挂牌，显著提升了市场的AI含金量。这一现象表明，AI产业的资本化已进入新阶段，香港凭借其开放的融资环境、完善的监管体系以及连接东西方的桥梁作用，正成为内地AI企业走向全球、获取国际资本支持的首选平台。资本市场对AI赛道的情热持续高涨，AI相关主题指数在过去一年中表现强劲，显著跑赢大盘，如恒生人工智能主题指数在2025年录得了高达34.6%的涨幅。
- 投资逻辑转变：从“技术叙事”到“商业价值”。资本市场的偏好正发生深刻变化。早期投资更关注技术的颠覆性与模型的参数规模，而当前，投资逻辑已转向关注AI技术能否有效解决垂直领域的具体痛点、能否创造可衡量的商业价值以及是否具备可持续的盈利模式。市场不再仅仅为“可能性”买单，而是更加看重“确定性”，即企业的客户基础、收入增速、毛利率和续费率等核心经营指标。
- 企业AI预算激增，应用从“实验”走向“标配”。随着AI技术成本的下降和应用门槛的降低，全球企业正以前所未有的力度增加AI领域的预算投入。AI应用正迅速从少数头部企业的“创新实验室”走向各行各业的“标准配置”，广泛渗透到研发、生产、营销、客服、办公自动化等核心业务流程中，以提升效率、优化决策并创造新的增长点。

港股AI服务上市公司商业模式及技术路径对比分析

当前港股AI服务上市公司的商业模式呈现出多元化特征：

- **平台赋能型**：以**第四范式**为代表，核心是提供一个强大的AI开发与运行平台，降低企业应用AI的门槛，通过平台授权、订阅费和技术服务费获取收入。这种模式的规模效应强，但前期平台研发投入巨大。
- **解决方案驱动型**：以**创新奇智**和**滴普科技**为代表，更侧重于针对特定行业的痛点，提供“交钥匙”式的端到端解决方案。这种模式的客户粘性高，项目价值大，但对行业理解和交付能力要求极高。
- **技术授权型**：以**智谱AI**、**MiniMax**和**云知声**为代表，核心是将其领先的底层AI能力（如大模型、语音识别）通过API或MaaS平台进行输出，按调用量或订阅方式收费。这种模式能够触达广泛的开发者和企业，但面临激烈的技术竞争。
- **AI内嵌型**：以**金蝶国际**为代表，将AI能力作为其现有成熟软件产品（如ERP）的增值模块，提升产品竞争力和客户价值。这种模式的商业化路径清晰，能够利用现有渠道和客户基础快速变现。
- **软硬一体型**：以**云知声**为代表，不仅提供软件和云服务，还设计和销售搭载其AI算法的芯片和硬件模块。这种模式能够构建更深的护城河，但对供应链和硬件研发能力提出了更高要求。

各公司的技术路径选择也反映了其不同的战略定位：

- **底层基座模型派**：**智谱AI**和**MiniMax**是坚定的底层技术自主研发者，致力于打造性能卓越、安全可控的通用大模型。它们的成功与否直接关系到中国在AI核心技术领域的国际竞争力。
- **行业模型深耕派**：**第四范式**和**明略科技**则认为通用模型无法完全满足复杂的行业需求，它们的技术路径是构建深度融合了行业知识的专有模型和知识图谱，追求在特定领域的“认知深度”。
- **特定模态专精派**：**商汤科技**起家于并持续领先于计算机视觉领域，而**云知声**则在智能语音赛道建立了深厚的壁垒。它们在特定信息处理模态上的技术优势是其核心竞争力。
- **应用工程与整合派**：**滴普科技**和**金蝶国际**更侧重于AI技术的应用工程能力，即将领先的AI技术（无论是自研还是第三方）与企业的数据和业务流程进行高效整合，快速创造商业价值。

一、板块行情回顾

二、大模型赛道资本狂飙，商业化预期高涨

三、智算中心产业爆发式增长，算力租赁迎来风口

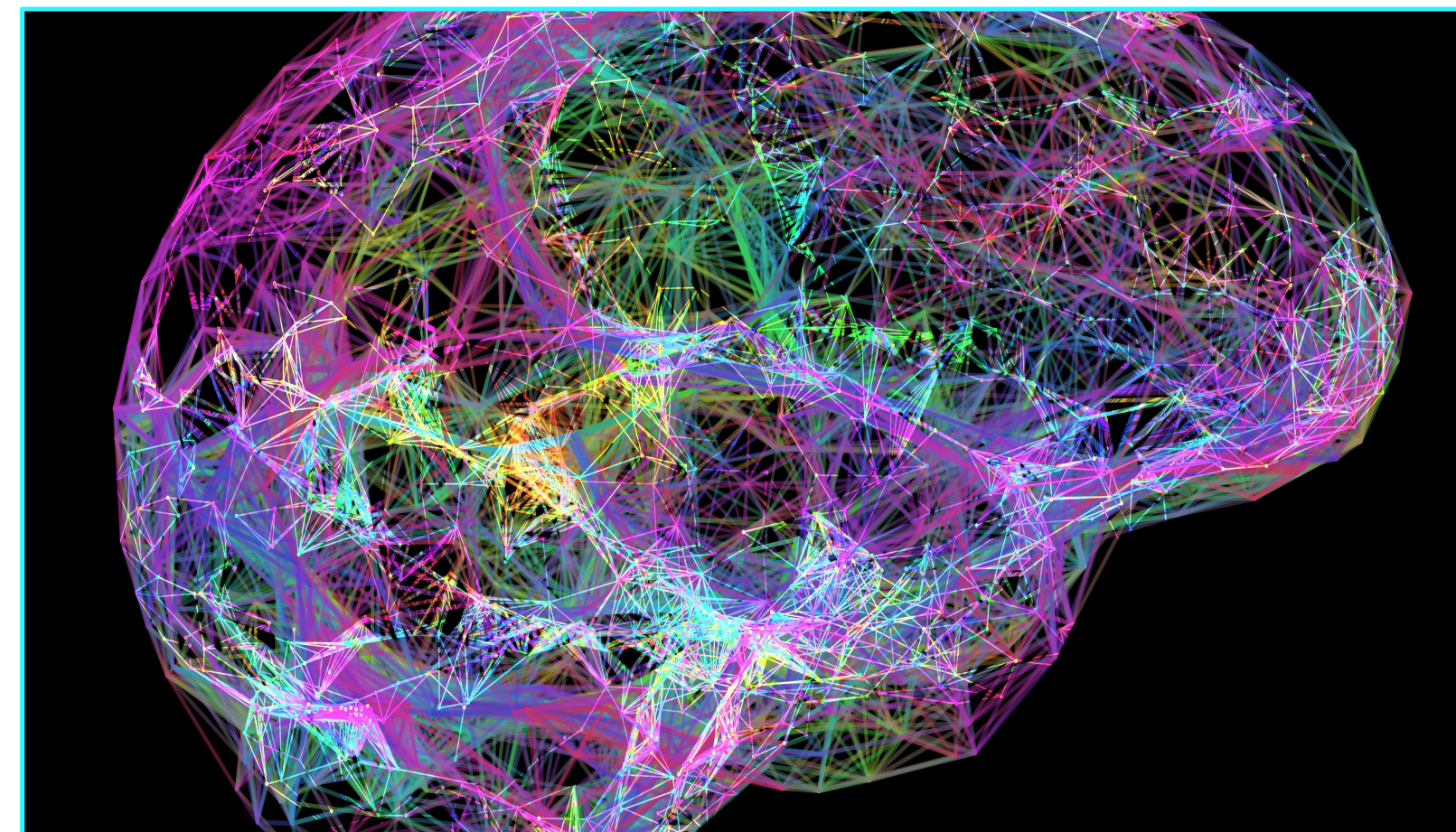
四、AI应用故事还未结束，Agent技术重塑业务价值

五、行业展望与战略建议

六、风险提示

大语言模型技术对AI市场规模增长起到核心驱动作用

LLM技术能够赋予AI Agent理解意图、逻辑思考与持续进化的核心能力，使其具备语言理解、生成和推理能力，成为现代AI Agent的理想“大脑”，



一 自然语言接口

充当Agent的“语言中枢”，实现流畅、准确的人机交互，精准理解用户指令意图。

推理与规划器

作为“逻辑大脑”，将复杂的长期任务拆解为可执行的子步骤（思维链/CoT）。

工具调用者

充当“手眼延伸”，自主判断并调用外部API、计算器或搜索工具以扩展解决能力。

世界模型与知识库

作为“记忆库”，提供海量事实性知识与通用常识，支持Agent的背景推理。

训练方式

通常在预训练LLM基础上，通过指令微调或上下文学习等技术，赋予模型角色扮演能力，使其能够稳定、可靠地充当AI Agent的核心驱动。

全球大模型格局已定

截至2026年，全球大模型格局已定，国际与国产并驱。目前已形成较为稳固的竞争格局，但是不同模型在性能与定位上依然差异明显。而根据模型产品用户数量、产生收入以及模型能力等综合因素考量，大致可以分为三个梯队，不同梯队细分领域及生态中都拥有属于自身的核心竞争优势。

表：全球通用大模型格局

第一梯队：旗舰级模型（核心业务首选）	第二梯队：性价比模型（部署成本低）	第三梯队：轻量模型（处理日常任务）
Claude Opus 4.6 ：编程与复杂推理能力强大，最高分，适合专业开发者与企业代码审查。	Kimi K2.5 ：国产编程最强，万亿 MoE 架构，完全开源，编程辅助性价比无敌。	GPT-5.4 Mini/Nano ：生态兼容最好，适合文本分类与日常摘要。
GPT-5.4 ：最均衡的通用模型，知识储备全面，工具调用与 Agent 开发能力最深，适合普通用户日常使用及创意工作者。	GLM-5.1 ：全能均衡，订阅费用低，最适合个人用户日常使用。	Gemini 3.1 Flash Lite ：支持超长上下文且价格极低，长文档摘要首选。
Gemini 3.1 Pro ：多模态与超长文档处理能力强大，支持 100 万 token 上下文，原生支持文本、图片、音频、视频全理解。	MiniMax M2.7 ：响应速度最快，适合实时对话场景。	
DeepSeek V4 ：性价比较强，中文理解多项基准超越 GPT-5.4，数学与逻辑推理极强，且完全开源可本地部署。		

国内大模型能力已逐渐追上国际水平

目前国内大模型相比海外大模型在文本模态和编程能力差距相距不远，但在多模态领域仍有一定差距。国际通用大模型在通用推理与多模态能力上仍具优势。当前国产大模型核心优势是在中文理解、性价比及本地化服务上，部分能力甚至达到国际领先水平。

表：大模型文本模态能力排名

发布公司	大模型名称	Arena 排行榜
谷歌	gemini-2.5-pro	1452
Anthropic	claude-sonnet-4-5-20250929-thinking-32k	1448
Anthropic	claude-opus-4-1-20250805-thinking-16k	1448
OpenAI	gpt-4.5-preview-2025-02-27	1442
Anthropic	claude-opus-4-1-20250805	1439
OpenAI	chatgpt-4o-latest-20250326	1438
OpenAI	gpt-5-high	1437
Anthropic	claude-sonnet-4-5-20250929	1436
OpenAI	03-2025-04-16	1433
阿里巴巴	qwen3-max-preview	1433
百度	ernie-5.0-preview-1022	1432
智谱AI	glm-4.6	1429
OpenAI	gpt-5-chat	1425
阿里巴巴	qwen3-max-2025-09-23	1424
Anthropic	claude-opus-4-20250514-thinking-16k	1423
DeepSeek	deepseek-v3.2-exp-thinking	1421
阿里巴巴	qwen3-235b-a22b-instruct-2507	1420
XAI	grok-4-fast	1420
DeepSeek	deepseek-r1-0528	1417

表：大模型编程能力排名

发布公司	大模型名称	Arena 排行榜
OpenAI	GPT-5(high)	1473
Anthropic	Claude Opus 4.1 thinking-16k(20250805)	1458
Anthropic	Claude Opus 4.1(20250805)	1451
Anthropic	Claude Sonnet 4.5 (thinking32k)	1420
MiniMax	MiniMax-M2	1405
谷歌	Gemini-2.5-Pro	1399
智谱AI	GLM-4.6	1395
DeepSeek	DeepSeek-R1-0528	1393
Anthropic	Claude Sonnet 4.5	1387
Anthropic	Claude Opus 4(20250514)	1383
智谱AI	GLM-4.5	1379
智谱AI	GLM-4.5-Air	1366
阿里巴巴	Qwen3-Coder	1365
Anthropic	Claude Sonnet 4(20250514)	1362
DeepSeek	DeepSeek-V3.1-thinking	1360
Anthropic	Claude 3.7 Sonnet(20250219)	1358
Anthropic	Claude Haiku 4.5(20251001)	1354
阿里巴巴	Qwen3-235B-A22B-Instruct-2507	1352
DeepSeek	DeepSeek-V3.1	1338
阿里巴巴	qwen3-coder-plus-2025-09-23	1334

