

评级： 增持（维持）

分析师：康雅雯

执业证书编号：S0740515080001

Email: kangyw@zts.com.cn

分析师：朱骏楠

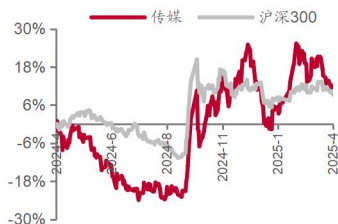
执业证书编号：S0740523080003

Email: zhuqn01@zts.com.cn

基本状况

上市公司数	131
行业总市值(亿元)	15,067.40
行业流通市值(亿元)	13,714.81

行业-市场走势对比



相关报告

- 1、《【中泰传媒】呼市发布育儿补贴细则，把握出版、AI+教育投资机会》2025-03-17
- 2、《GDC2025 召开在即，游戏+AI 迎新进展》2025-03-16
- 3、《【中泰传媒】北京市教委发布人工智能教育工作方案，关注 AI+教育投资机会》2025-03-09

重点公司基本状况

简称	股价 (元)	EPS					PE					评级
		2022A	2023A	2024E	2025E	2026E	2022A	2023A	2024E	2025E	2026E	
南方传媒	16.43	1.06	1.46	0.98	1.07	1.15	15.50	11.25	16.71	15.35	14.35	买入
恺英网络	16.45	0.49	0.70	0.75	0.85	1.00	33.57	23.50	21.99	19.31	16.38	买入
世纪天鸿	10.76	0.17	0.12	0.15	0.18	0.21	63.29	89.67	72.95	60.62	51.16	增持

备注：股价取自 2025 年 4 月 3 日收盘

报告摘要

核心观点：AI Agent 积木搭建完成，迎来爆发临界点

以产业演进角度看，AI Agent 的发展路径类似于搭积木过程，过去各模块（大模型智能性、多模态推理、Coding 能力、工具调用 Tool Use、Token 经济、算力支撑）零散分布，单一技术无法形成闭环，Agent 形态始终停留在“缺乏行动能力”的状态。当前，以 MCP 为代表的工具调用协议标准化落地、OpenAI 的 Agent SDK 框架逐步完善、大模型 Coding 能力提升至产业应用级别、Token 调用成本大幅下降等因素叠加，各项技术模块已逐渐完整拼接。

我们认为，AI Agent 已经走到从组件齐备到整体协同运转的“临界点”，产业生态将步入爆发式增长阶段。预计将有越来越多的工具、软件开发接口，供大模型调用以提升其用户使用体验。且随着 Token 价格的下降，Agent 渗透将加速。以 MCP 聚合平台 Smithery 为例，平台上已经有超过 3500 个 MCP 服务器推出，涉及多个领域。

■ 本篇报告对上述提及的技术演进历程做了详细分析，同时就 AI Agent 可能带来的算力消耗做了敏感性分析测算。

技术演进：

- **大模型**：基础对话→需求对齐→智能涌现→多模态+长时记忆→强 Coding 与推理能力。
- **工具调用**：RAG（检索增强生成，扩展大模型记忆能力）→Function Calling（函数调用，通过强化学习大模型可以调用外部一切）→MCP 协议（统一了各大模型与各工具间的通信格式）→计算机操作（以 OpenAI 的 Operator 与 Anthropic 的 Computer Use 为代表）。
- **技术迭代**：COT 的提出（思维链，让模型具备规划能力）；LangGraph 等的 Agent 开发框架的完善，增强大模型处理复杂问题、多重推理的能力；Cursor（估值 25 亿美元）、Cline 等 AI 代码编辑器的推出，加快 AI Agent 所需的工具模块的开发。

算力消耗测算：

- 以 GPT-4o 为基准，假设一个日活规模达到 10 亿的全球级 AI Agent 应用，每日所需算力约相当于 14.15 万张 NVIDIA H100 SXM GPU。随着未来 Agent 应用向多模态、复杂推理与高频次 Tool Use 方向演进，这一需求预计还将进一步指数级攀升。
- 两个案例：1) Manus 一次标准复杂度任务执行预计消耗 20 万 Tokens；2) Cline + MCP + claude sonnet 3.7 组合一次标准旅行规划消耗 93.34 万 Tokens。

相关公司推荐关注：

- 我们将产业链相关标的分为：**1、硬件与基础设施**（NVIDIA、AMD、Intel、TSMC、中芯国际、Amazon、Microsoft、Alphabet、CoreWeave、阿里巴巴）；**2、模型开发与研究**（Alphabet、Microsoft、Meta、阿里巴巴、腾讯控股、科大讯飞）；**3、平台与工具**（Hugging Face-未上市、Brevian-未上市、Anysphere-未上市、Coze-字节跳动旗下、Scale AI-未上市、Turing-未上市）；**4、应用**（特斯拉、Salesforce、Adobe、Palantir、快手、阿里巴巴、腾讯控股、Boss 直聘、美图、