表：多模态-视觉模态大模型能力排名

发布公司	大模型名称	Arena 排行榜
谷歌	gemini-2.5-pro	1249
OpenAI	chatgpt-4o-latest-20250326	1240
OpenAI	gpt-4.5-preview-2025-02-27	1228
谷歌	gemini-2.5-flash-preview-09-2025	1224
OpenAI	gpt-5-chat	1222
OpenAI	03-2025-04-16	1221
OpenAI	gpt-4.1-2025-04-14	1216
谷歌	gemini-2.5-flash	1215
Anthropic	claude-opus-4-20250514-thinking-16k	1211
OpenAI	gpt-5-high	1209
Anthropic	claude-sonnet-4-20250514-thinking-32k	1206
阿里巴巴	qwen3-W-235b-a22b-instruct	1204
OpenAI	04-mini-2025-04-16	1203
OpenAI	gpt-4.1-mini-2025-04-14	1203
Anthropic	claude-3-7-sonnet-20250219-thinking-32k	1196
OpenAI	01-2024-12-17	1195

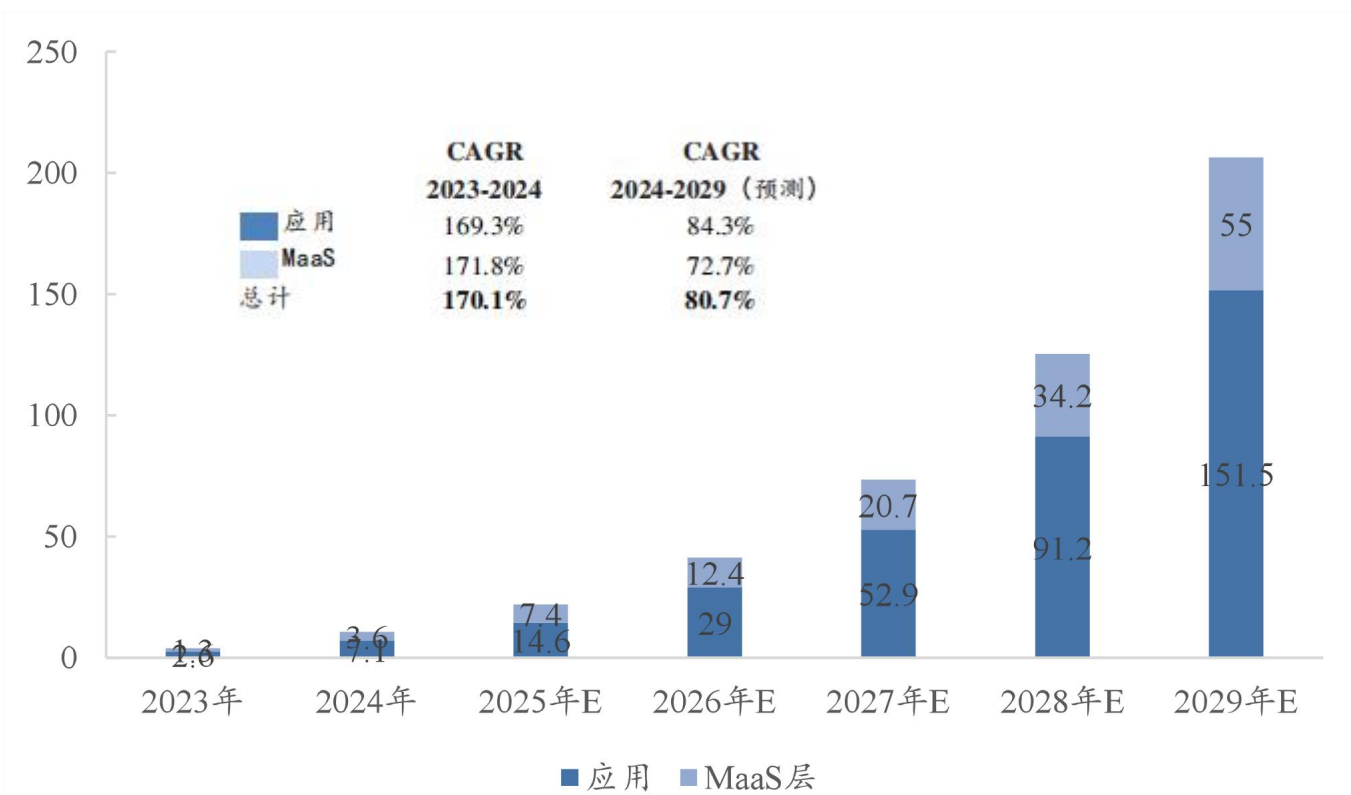
大模型市场规模快速攀升

AI大模型市场正以高达70%-80%的复合年增长率扩张。在技术端模型参数规模呈指数级增长，并叠加多模态融合技术，大幅拓宽了AI的能力边界。在应用端，千行百业的AI Agent持续落地，国内以通义、豆包等为代表的大模型日均调用量快速攀升。放眼全球，所有科技巨头资本开支维持在历史高位，充裕的资金供给为行业高速扩容奠定了基础。

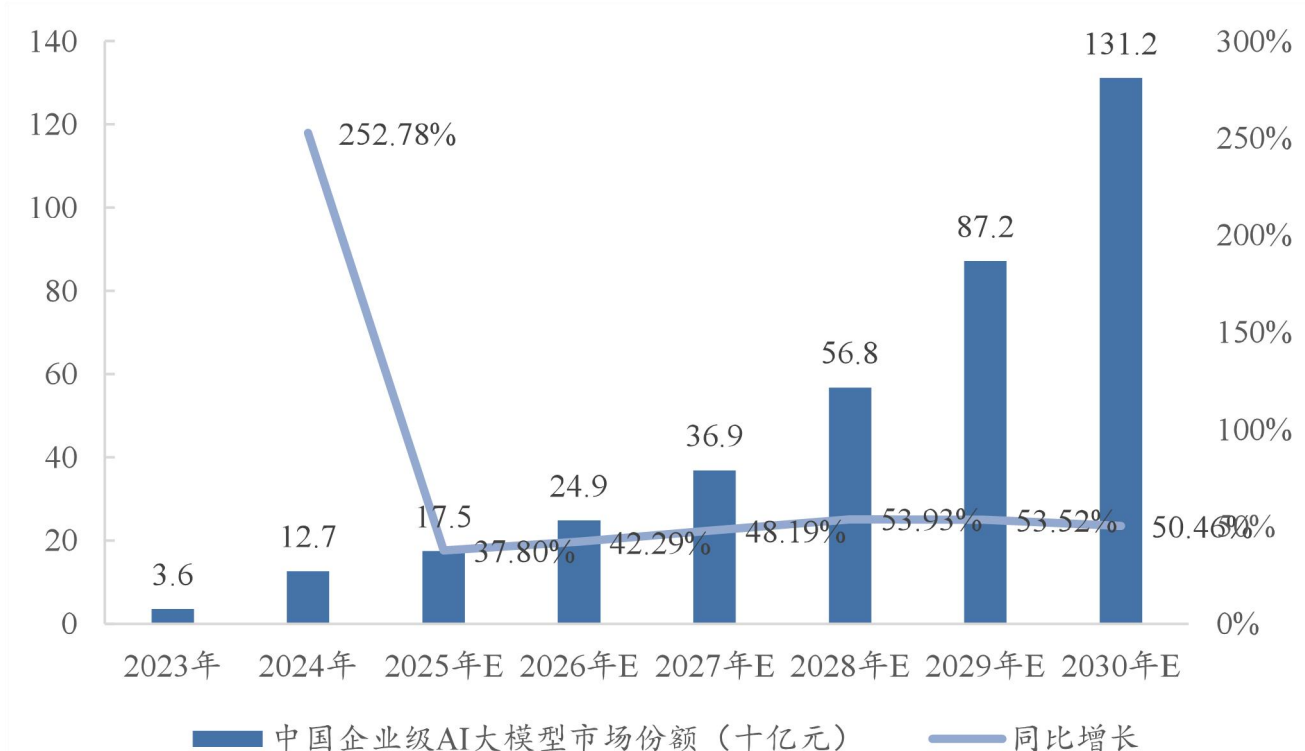
核心驱动因素

- **AI能力普遍化：**AI能力平台化极大地降低了企业使用先进AI能力的门槛。企业无需投入巨资进行模型自研，通过API灵活调用所需功能，这大大拓宽了市场对大模型的需求。
- **模型能力的持续增强：**头部模型厂商通过不断迭代，提供性能更强、功能更多样（如多模态、长文本）的大模型，持续吸引开发者和企业客户付费使用。
- **开源与闭源的共生发展：**开源模型的问世促进了大模型技术普及和创新，同时闭源模型则在性能上持续突破，两者共同做大了整个模型层市场。

图：全球大模型市场规模（十亿美元）



图：中国企业级大模型市场的规模（十亿元）



大模型的通用性带来广阔潜在市场空间

AI在释放社会生产力的同时，也丰富了创造力。如今AI已渗透到企业和个人生活的各个层面，从社交媒体内容推荐系统、chatbot、智能个人助理，到自动驾驶系统、智能风控模型、AI辅助诊断等，AI已成为推动社会与产业智能化转型的核心力量。和前几代AI相比，过去三年的大模型是重要的技术范式变化，也是社会发展的必然需要。传统AI基于小模型针对不同需求场景进行定制化训练，目标是达到能够执行人类所有思维任务的智能水平。而过去三年，大模型领域取得了飞跃式的发展，模型规模和智能水平都获得了大幅提升，叠加多模态能力拓展以及商业价值释放，行业呈现出高速演进的态势。目前大模型应用的主要赛道包括生产力、娱乐、视觉生成、音频生成和通用2B服务。由于大模型技术的通用性特征，这些市场都实现了用一套高度可扩展的模型，来撬动下游规模化与个性化并存多元需求场景，从大型行业客户到中小创作者实现了全覆盖，业务ROI实现转正。

图：大模型应用市场格局



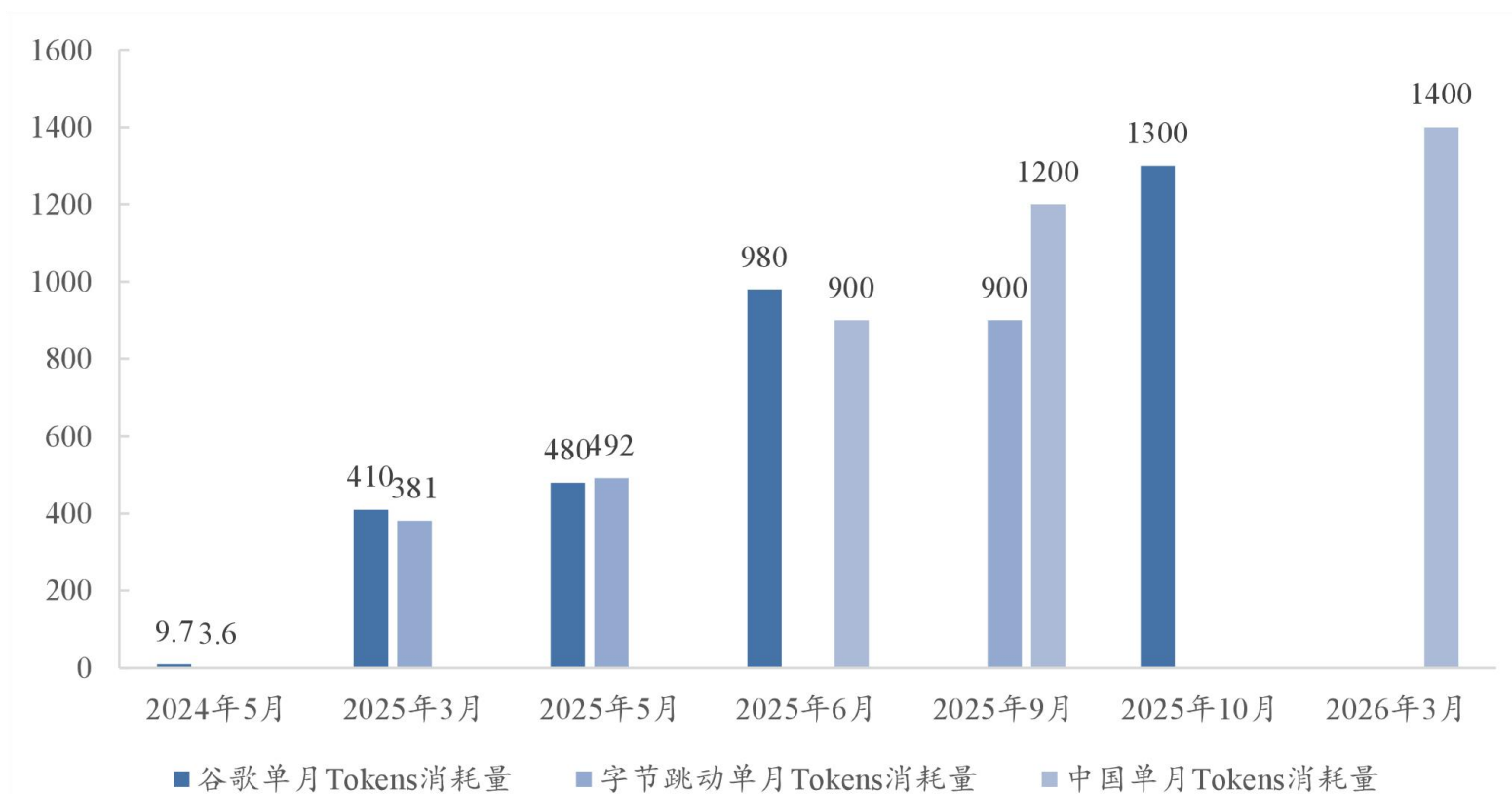
Token消耗将持续爆发增长

过去三年中国日均Token调用量在两年内从1000亿提升到140万亿以上，增长超过千倍。Tokens消耗增量主要来自四大场景：AI搜索、AI视频、AI应用、大模型API调用。其中AI搜索仍然为核心Tokens消耗场景，例如Gemini大模型API在25年10月达到每分钟70亿Tokens消耗（月消耗约300万亿）。

中国国家数据局披露2025年9月底我国日均Tokens消耗达40万亿、字节跳动日均消耗超30万亿。预计中国企业活跃智能体数量将在2031年突破3.5亿，年复合增长率超过135%。随着智能体任务密度和复杂度提升，智能体Token消耗年均增幅有望超过30倍。

Token调用未来将持续实现爆发式增长。随着大模型、AIAgent、多模态应用加速落地并走向规模化商用，各类智能应用的交互频次、任务复杂度、上下文长度持续攀升，叠加企业端批量部署数字员工、C端AI工具渗透率快速提升，模型调用量迎来指数级扩容。Token作为AI算力与交互的基础计量单位，其需求将同步大幅走高。同时行业级应用走向常态化，长文本处理与多轮复杂任务成为主流，进一步推高单场景Token消耗。在当前供需格局下，Token赛道已经进入量价齐升的高增长周期。

图：全球Tokens消耗量呈现爆发态势（万亿Tokens）



Token中间层商业化模式开始出现

AI应用的商业化已经延伸到运营Token调用的能力。Token作为大模型处理信息的最小信息单元，也是模型API计费、结算和消耗的基础。随着调用量放大，Token本身开始像一种“库存”被采购、路由、拆分、转售。目前Token运营正在形成一个新的中间层市场，但不能简单理解成倒卖API额度。这类企业利润来自转售利差的占比较低，但通过推理加速、统一接口、企业端Prompt工程、Agent编排、模型选型和业务系统集成等能力，能够让利润增厚。

图：Tokens运营产业链图解



Token分销当前有三种盈利方式。

第一种是转售利差。平台向上游模型厂商批量采购API额度，再向下游客户加价销售。例如在供应商成本上加收约5.5%的溢价。

第二种是技术溢价。平台通过自研推理加速引擎降低单Token运行成本，在售价接近甚至低于官方价格时，依靠算力效率差获取毛利。例如将语言模型推理速度提升10倍，文生图效率提高3倍，使大模型API调用成本降低。

第三种是企业增值服务。企业部署AI的成本并不只在Token单价，还包括Prompt工程、多模型选型、业务系统集成、工作流编排、运维调度和员工AI能力建设。

大模型基础架构企业增长空间测算

目前智谱AI和Minimax代表了中国大模型创业公司的两条截然不同的商业化路径。

智谱：聚焦B/G端，MaaS与私有化部署双轮驱动

智谱商业模式主要面向企业和政府客户，业务稳定性较高。

- **MaaS API平台服务**：通过开放平台提供模型API调用服务，这是其实现规模化、标准化收入的核心。2025年，其API服务收入从0.48亿元大幅增至1.9亿元，并且其MaaSAPI平台的年度经常性收入实现了60倍的增长，达到17亿元（约2.5亿美元）。
- **本地化部署**：主要为大型企业或政府机构提供定制化的模型解决方案和私有化部署服务。这是公司主要来源，贡献了超过80%收入。该类型客户客单价高，但业务更偏向于项目制，因此增长的可扩展性相对较弱。

Minimax：C端应用引领，开放平台协同，拓展全球市场

Minimax的策略则更加多元化，呈现出C端应用+B端平台+全球化的立体打法。

- **AI原生产品（C端）**：这是Minimax的特色和核心差异点，旗下拥有AI社交、创作工具等面向消费者的产品矩阵。其收入主要来自用户订阅和应用内购买。
- **开放平台与企业服务（B端）**：与智谱类似，提供API服务。目前开放平台收入增长迅猛，2025年同比增长197.8%，API接口收入在Q4同比增长277.6%。该部分业务毛利率高达69%，是重要的利润来源。
- **全球化战略**：Minimax从创立之初就具备全球化视野，其收入结构也证明了这一点。2025年，其国际市场收入占比高达73%显示出强大的海外市场拓展能力。

表：大模型企业收入结构拆解

收入结构拆解	
业务板块	测算关键指标
C端订阅	用户数 × ARPPU × 付费率
B端企业	客户数 × 客单价 × 续费率
API调用	Token消耗量 × 单价
云服务/部署	服务器规模 × 单位收入

大模型基础架构企业增长空间测算

Minimax2026年收入测算：

2025年总收入7904万美元，C端占40%（约3162万美元），B端占60%（约4742万美元）。

C端收入预测：

付费用户数增长：假设在产品创新和市场推广下，付费用户数（65.03万名）增长150%。

ARPPU增长：假设ARPU提价和增值服务推动ARPPU增长30%。

• 2026年C端收入：3162万美元×(1+150%)×(1+30%)=1.02亿美元

B端收入预测：

付费客户数增长：假设开发者生态爆发，客户数增长80%。

调用量增长：假设存量客户调用量大增，ARPC增长80%。

API价格增长：行业持续涨价30%

• 2026年B端收入：4742万美元×(1+80%)×(1+80%)×(1+30%)=2.0亿美元

2026年总收入预测：1.02亿+2.00亿=3.00亿美元

2026年收入增速：(3.00/0.7904)-1=382%

表：Minimax收入及利润一致预期表

一致预测

关键指标	2022A	2023A	2024A	2025A	2026E	2027E	2028E
营业总收入(百万) ②	--	3.46	30.52	79.04	226.47	628.76	1,504.80
增长率(%)	--	--	782.17	158.95	186.54	177.63	139.33
归母净利润(百万)	-73.73	-269.25	-465.24	-1,871.62	-665.59	-598.25	-563.46
增长率(%)	--	-265.19	-72.79	-302.29	64.44	10.12	5.82
EPS(稀释)	-0.74	-2.56	-4.28	-17.23	-2.12	-1.91	-1.80
ROE(摊薄)(%)	96.54	78.75	58.20	70.68	-82.13	-84.38	-276.81
ROA(%)	-99.67	-135.28	-75.28	-188.13	-42.37	-48.77	-27.27
PE	--	--	--	--	-39.89	-44.38	-47.12
PEG	--	--	--	--	-0.62	-4.39	-8.10

大模型基础架构企业增长空间测算

智谱2026年收入测算：

2025年总收入7.24亿元。MaaSAPI占26%（1.9亿元），私有化部署占74%（5.34亿元）。

MaaSAPI收入预测：

付费客户数增长：假设开发者生态爆发，客户数增长150%。

调用量增长：假设存量客户调用量大增，ARPC增长80%。

API价格增长：行业持续涨价30%

• 2026年B端收入： $1.9\text{亿元} \times (1+150\%) \times (1+80\%) \times (1+30\%) = 11.11\text{亿元}$

私有化部署收入预测：

客户数增长：假设客户数增长40%。

签订合同的平均年化价值增长：假设ACV增长50%。

• 2026年私有化部署收入： $5.34\text{亿元} \times (1+40\%) \times (1+50\%) = 11.21\text{亿元}$

2026年总收入预测： $11.11\text{亿} + 11.21\text{亿} = 22.32\text{亿元}$

2026年收入增速： $(22.32/7.24) - 1 = 208\%$

表：智谱收入及利润一致预期表

一致预测							
关键指标	2022A	2023A	2024A	2025A	2026E	2027E	2028E
营业总收入(百万) ②	57.41	124.54	312.41	724.33	2,542.47	5,971.81	15,597.10
增长率(%)	--	116.93	150.86	131.85	251.01	134.88	161.18
归母净利润(百万)	-103.89	-444.90	-1,394.04	-1,918.48	-4,646.75	-4,369.54	-2,763.69
增长率(%)	--	-449.58	-275.21	-58.91	-142.21	5.97	36.75
EPS(稀释)	-7.80	-29.46	-87.20	-12.03	-10.42	-9.80	-6.20
ROE(摊薄)(%)	79.06	80.19	74.72	58.06	--	--	--
ROA(%)	-38.11	-47.27	-80.70	-100.95	-77.62	-75.56	-21.11
PE	--	--	--	--	-84.57	-89.93	-142.19
PEG	--	--	--	--	0.59	-15.07	-3.87

如何为亏损且收入不高的AI大模型巨兽定价？

传统估值模型失灵

- 市盈率完全不适用。当前与AI相关的公司和业务都处于巨额亏损中，P/E估值法没有意义。如果使用远期市盈率，例如预测公司在2030年实现盈利，并给予一个P/E倍数，但这依赖于长达数年的、极不确定的财务预测，其参考价值非常有限。
- 折现现金流理论可用，但实际操作中困难重重。首先是收入预测的巨大不确定性，AI的商业模式仍在快速演变，未来10年的收入增长曲线极难预测。无法确定是线性增长、指数增长还是S型增长。杀手级应用出现、竞争格局演变等变量都让收入预测失真。同时成本与资本支出的不可预测性，AI公司的主要成本是研发人员薪酬和算力成本（购买GPU），需要持续高研发投入，而GPU的价格和供应受到全球供应链的严重影响。未来的资本支出规模难以精估算。

市销率过于简单

在P/E和DCF失效的背景下，市销率成为了目前市场衡量这类公司最常用的指标。其逻辑在于对早期公司，收入增长比盈利更重要。

- P/S的优点：简单直观，可用于亏损公司，能够反映市场对公司未来增长潜力的预期。
- P/S的缺点与风险：忽略了盈利能力和成本结构：高收入不等于高质量的收入。一家毛利率80%的公司和一家毛利率20%的公司，即便P/S相同，其内在价值也天差地别。对于AI公司，高昂的推理成本可能长期侵蚀毛利，这是P/S无法反映的。同时AI相关业务的P/S倍数在当前市场中，缺乏可供参考的历史基准。

当市场情绪狂热时，投资者愿意为规模增长给予更高倍数，P/S倍数可以脱离基本面支撑，但会积累回调风险。目前部分AI技术上市公司的远期市销率（ForwardP/S）已经高达500倍。

如何为亏损且收入不高的AI大模型巨兽定价？

考虑到以上因素，当前市场对大模型公司定价高度依赖对远期ARR的预期，并普遍采用基于ARR的倍数法（P/ARR或EV/ARR）。

- OpenAI公司ARR已经达到250亿美元。2026年3月，OpenAI完成单轮融资1220亿美元，当前估值达到8520亿美元，P/ARR约34倍。OpenAI的月收入突破20亿美元，其中企业客户贡献的收入占比超过40%。ChatGPT的周活用户突破9亿，订阅用户达到5000万，企业付费客户900万个。OpenAI历史连续三年实现3倍暴涨，2023年底ARR达到20亿美元，2024年底ARR达到60亿美元，2025年底ARR成功突破200亿美元，年增速达233%。
- Anthropic官方披露公司ARR突破300亿美元，目前估值已经突破9000亿美元，P/ARR约30倍。根据Similarweb数据显示，Claude在移动端的日活用户已达1130万。根据《华尔街日报》及行业研究机构SemiAnalysis的最新财务披露，Anthropic的ARR2025年初为10亿美元，2025年底达到90亿美元，年增速达800%。2026年2月达到140亿美元，2026年4月已经突破300亿美元，预计2026年5月冲高至440亿至450亿美元区间。
- MiniMax旗下AI原生产品平均月活跃用户达2760万，拥有约177万名核心付费用户。年度经常性收入(ARR)增长正加速，4月ARR较2月增长超过100%，最新ARR已突破1.5亿美元。管理层目标2026年底实现ARR10亿美元。
- 智谱拥有24.2万名付费开发者，MaaSAPI平台年度经常性收入（ARR）达到17亿元（约2.5亿美金）。管理层业绩会上表示年底目标达到10亿美金。