多领国)。

- **A股传媒相关标的:** 1、AI教育: 南方传媒、豆神教育、荣信文化、世纪天鸿; 2、AI陪伴: 奥飞娱乐、上海电影、汤姆猫; 3、AI工具: 果麦文化、昆仑万维、易点天下; AI娱乐: 恺英网络、巨人网络、盛天网络。
- 风险提示: 1) AI大模型与AI应用发展进度不及预期风险; 2) 技术、环境变化导致测算依据失效的风险; 3) 数据信息更新不及时风险。

内容目录

核心技术跃迁，AI Agent 时代全面到来	4
何为 AI Agent?	4
AI 大模型：对齐→涌现→长时记忆、多模态→推理	4
可调用工具及协议不断完善（RAG、Function calling、MCP）	6
推理赋予 AI Agent 处理复杂问题的能力	9
Coding（代码）就是大模型的自然语言	10
AI Agent 实现路径探寻	12
模型自治 + 外部协同	12
AI Agent 应用算力消耗假设	14
一个全球爆款 AI Agent 应用需要多少张 H100 支持?	14
案例分享：Manus 与 Cline + MCP + claude sonnet 3.7	17
AI Agent 应用场景展望与投资建议	19
现有应用场景分析	19
投资建议	20
风险提示	22
1、AI 大模型发展进度不及预期风险	22
2、AI 应用发展进度不及预期风险	22
3、因技术与环境变化导致测算依据失效的风险	22
4、报告中数据信息更新不及时风险	22

图表目录

图表 1：AI Agent 突破关键技术点	4
图表 2：GPT-4 Tokens 价格变化（美元/百万 Tokens）	5
图表 3：大模型上下文窗口支持 Token 量发展趋势	6
图表 4：RAG 与 LLM 配合使用的概念流程	7
图表 5：Function Calling 实现原理图	8
图表 6：MCP 实现原理	9
图表 7：利用 Grok3 进行 DeeperSearch 处理问题	10
图表 8：LangGraph 与 LangChain 实现路径对比（模拟制定旅行规划）	10
图表 9：Cursor 的订阅价格	12
图表 10：Cursor 年度经常性收入（百万美元）	12
图表 11：Cursor 完成 1 亿美元 ARR 的时间	12
图表 12：AI Agent 实现路径探寻	13
图表 13：H100 SXM GPU 参数一览	14
图表 14：DeepSeek、ChatGPT、Claude 日均交互次数（次）	15
图表 15：Token 消耗量敏感性分析	15
图表 16：代表性超级 APP 的日均交互次数（次）	16
图表 17：每日算力消耗量敏感性分析	16
图表 18：H100 SXM GPU 需求量敏感性分析	17
图表 19：Manus 不同复杂度任务积分消耗	18
图表 20：大模型进行任务规划示意图	19
图表 21：大模型进行任务规划结果	19
图表 22：Manus 的应用分类	20

核心技术跃迁，AI Agent 时代全面到来

何为 AI Agent?

■ 何为 AI Agent?

- OpenAI 官网定义: AI Agent 是可以智能完成任务的系统——从执行简单的工作流程到追求复杂、开放式的目标。
- Anthropic 官网定义: AI Agent 是让 LLM (大语言模型) 动态掌控工具使用方式的系统, 可自主决定如何完成任务。

■ 一句话形象总结: AI Agent 就是给 AI 大模型安上可操控的“四肢”。

初期 AI 大模型只会告诉用户处理问题的方式或解决问题的办法(只有大脑), 但实操环节还需人为介入, 且需要用户来设计解决问题的路径。发展到 AI Agent, 其不仅可以提出处理方案, 还可以自主编排处理流程, 自主调用工具, 最终解决问题。有了 Agent, 将大大扩展个体或机构的能力边界。

■ 技术快速迭代创新, 推动 AI Agent 时代的到来:

- 1、更加智能的大模型, 长上下文记忆, 且调用的 token 成本持续下降。
- 2、大模型可调工具不断丰富。
- 3、辅助框架进一步完善 (RAG 技术、Function calling、LangGraph 框架、MCP 协议、Agent SDK)。