从ARR角度考量美国头部AI公司股价对应的ARR倍数（P/ARR）大约在30倍至35倍左右。因此根据Minimax以及智谱的2026年预期ARR来计算，市场预期合理估值约在**300亿美元（2350亿港元）至350亿美元（2742亿港元）**。因此我们认为**Minimax**目前二级市场估值相对合理，投资人若需要配置大模型企业，可考虑增配。

一、板块行情回顾

二、大模型赛道资本狂飙，商业化预期高涨

三、智算中心产业爆发式增长，算力租赁迎来风口

四、AI应用故事还未结束，Agent技术重塑业务价值

五、行业展望与战略建议

六、风险提示

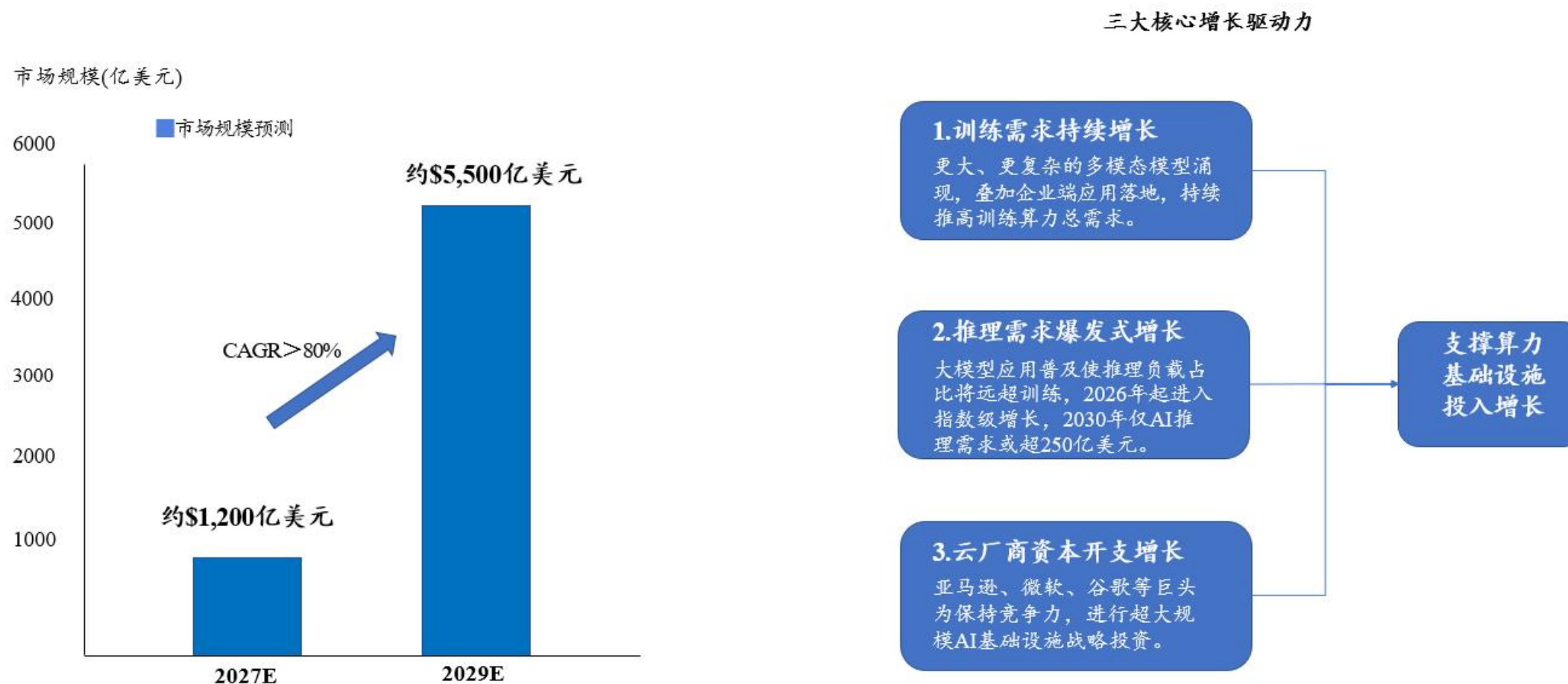
基础设施层：算力刚需奠定底层千亿级基本盘

基础设施层是AI为整个生态提供源动力。其市场规模主要由对算力的需求驱动。根据弗罗斯特沙利文预测，到2027年，全球AI基础设施市场（包括服务器、存储、网络等）规模将达到约1200亿美元。到2029年其规模可能达到5500亿美元。

增长驱动力：

- 训练需求的持续增长：尽管单个模型的训练效率在提升，但更大、更复杂的多模态模型的出现，以及开源社区和各国企业的“百模大战”，仍在不断推高对训练算力的总体需求。
- 推理需求的爆发式增长：随着大模型应用的普及，推理工作负载的占比将远超训练。从技术验证到价值兑现的转变，意味着推理需求将从2026年开始进入指数级增长阶段。据预测，到2030年，仅AI推理市场的需求就可能超过250亿美元。
- 云厂商的资本开支增长：亚马逊、微软、谷歌等云巨头为保持在AI领域的竞争力，正在进行大规模的AI基础设施投资，这构成了市场需求的重要组成部分。

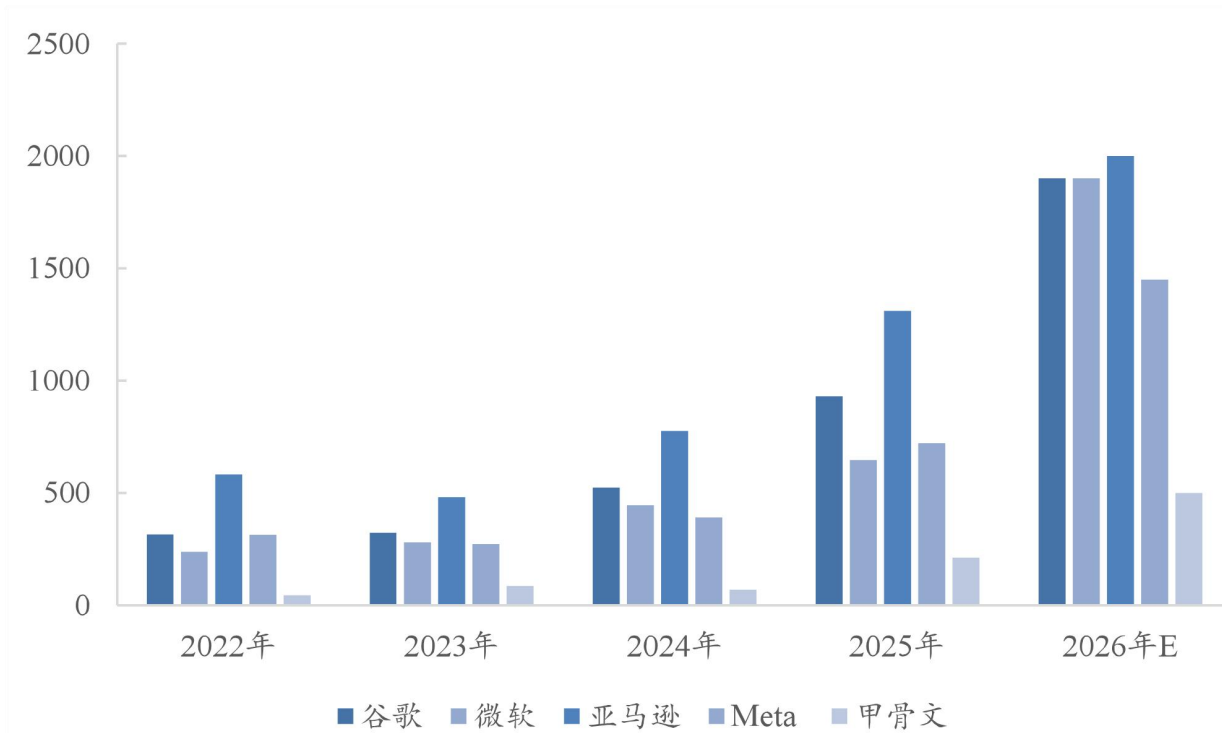
图：全球AI基础设施市场规模预测(2027E-2029E)



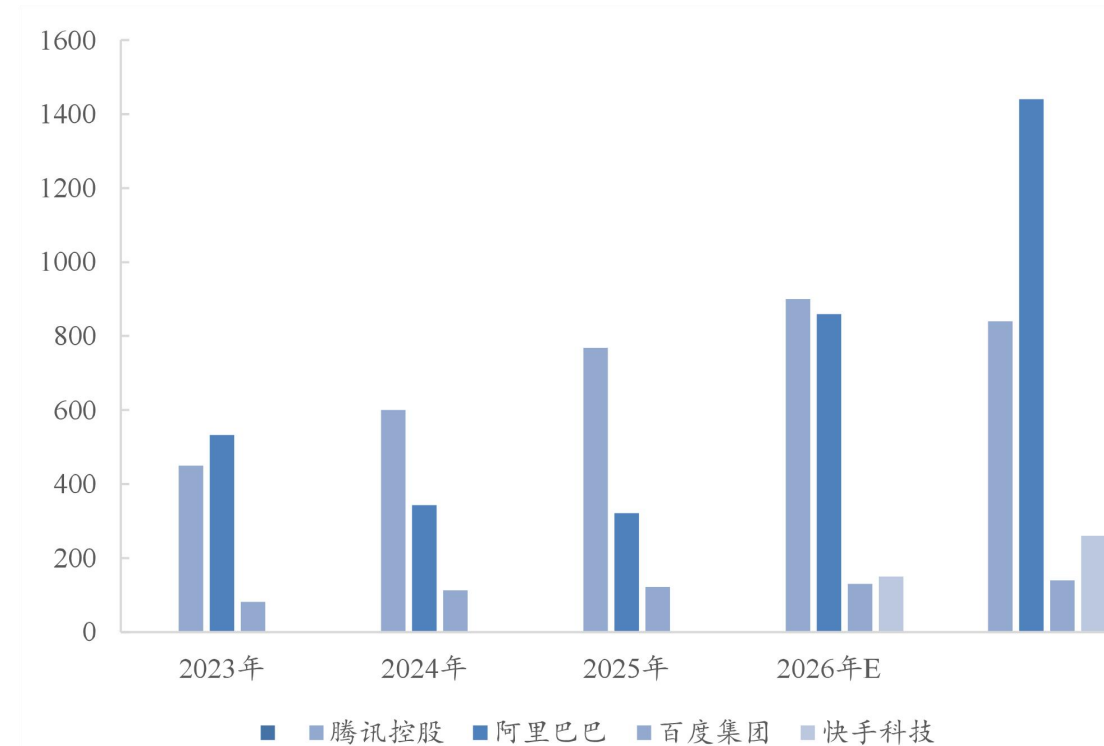
科技大厂Capex逐年攀升

科技大厂的Capex更多体现未来竞争格局。从Capex/经营性现金流来看，目前META、亚马逊、微软在2026已经接近80%。AI算力投资已带来较为显著的折旧摊销成本提升。科技巨头的商业模式逐步向重资产模式转变，从折旧摊销费用/运营总支出提升明显就可以明显看出这一趋势。我们认为未来科技巨头的AI Capex的ROI，将会成为市场关注的核心焦点。

图：科技大厂CAPEX攀升幅度惊人



	2022年	2023年	2024年	2025年	2026年E
谷歌	315	323	525	930	1900
微软	239	281	445	646	1900
亚马逊	583	481	777	1310	2000
Meta	314	273	392	722	1450
甲骨文	45	87	69	212	500



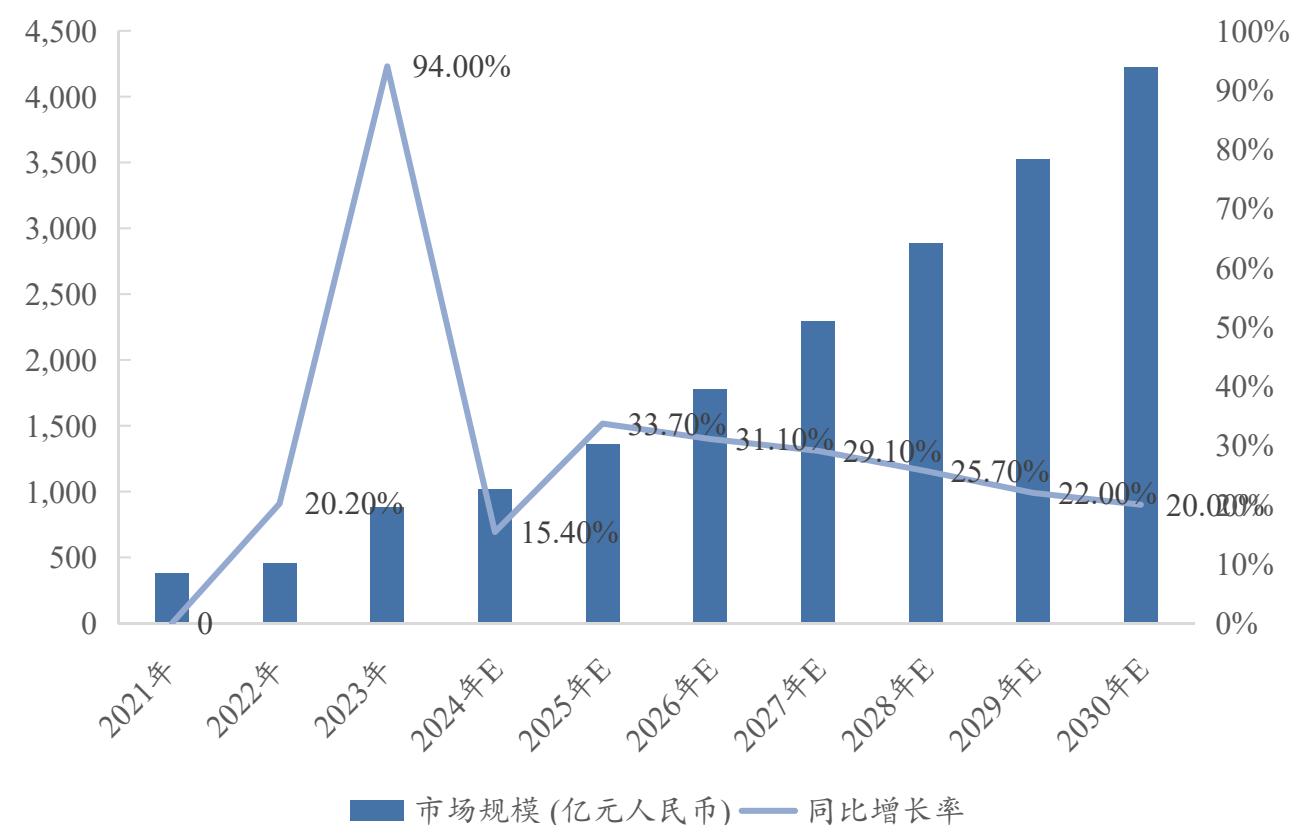
	2022年	2023年	2024年	2025年	2026年E
腾讯控股	450	600	768	900	840
阿里巴巴	533.1	343.3	320.9	859.7	1,440
百度集团	81.5	113	122	130	140
快手科技				150	260

智算中心市场规模迎来新增长周期

当前AI产业高速迭代发展，智算中心作为算力供给的核心载体，是整个AI产业落地与扩张的底层基石。伴随大模型训练、AI应用商业化进程持续提速，行业算力缺口长期存在。因此智算中心的刚需属性不断强化，赛道成长逻辑清晰、景气度上行明确，具备中长期布局价值。根据《中国智算中心产业发展白皮书（2024年）》报告指出，2023年至2028年期间，中国智算中心市场的复合年均增长率预计将达到26.8%。

由于传统数据中心服务主要为计算机系统运行提供特殊的基础设施环境，旨在保障包含服务器、存储和网络设备在内的计算机系统安全、稳定、可靠地运行，可用于存储、处理大量数据。随着AI技术普及，对于数据中心能力提出了更高要求。而智算中心服务相较于传统数据中心服务具有定制化、智能化特点，借助IT技术实现智能化，不再局限于向业务端单向地提供信息和输出数据，而是直接利用整合好的数据进行业务生产，承担核心运营支撑、信息资源服务、核心计算、数据存储和备份等功能，能够根据行业客户差异化需求，提供专业化服务，因此备受市场青睐。

图：中国智算中心市场规模预测



图：当前中国第三方数据中心企业分部



资料来源：中国智算中心产业发展白皮书（2024年）、国元证券经纪（香港）整理

智算中心相关标的推荐

金山云 (KC.O) 已覆盖

公司业务规模增长路径拥有较高可见度，我们认为随着小米金山加大AI投入，基础大模型、智能汽车、WPS均需要大量AI训练和推理算力，非小米客户AI需求亦旺盛，而行业云则稳健增长。金山云已初步打通“AI需求+生态协同”的高质量增长路径。AI拉动与生态协同效应将在未来持续释放。目前行业进入新一轮扩容周期，一方面，年初至今openclaw热度持续、token需求爆发，云商定价权已经开始修复，另一方面，小米大模型迈向万亿参数，生态内订单具备强确定性，此外公司客户不断向自动驾驶、机器人、高端制造及金融科技等多场景拓展。我们看好行业高景气下公司新发展周期，未来上升空间可见度高。

粤港湾智算 (1396.HK) 已覆盖

公司主营业务从传统地产开发已经完成向AI算力生态运营转型。公司2025年收购天顿数据完成业务重组，实现基建与AI业务双轮驱动模式转型。目前天顿数据处于中国智算建设运营商第一梯队，是国内最早布局国标A级第三方数据中心的企业之一，能够提供“绿能智算”全生命周期服务。同时天顿数据作为少数能够搭建万卡集群数据中心的企业，精准聚焦人工智能模型训练与推理核心场景，可直接满足千亿参数级大模型训练、大规模AI推理等高端算力需求，填补了市场对顶尖AI算力服务的供给缺口，并精准聚焦人工智能模型训练与推理场景，占据AI产业链的核心环节。其独特的技术壁垒与业务定位具备稀缺性。

浪潮数字企业 (0596.HK) 已覆盖

公司智算中心业务归属物联网解决方案板块，是公司“AIFirst”战略核心落地载体。在保持原有云SaaS业务持续推进的背景下，聚焦AI算力基础设施建设与运营，面向政府、央国企、大型民企提供全栈智算中心解决方案，在2025年智算中心智算中心业务签单46.9亿元，同比增长72%，占公司总签单超60%，成为公司第一大签单引擎，国内主要以运营商、互联网厂商、金融、能源行业央国企等为主，海外主要以出海互联网厂商、算力租赁公司为主。当前已承建10多个国家绿色智算中心以及承建40多个国家A级数据中心。

一、板块行情回顾

二、大模型赛道资本狂飙，商业化预期高涨

三、智算中心产业爆发式增长，算力租赁迎来风口

四、AI应用故事还未结束，Agent技术重塑业务价值

五、行业展望与战略建议

六、风险提示

AI应用层：垂直深化与智能体崛起是当前技术发展趋势

- 从通用能力到垂直深化。AI的应用正在经历一场从追求“大而全”的通用能力，转向深入解决“小而美”的垂直领域行业痛点的关键转变。虽然通用大模型提供了强大基础能力，但真正的商业价值在于将其与特定行业的知识图谱、业务流程和专有数据深度结合，形成能够即插即用、创造实际效益的行业解决方案。
- 自主智能体的兴起。2026年自主智能体技术成为企业级AI服务领域最前沿、最受关注的发展方向。它不仅仅是执行指令的工具，而是能够理解复杂目标、自主规划、调用工具并执行多步任务的“数字员工”。企业级智能体正在研发、客服、数据分析、市场营销等核心业务场景中展现出巨大潜力，有望实现业务流程的根本性重塑。模型迭代加速与多模态融合。大模型的技术迭代速度依然惊人，AI正从感知智能，向能够理解和预测世界状态的认知智能跨越。同时，文本、图像、语音、视频等多模态信息的融合处理能力日益成熟，使得AI能够更全面地理解和交互，为企业应用创造了更丰富的可能性。
- AI与产业的深度融合。技术不再孤立发展，AI与硬件、软件、数据和行业场景的深度融合成为主流趋势，预示着AI将走出数字世界，通过机器人等载体与物理世界进行交互，为制造业、物流等实体产业的智能化升级开辟了全新路径。

表：核心AI应用赛道及其商业化进展

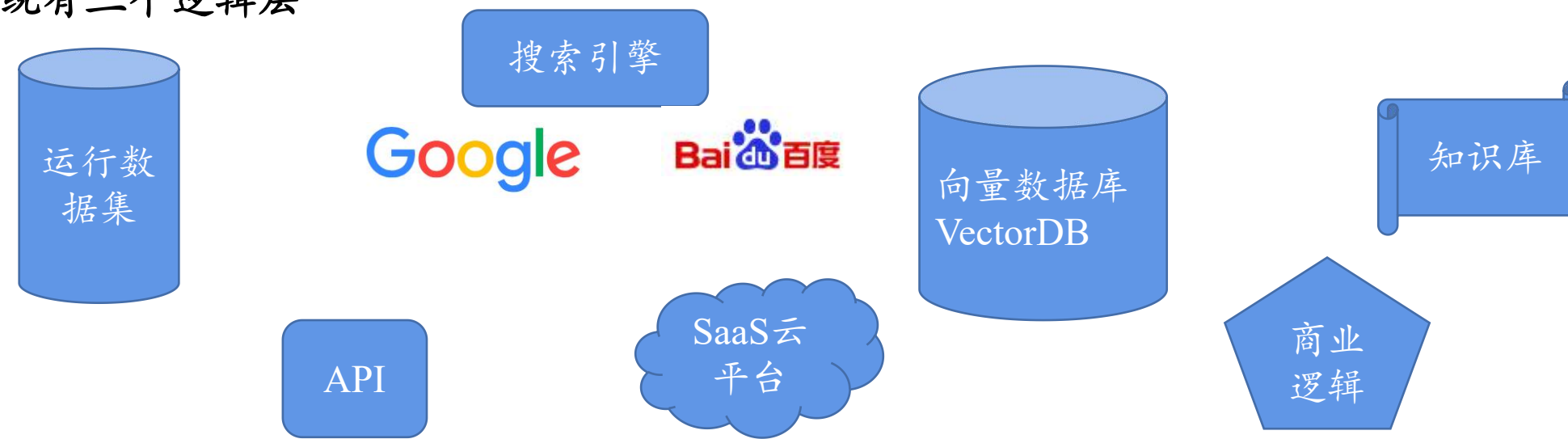
AI应用类型	当前贡献的年化收入	AI应用商业化所处阶段	AI应用大规模商业化方式	代表公司或产品
AI广告	预计数百亿美元量级	成熟稳定提升期	用户使用时长提升、广告推荐效率提升	推荐系统：META、谷歌、字节跳动、快手、腾讯阿里巴巴等 广告投放系统：Applovin、汇量科技、万咖壹联
AI Chatbot	预计百亿美元量级	商业化加速期	SaaS订阅+根据用量增值收费	OpenAI ChatGPT、谷歌Gemini、豆包、千问、猎豹移动
AI编程	预计50-60亿美元	商业化加速期	SaaS订阅+根据用量增值收费	Anthropic、Anyshpere、Cognition、微软
AI视频音乐	预计10亿美元量级	商业化初期	形成AIUGC视频内容社交平台后商业化作为AI视频制作工具SaaS订阅收费	谷歌、OpenAI、快手、Minimax、阅文集团、Synthesia
AI设计	预计数十亿美元	商业化初期	作为AI设计工具SaaS订阅收费	Adobe、Canva、Figma、Midjourney、美图
企业服务/数据治理	预计数十亿美元	商业化初期	SaaS订阅+根据用量增值收费	企业服务：微软、ServiceNow、Salesforce、SAP 数字化底座：Palantir、Datadog、Snowflake、范式智能
行业垂类	预计数十亿美元	商业化初期	SaaS订阅+根据用量增值收费	AI招聘：同道猎聘 AI客服：阿里巴巴、百度、京东、北森控股 AI营销：Clay、水滴公司、百融云、天润云 AI教育：Duolingo、粉笔 AI SaaS：金蝶国际、微盟集团

AI Agent智能系统运作模式分析

Agent智能系统在三个逻辑层面上运作：工具层、推理层和行动层。每一层在使智能体能够有效检索、处理和行动信息方面都扮演着特定角色。理解这些层之间的交互对于设计功能性和可扩展的系统至关重要。

- 1.工具层：**系统的基础，此层与外部数据源和服务交互，包括API、向量数据库、操作数据、知识库和用户交互。它负责获取系统所依赖的原始信息，设计良好的工具确保智能体高效检索相关且高质量的数据。
- 2.行动层：**有时也称作编排层，此层负责协调大型语言模型与工具之间的交互，在适用情况下，它还处理与用户的交互，它接收来自LLM的关于下一步行动的指令，执行该行动，然后将结果提供给推理层的LLM。
- 3.推理层：**系统智能的核心。此层使用大型语言模型处理检索到的信息，它根据上下文、逻辑和预设目标确定智能体接下来需要做什么。推理不充分会导致错误，例如重复查询或行动不一致，使得Tokens使用效率降低。

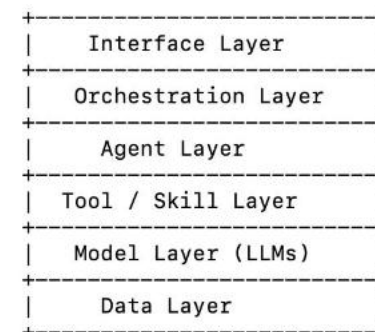
图：AI Agent智能系统有三个逻辑层



编排层Action/OrchestrationLayer

决定 **哪个Agent** 执行什么任务；
 保证任务的 **执行顺序、依赖关系和数据流转**；
 管理任务间的上下文、状态、错误处理；
 支持并发、多步推理和分支逻辑。

典型的多层架构分层图，适用于多Agent系统/AI编排系统



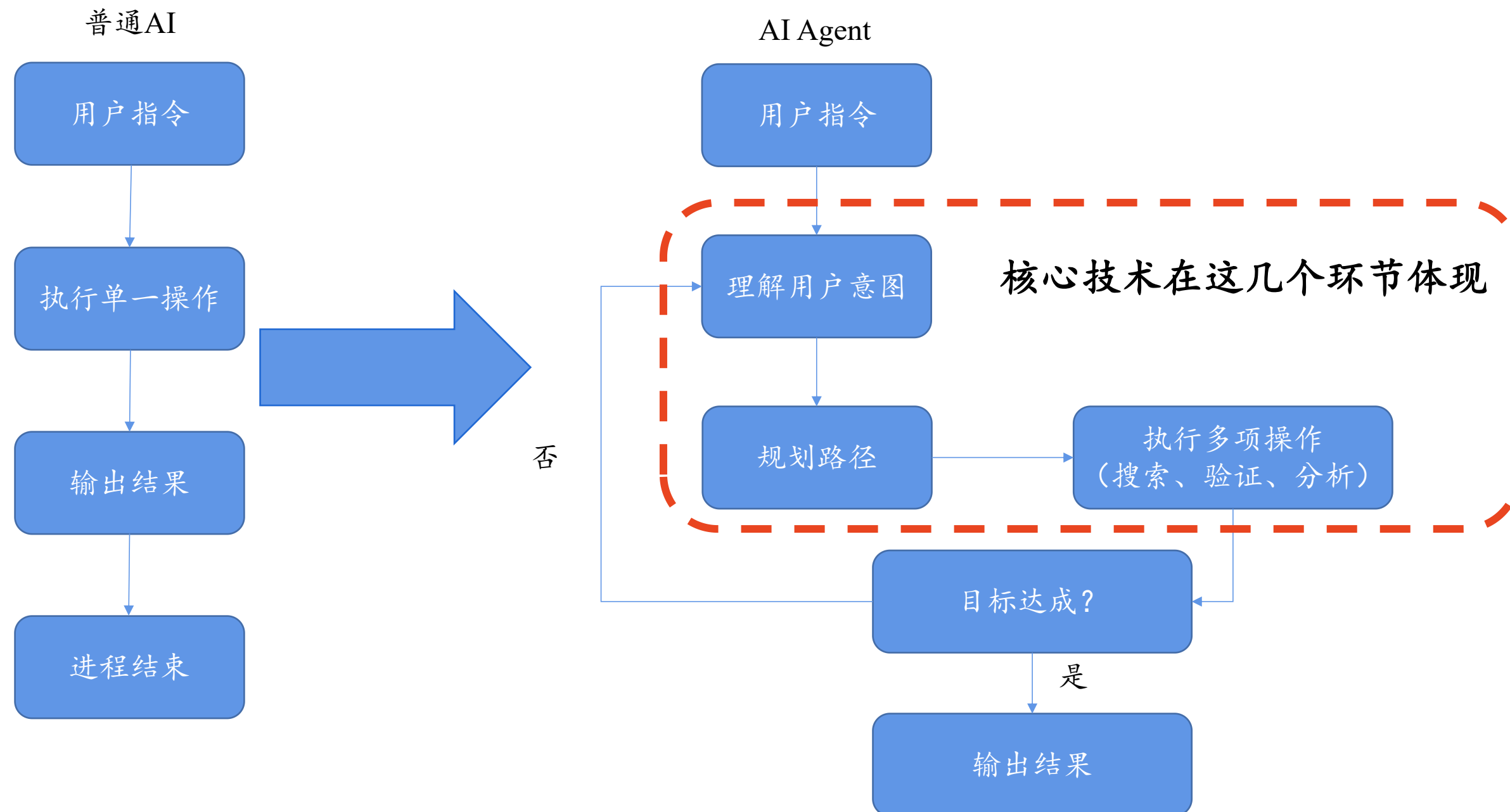
UI/API/输入输出层
 负责协调Agent/Tool/Memory
 功能明确的Agent
 代码执行、搜索、数据库、爬虫等
 GPT/Claude/自训练模型等
 知识库、RAG、文档、图谱



AI Agent是当前可行的商业化应用模式

AI Agent是当前大模型应用落地的核心形态，是能自主感知、决策、执行任务的智能体，具备高落地性的应用模式。相较于通用大模型单纯提供对话能力，AI Agent凭借自主规划、工具调用、流程执行与多轮任务闭环的核心能力，深度契合企业降本增效、个人效率升级的真实诉求。目前行业技术框架逐步成熟，场景适配度持续提升，在办公自动化、客户服务、内容生产、垂直行业运维等领域已跑出可复制的落地案例与盈利模型，商业模式清晰、投入产出可量化。叠加企业数字化转型需求持续释放，AI Agent摆脱了早期AI应用难变现困境，成为现阶段AI产业从技术研发转向规模化商业变现的核心赛道。

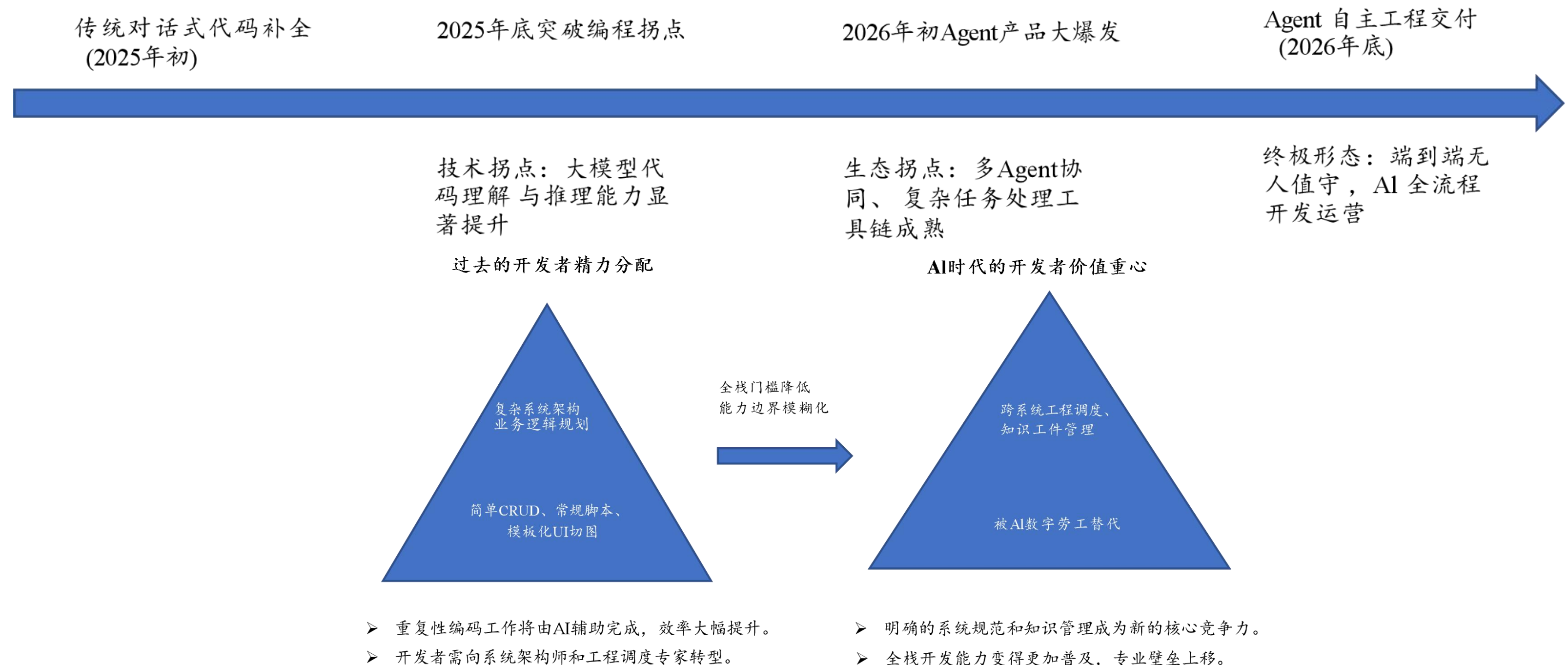
图：AI Agent技术框架已经成熟



AI Agent商业化明显加速，26年有望保持

当前AI Agent赛道商业化进程正呈现明显加速态势。伴随工具功能迭代、开发者渗透率提升以及企业付费意愿增强，行业商业模式已得到一定验证。我们预计2026年赛道整体营收仍将延续高速增长态势，全球及国内市场规模均有望实现大幅扩容。结构上来看，头部产品年化收入持续攀升，B端企业订阅、私有化部署等多元变现路径不断成熟，行业高景气度具备较强持续性。

图：AI Agent技术发展历程



企业级AI Agent发展将迎来加速

相比传统软件落地，部署企业级AI Agent需要一定的前期准备工作：1) 传统软件仅依托于类型较为单一的结构化数据，而AI需要依靠非结构化、多类型数据，其清洗和数据库搭建任务繁杂；2) AI Agent与 workflow 结合、安全合规要求高。我们预计企业级AI Agent平台加速落地时间点或在2026年下半年。目前大模型公司、云计算巨头均在探索企业端AI Agent市场，行业竞争加剧，未来竞争壁垒将来自客户数据的获取能力以及行业Knowhow能力。AI时代研发商需要深度理解企业客户数据，以及对企业客户的行业特性具备较深认知才能构建有效的AI产品。一旦建立合作关系，考虑到迁移成本较高、研发周期长等原因，基本上更换服务合同的情况将不会出现。

表：当前企业级AI Agent平台分布

公司	主要方向	企业级AI领域的关键工作及积累
Palantir	基于企业数据分析平台，理解用户 workflow+数据整形	多年为国防、制造、能源客户进行数据分析、治理、安全。AI平台完成企业客户数据深度整形并实现AI功能
SAP	基于企业ERP平台，理解用户 workflow且财务数据为企业核心	拥有ERP领域多年积累的客户，具备业务理解能力且财务等数据为企业运营中核心数据，拥有较大的AI应用落地潜力 工作：在企业客户原有ERP平台上集成AI功能，嵌入 workflow
Snowflake	基于云原生数据平台，具备数据整形能力	作为云原生数据平台，为企业构建数字化底座，为AI平台落地铺平道路，AI功能套件 Snowflake Cortex、Snowpark系列AI服务
第四范式	企业级人工智能平台与解决方案，覆盖金融、零售、制造、能源等行业。	推出“先知”企业级AI生产服务赋能平台；构建覆盖14个行业的AI大模型；与宁德时代等行业巨头达成战略合作；技术核心为AutoML。
创新奇智	以AI制造为核心企业级AI产品及解决方案，提供全栈式服务	拥有ManuVision、MatrixVision、Orion三大自研AI平台；申请超600项AI专利；深度聚焦制造业智能化转型。
商汤科技	计算机视觉、通用人工智能、智能汽车、智慧城市	推出“日日新 SenseNova”大模型体系；拥有亚洲最大AI智算中心之一；拥有超8000项AI专利及申请，技术实力雄厚。
滴普科技	企业级数据智能与生成式AI应用，提供从数据底座到AI应用的全栈服务	拥有FastData数据融合平台和FastAGI大模型应用平台；打通数据与AI链路；专注于消费零售、制造、医疗等行业。
明略科技	企业级数据智能与自主智能体 (Agentic AI) 软件服务	港股“全球Agentic AI第一股”；提出“可信智能体+可信数据”模式；推出企业级专有大模型DeepMiner；拥有超2300项专利。
智谱AI	以大模型为核心业务，专注于B端企业服务和安全合规	港股“全球大模型第一股”；坚持底座大模型自研，对标OpenAI；深度布局政企服务市场，强调安全可控。
MiniMax	安全可控的通用大模型研发与企业级API服务	对标Anthropic，强调模型安全、长文本处理、强推理能力；核心技术自主可控，服务开发者和企业。
金蝶国际	企业管理软件与云服务，通过AI赋能企业数字化转型	推出“金蝶AI苍穹”企业级AI平台及AI管理助手；将AI深度融合于ERP、财务等核心业务流程，拥有庞大客户基础。
云知声	语音交互、边缘AI芯片及全栈交互式AI解决方案	“云-端-芯”一体化布局；语音识别和边缘AI芯片龙头；推出“山海大模型”并提供MaaS服务；深耕医疗、金融等领域。
迈富时	企业级GEO应用平台，专注于AI Agent营销中台	提供全栈自研GEO技术；AI产品覆盖企业营销全链路；服务超80家世界500强企业；申请超750项AI相关专利。

当前AI Agent已经落地的场景分布

金融行业：金融行业因其高度依赖数据、流程标准化、对效率和风控要求极致的特点，成为AI Agent最理想的应用场景之一，也是最早实现大规模部署和价值验证的行业。

医疗保健：在医学影像辅助诊断、电子病历结构化、药物相互作用检查等场景，AI Agent正在进入医院核心的HIS系统，提升诊断效率和医疗安全。其应用显示出极高的用户复购率，证明了其临床价值。

电商与零售：渗透率同样超过30%，主要应用于智能客服、个性化推荐、动态定价和库存管理，以提升客户体验和销售转化率。

电信行业：在网络优化、故障排查、客户服务自动化等方面，电信运营商是企业级AI Agent采纳率最高的行业之一。

软件开发：从代码生成、Bug修复到自动化测试和项目管理，AI Agent正在成为软件工程师的得力助手，重塑软件开发生命周期。

广告行业：Agent的出现将广告行业从“人工操作+经验驱动”改成数据驱动+全链路自动化+规模化个性化。创意逻辑从“精品少数”到“海量差异化”，全链路效率从“几周”压缩到“几分钟”，AI接管调价、调预算、换素材。

探索AI Agent的“数字劳动力”价值来对AI应用业务进行定价



面对传统估值方法的失灵，资本市场和行业研究者正积极探索能够真正反映AI Agent内在价值的新范式。软件的价值在于其功能，它赋能于人，提升人的效率。而随着技术从“内容生成”向“任务执行”的演进，AI Agent正将软件行业的定位从“工具提供者”转变为“劳动力提供者”。

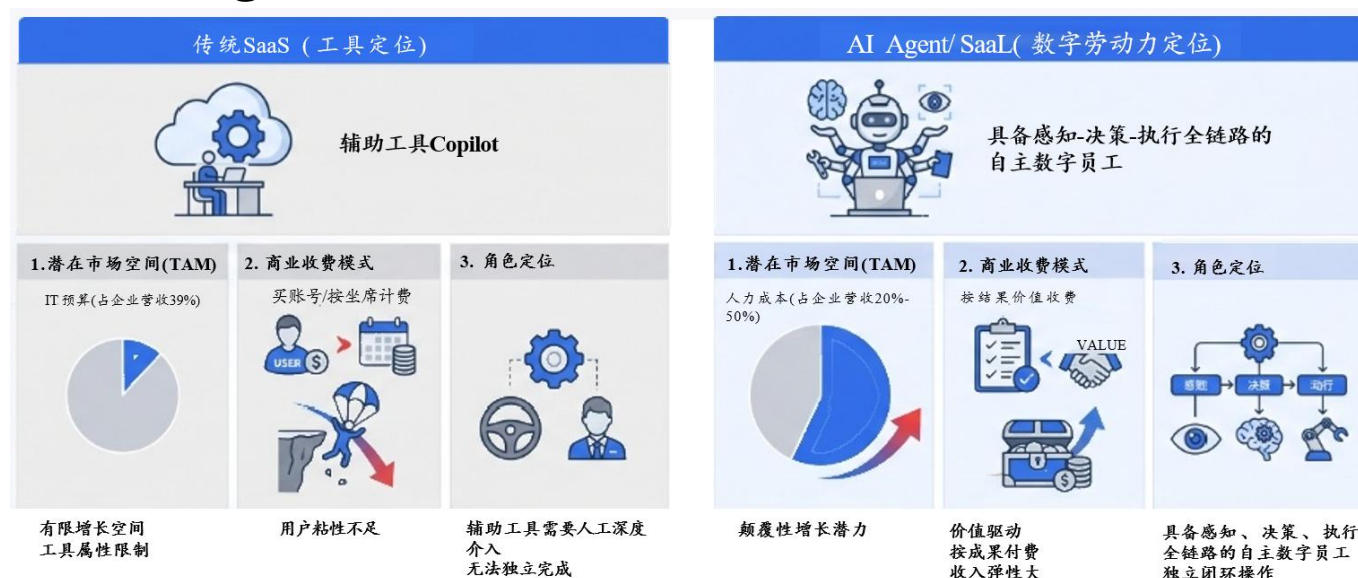
这种从“赋能”到“替代”转变，意味着估值的核心逻辑发生了根本性颠覆。传统的SaaS行业建立在“按订阅收费”的基础上，而随着Agent能力的增强，每增加一个强大的AI功能，就可能减少一个或多个付费坐席需求，将AI Agent定义为“数字劳动力”意味着估值天花板的量级跃迁。

DAU过去一直作为互联网的核心估值公式，底层逻辑是因为人的时间是有限的，因此使用时间是稀缺资源，对于用户使用习惯、使用时长的占用成为了商业化的基础。但由于AI是生产力，因此作为度量生产力的核心指标，二级市场普遍转向TPD估值法——即**每日Token消耗量**。Token消耗正在被越来越多的AI厂商视作反映AI规模、采用度和市场领导力的信号，但这一指标还是不能有效体现业务价值、效率或可持续性。而当人类进入智能体时代，衡量一个平台和生态的繁荣，更应该关注Agent数量和创造价值，

Agent潜在市场空间广阔，传统企业IT预算通常仅占总营收约3%，而企业人力成本往往占据营收的20%至50%。当软件从成本转变为替代真人的投资时，其潜在市场空间被打开，增长空间从百亿美元级跨越至数万亿美元。

商业模式从SaaS向SaaS L跃迁：收费模式从“买账号”彻底重构为“买结果”。AI Agent不再是按席位计费的辅助工具，而是基于完成工作任务的价值需求来计价，根据完成任务量或替代人工价值直接收取费用。

图：AI Agent作为数字劳动力的定位



探索AI Agent的“数字劳动力”价值来对AI应用业务进行定价



随着技术演进，预计到2026年，全球将迈入智能体全面铺开阶段，智能体将具备更稳定的规划、调度与任务执行能力，重塑AI Agent的商业价值。

- 从传统的固定订阅收费转向价值定价。Agent输出企业直接承诺为客户节省费用并将工具产生的收益变成KPI，或者提供闭环成交的销售平台，直接对转化结果负责，甚至可以按创收比例提取“工资”。
- 由于Agent的边际扩张成本极低。传统SaaS市场的复合年增长率已降至约13%-15%，而具备AI能力的AI API和AISaaS市场复合年增长率我们预计将达到35%以上。

因此基于赋予AI Agent任务的估值模型有望成为量化AI Agent商业价值的考量，可以通过计算和预计其自动化任务在全生命周期内创造的净经济价值总和来进行评估。

例如单次任务的价值可以量化为公式：**任务净价值=(替代的人力成本+任务带来的增量收益)-(Agent执行成本+错误成本)**
替代人力成本与增量收益：Agent的价值锚定为“替代人类时薪×年工作时长×3年”，直接将智能体价值与被替代的人力资本对标。此外，7x24小时无休工作带来的额外销售转化。执行成本：系统级智能体在处理长链条、多步骤任务时，需要消耗大量Token，因此算力成本投入ROI需要纳入考量。

根据上市公司百融云资料显示，其“硅基员工”在特定场景下可实现“1个Agent等于10个员工”的效率替代，在理想场景下（如续保提醒、信息核验）可抵消10人班次的产出。当前一名客服一年收入约5万左右，那么一名Agent一年创造的经济价值在考虑成本后，将在30万到40万之间

图：AI Agent作为数字劳动力其价值可以从费用端进行衡量



我们认为值得关注的AI Agent应用企业

公司	覆盖状态	核心AI Agent产品/平台	核心定位	关键架构/功能特性	目标客群与场景
金蝶国际	已覆盖	苍穹AI Agent平台2.0、小K入口、五大核心智能体（如金钥财报）	企业管理AI原生平台	“平台-数据-SaaS-智能体”四层架构；CosmicAI助手具备感知、记忆、思考、行动能力	大、中、小型全客群；财务、人力、制造等全管理流程
第四范式	未覆盖	先知AI平台（Sage）、SageGPT	企业级AI操作系统	“AI Agent+垂直世界模型”；HyperCycle自动化决策流程；多模态交互	金融、能源、制造等14+行业；信贷风控、设备运维、生产排程等复杂决策场景
迈富时	未覆盖	AI-Agentforce智能体中台、Tforce营销大模型	营销销售数字员工孵化调度中心	全链路六层架构；营销、销售、经营决策、搜索四大智能体矩阵；高复用低定制	零售、金融、汽车等多行业企业；营销内容生成、销售线索挖掘、经营分析
中国软件国际	已覆盖	CoMi Agent平台2.0、AI Agentforce智能体中台、垂直场景智能体	全栈全场景AI融合服务商	基础设施+平台+应用三层布局；与ERP/IT服务深度融合；提供“三步九步”AI服务方法论	政企及大型企业客户；公文、会议、合同等办公场景及业务系统智能化
微盟	已覆盖	WAI SaaS、WAI Pro、WIME、星启GEO	智慧商业全场景AI伙伴	从Workflow Agent升级为Skill 调度Agent；覆盖ToB与ToC；深度集成微信等生态	中小商家、品牌零售客户；电商店铺运营、内容创作、导购销售、搜索优化
百融云	已覆盖	百融智能、“结果云”平台	打造覆盖全业务场景的专业“硅基员工”	以“结果云”为核心载体，采用“百基—百工—百汇”三层架构，以RaaS模式替代传统SaaS，按任务、岗位薪酬或价值分成计价	以金融机构为基本盘、以泛金融与高流程化B端为核心扩张层
水滴公司	已覆盖	水守大模型、KEYI.AI、保小慧AI智能客服、Bangbang数字员工	构建“AI原生”保险科技公司	以自研保险垂直大模型为底座，用AI Agent重构保险全链路	大、中、小型保险公司；覆盖保险全链路

一、板块行情回顾

二、大模型赛道资本狂飙，商业化预期高涨

三、智算中心产业爆发式增长，算力租赁迎来风口

四、AI应用故事还未结束，Agent技术重塑业务价值

五、行业展望与战略建议

六、风险提示

行业展望与战略建议：最终还是需要看生态

未来AI的竞争重心将转向系统性的生态构建，头部企业不会是拥有最多算力的公司，而是成功构建了最繁荣、最稳固生态的建设者。生态拥有高效运转的基础设施（AI平台）、源源不断的人口流入（用户）、繁荣的商业活动（应用与服务）、以及鼓励创新的文化氛围（开发者社区），所产生的引力和雪球效应会越来越大，通过网络效应实现了价值的指数级增长，构建商业壁垒，能够为企业带来断崖式的竞争优势和长期稳定的商业回报。

一个成功的AI生态通常由以下几个关键部分构成：

1、技术与数据

数据是AI时代的核心要素，是驱动模型迭代和优化的基础；而AI能力则是将数据转化为智能的引擎。在一个有效的闭环中，技术不仅用于处理数据，其应用本身也会产生新的、更高质量的数据。高质量、全链路的业务数据是AI学习和提升的关键。

2、场景与应用

再强大的AI技术，如果不能与具体的业务场景深度结合就无法产生商业价值。场景和应用是AI能力触达用户、解决实际问题的桥梁。一个成功的生态系统，必然拥有一系列能够为用户提供不可替代价值的核心应用。

3、用户与流量

用户是所有商业价值的最终创造者和支付者。庞大且活跃的用户基础，不仅能为商业变现提供基础（如订阅、增值服务等），同时也是优化AI模型、洞察新需求、孵化新应用的因素，是生态系统保持活力和实现商业化的关键。

4、开发者与合作伙伴

一个开放、协作的平台是构建闭环的重要方式。通过开放API、提供开发工具和激励政策，吸引第三方开发者和合作伙伴加入，能够极大地丰富生态内的应用和服务。这种开放与协作的文化是生态长期健康的基座。

5、商业模式与价值变现

闭环的终极目标，一个无法实现商业变现的闭环是不可持续的。生态系统必须设计出清晰、可行的商业模式，将生态内流动的价值有效转化为企业的收入和利润。这可能包括SaaS订阅、按需付费、收入分成、增值服务等多种形式。

行业展望与战略建议：最终还是需要看生态

面对当前复杂的局面，我们认为需要从以下几点来寻找合适标的：

- **商业化进程，用现金流增长证明：**当前宏大叙事已经讲完，估值已经上升到极高水平，因此现在必须在下半年看到实在的业绩增长才能支撑牛市持续。无论是B端规模化，还是C端盈利，都需要在未来12-24个月内看到明确的、可持续的进展。自由现金流的回正，将是市场延续的最强信号。
- **构建差异化的护城河：**在通用能力趋同的背景下，必须强化自身的“独门绝技”。一方面深化行业解决方案，形成数据和know-how壁垒，另一方面在产品体验和社区生态上持续创新，将用户粘性转化为不可逾越的护城河。
- **精细化成本管控：**在追求技术领先的同时，必须关注投入产出比。优化模型、提升推理效率、探索更经济的算力方案，将是长期发展和技术演化的关键，中期亏损率能否下降影响市场信心。巨额的研发和算力投入虽然必要，但关键在于投入产出比，能否以更低的成本训练出性能更强的模型，以及能否通过技术优化（如模型蒸馏、量化）大幅降低推理成本。
- **需要注意解禁抛压：**由于今年上半年AI企业扎堆上市，下半年面临较大解禁压力。
- **分散投资，组合管理：**考虑到单一公司的巨大不确定性，不应将赌注押在一家公司上。通过投资组合的方式，同时布局AI产业链上的基础设施、模型、应用等不同环节，是更为稳妥的策略。

总体而言，我们认为当前市场存在一定泡沫，同时二级市场也能认识到这一基本情况。但考虑到美国凭借长期积累的科研优势、资本实力与顶尖人才体系率先占据高地，而中国通过庞大的市场规模、工程化能力与应用落地速度方面展现出独特优势，如今终于能够在全球人工智能版图中占据领先地位。从国家层面来看，政府依然会大力支持人工智能产业发展。因此短期板块估值可能会面临一定大幅波动，但AI大模型行业仍处于技术、商业模式和竞争格局剧烈演变的早期阶段，技术还在突破、商业化进程还在持续和宏观政策仍然支持，长期发展依然向好。细分领域将呈现轮番牛市行情，此前是AI应用端，而当前是基础大模型板块，我们预计接下来基础设施板块、AI Agent、AI平台有望出现多轮行情机会，但最终AI生态才是更牢固的护城河，只做单一技术很容易被复制，陷入价格竞争红海，但生态往往难以被超越。参考美股科技巨头企业，基本上都拥有自己的生态。

板块相关标的及估值

细分领域	代码	证券名称	总市值 亿港元	EPS (港元)			市盈率PE			市净率PB 2026TTM	市销率PS 2026TTM	年初至 今涨幅
				2025	2026E	2027E	2025	2026E	2027E			
大模型	2513.HK	智谱	4,583.3	-13.3	-12.0	-11.5	-87.7	-85.8	-89.6	-51.2	571.5	795.0%
	0100.HK	MINIMAX-W	2,692.6	-134.1	-14.8	-14.7	-18.5	-57.8	-58.2	-13.1	437.8	381.2%
	1024.HK	快手-W	1,918.8	4.8	4.1	4.6	9.3	10.9	9.5	2.2	1.2	-25.5%
	0700.HK	腾讯控股	45,190.0	27.4	30.6	34.4	17.8	16.2	14.4	3.5	5.4	-25.9%
	9988.HK	阿里巴巴-W	25,026.9	7.5	5.6	6.4	18.3	23.4	20.5	2.2	2.2	-11.2%
	2506.HK	讯飞医疗科技	83.2	-0.6	-0.1	0.4	-114.2	-590.5	165.0	8.4	8.2	1.2%
	9888.HK	百度集团-SW	3,280.0	1.9	6.7	7.9	54.3	18.0	15.2	1.1	2.3	-5.6%
	0020.HK	商汤-W	810.3	-0.1	0.0	0.0	-41.1	-50.5	-89.1	2.7	14.6	-16.2%
	6651.HK	五一视界	232.7	-0.5	-0.4	0.0	-113.0	-149.9	-3,398.8	29.0	60.4	31.9%
	申请中	中科闻歌	预计400亿									0.0%
AI平台搭建	6682.HK	范式智能	197.5	-0.1	0.5	1.1	-472.1	78.3	35.1	2.4	2.5	-24.0%
	1384.HK	滴普科技	190.9	-3.5	0.0	0.2	-18.5	2,091.7	242.5	20.8	41.6	-9.8%
	2718.HK	明略科技-W	350.6	-152.1	-1.2	0.3	-4.9	-197.0	731.6	13.3	22.2	17.9%
	6636.HK	极视角	135.5	-0.5	0.0	0.0	-267.3	N/A	N/A	56.8	41.9	150.5%
	2158.HK	医渡科技	67.9	-0.1	-0.1	0.0	-46.3	-105.0	637.7	1.6	8.3	3.7%
	9678.HK	云知声	251.0	-5.1	-1.5	1.2	-68.8	-234.1	291.4	32.7	18.7	-31.5%
	0068.HK	群核科技	524.4	-1.1	0.0	0.0	-110.7	N/A	N/A	-11.3	57.7	159.8%
	0596.HK	浪潮数字企业	26.8	0.2	0.6	0.7	10.0	3.8	3.1	0.8	0.3	-65.1%
垂直行业AI解决方案	2723.HK	深演智能	0.0	0.1	0.0	0.0	0.0	N/A	N/A	0.0	0.0	0.0%
	0268.HK	金蝶国际	315.9	0.0	0.1	0.3	386.7	60.7	34.9	3.5	4.1	-41.8%
	1762.HK	万咖壹联	30.6	0.0	0.1	0.1	42.5	23.6	13.5	2.0	0.6	94.2%
	1860.HK	汇量科技	216.1	0.3	0.6	1.0	45.1	21.3	14.1	9.6	1.4	-6.8%
	6608.HK	百融云-W	35.6	0.2	0.5	0.8	43.5	15.0	9.9	0.7	1.1	-50.2%
	2121.HK	创新奇智	22.5	-0.5	-0.3	-0.2	-8.2	-13.2	-20.1	1.7	1.3	-21.7%
	3317.HK	迅策	1,102.3	-0.3	0.0	0.6	-767.9	6,901.2	605.7	42.2	77.5	453.3%
	WDH.N	水滴公司	47.6	0.2	0.2	0.3	7.6	6.2	5.0	0.8	1.1	-16.4%
	0772.HK	阅文集团	267.0	-0.8	1.2	1.4	-31.1	21.6	18.4	1.4	3.3	-24.4%
	1357.HK	美图公司	199.5	0.1	0.3	0.4	32.9	14.5	11.2	3.2	4.7	-36.5%
AI基础设施及算力	3355.HK	飞速创新	189.0	1.9	0.0	0.0	28.4	N/A	N/A	7.9	5.8	4.1%
	3317.HK	迅策	1,102.3	-0.3	0.0	0.6	-767.9	6,901.2	605.7	42.2	77.5	453.3%
	1396.HK	粤港湾控股	108.1	0.1	0.6	0.8	174.1	16.4	12.0	2.5	9.7	236.0%
	KC.O	金山云	407.9	-0.3	-0.2	-0.2	-39.0	-41.1	-49.2	3.9	3.8	40.3%

一、板块行情回顾

二、大模型赛道资本狂飙，商业化预期高涨

三、智算中心产业爆发式增长，算力租赁迎来风口

四、AI应用故事还未结束，Agent技术重塑业务价值

五、行业展望与战略建议

六、风险提示

➤ AI概念股票可能波动较大

当市场调整当中，由于缺乏历史估值对标，AI概念股票波动可能很大。

➤ 业绩低于预期

目前大部分公司AI相关业务规模相对较小，受经济环境影响较大，业绩预测难度较大，经营业绩有低于预期的风险。

➤ 转型失败

部分企业进行AI转型时需重新配置资源，或面临“整合失败”风险。

➤ 行业竞争格局恶化的风险

当前行业处于发展初期，因此利用技术突破来获取用户的竞争较为激烈，需要企业持续的资源投入。

免责声明



一般声明

本报告由国元证券经纪（香港）有限公司（简称“国元证券经纪（香港）”）制作，国元证券经纪（香港）为国元国际控股有限公司的全资子公司。本报告中的信息均来源于我们认为可靠的已公开资料，但国元证券经纪（香港）及其关联机构对这些信息的准确性及完整性不作任何保证。本报告中的信息、意见等均仅供投资者参考之用，不构成对买卖任何证券或其他金融工具的出价或征价或提供任何投资决策建议的服务。该等信息、意见并未考虑到获取本报告人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐或投资操作性建议。投资者应当对本报告中的信息和意见进行独立评估，自主审慎做出决策并自行承担风险。投资者在依据本报告涉及的内容进行任何决策前，应同时考虑各自的投资目的、财务状况和特定需求，并就相关决策咨询专业顾问的意见对依据或者使用本报告所造成的一切后果，国元证券经纪（香港）及其关联人员均不承担任何责任。

本报告署名分析师与本报告中提及公司无财务权益关系。本报告所载的意见、评估及预测仅为本报告出具日的观点和判断。该等意见、评估及预测无需通知即可随时更改。在不同时期，国元证券经纪（香港）可能会发出与本报告所载意见、评估及预测不一致的研究报告。

本报告署名分析师可能会不时与国元证券经纪（香港）的客户、销售交易人员、其他业务人员或在本报告中针对可能对本报告所涉及的标的证券或其他金融工具的市场价格产生短期影响的催化剂或事件进行交易策略的讨论。这种短期影响的分析可能与分析师已发布的关于相关证券或其他金融工具的目标价、评级、估值、预测等观点相反或不一致，相关的交易策略不同于且也不影响分析师关于其所研究标的证券或其他金融工具的基本面评级或评分。

国元证券经纪（香港）的销售人员、交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。国元证券经纪（香港）没有将此意见及建议向报告所有接收者进行更新的义务。国元国际控股有限公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见不一致的投资决策。

除非另行说明，本报告中所引用的关于业绩的数据代表过往表现。过往的业绩表现亦不应作为日后回报的预示。我们不承诺也不保证，任何所预示的回报会得以实现。

分析中所做的预测可能是基于相应的假设。任何假设的变化可能会显著地影响所预测的回报。

本报告提供给某接收人是基于该接收人被认为有能力独立评估投资风险并就投资决策能行使独立判断。投资的独立判断是指，投资决策是投资者自身基于对潜在投资的目标、需求、机会、风险、市场因素及其他投资考虑而独立做出的。

特别声明

在法律许可的情况下，国元证券经纪（香港）可能与本报告中提及公司正在建立或争取建立业务关系或服务关系。因此，投资者应当考虑到国元证券经纪（香港）及其相关人员可能存在影响本报告观点客观性的潜在利益冲突。

本报告的版权仅为国元证券经纪（香港）所有，未经书面许可任何机构和个人不得以任何形式转发、翻版、复制、刊登、发表或引用。

分析员声明

本人具备香港证监会授予的第四类牌照——就证券提供意见。本人以勤勉的职业态度，独立、客观地出具本报告。本报告清晰地反映了本人的研究观点。本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

国元国际控股有限公司

香港中环康乐广场8号交易广场三期17楼

电 话：(852) 3769 6888

传 真：(852) 3769 6999

服务热线：400-888-1313

公司网址：<http://www.gyzq.com.hk>