图表 1: AI Agent 突破关键技术点



来源: 中泰证券研究所

AI 大模型: 对齐→涌现→长时记忆、多模态→推理

■ AI 大模型不断智能化

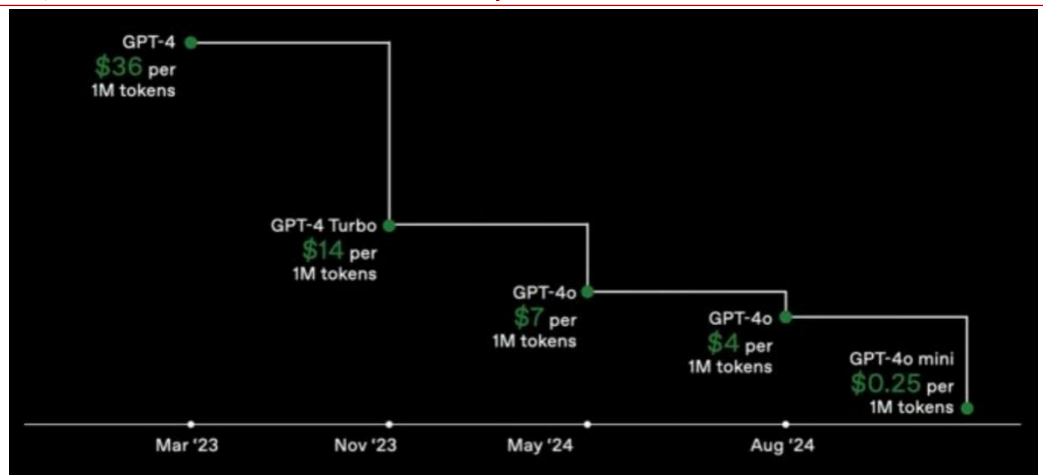
- 2018 年: 初代大语言模型 GPT-1 (固定长度的输入 512 Tokens) 推出, 还无法根据人类指令灵活响应。
- 2019 年: GPT-2 推出, 较好顺写文本, 但在长文本生成时会出现重复、跑题, 缺乏整体连贯性, 无法很好对齐用户意图。
- 2020 年: GPT-3 推出, 拥有 1750 亿参数, 远超 GPT-2 的 15 亿参数, 上下文输入输出窗口来到 2048 个 Tokens。由于参数的大幅提升, 模型开始出现 Few-Shot 学习等涌现能力, 但依旧无法很好对齐用户意图。
- 2021 年: DALL-E 模型发布, 从文本到图像生成显雏形。
- 2022 上半年: InstructGPT 推出, 微调 GPT-3, 采用监督学习和强化学习 (RLHF) 方式, 在仅有 13 亿参数情况下, 更好对齐用户意图。
- 2022 下半年: ChatGPT 正式推出, 至此为分水岭, AI 真正进入了大众视野, 其显著增强了对话的连贯性和上下文理解能力。

- 2023 上半年: Claude 1.0 和 GPT-4 共同推出, GPT-4 推理能力、专业知识进一步提升, 并支持图像输入, 是多模态 AI 的里程碑。
- 2023 下半年: Claude 2.0、GPT-4 Turbo、LLaMA 2、Gemini 推出, 国内阿里推出 Qwen, 大模型进入百花齐放阶段。其中 Claude 2.0 将上下文扩展到 100K tokens, 支持文件上传处理。2023 年 11 月上线 Claude 2.1, 进一步将上下文窗口扩展至 200K tokens, 模型记忆能力大幅提升。LLaMA 2 将其模型开源, 推动整个行业发展。Google Gemini 上线既具备多模态处理能力。
- 2024 上半年: GPT-4o、Claude 3、Gemini 1.5 Pro、Llama 3、DeepSeek-V2 等发布, 多模态进一步完善, GPT-4o 集成视频和语音功能, Llama 3 是首款能够处理图像和文本的开源模型, Gemini 1.5 Pro 支持多模态的同时还将上下文窗口提升至 1M Tokens。
- 2024 下半年~至今: OpenAI o1、Grok 3、Claude sonnet 3.7、DeepSeek-R1、Alibaba QwQ、Gemini 2.5pro 等发布, 标志着模型的推理思考能力大幅提升, OpenAI o1 通过链式思维 (chain-of-thought reasoning) 优化了推理能力, 特别在科学和编程任务上表现优异, 减少模型幻觉问题, Claude sonnet 3.7 在编程和工具调用能力上非常突出。

■ Token 调用成本大幅下降

- 以 GPT-4 系列为例, GPT-4o 每百万 Tokens 调用价格以及相较于 GPT-4 刚刚推出时下降了 89%, 若对比小模型 GPT-4o mini, 则价格比 GPT-4 下降了 99.3%。对比国产大模型 DeepSeek, V3 版本目前国内标准时间段 (北京时间 08:30-00:30) 价格为 2 元/百万 Tokens, 输出为 8 元/百万 Tokens, 较 GPT-4 调用成本更低。

图表 2: GPT-4 Tokens 价格变化 (美元/百万 Tokens)



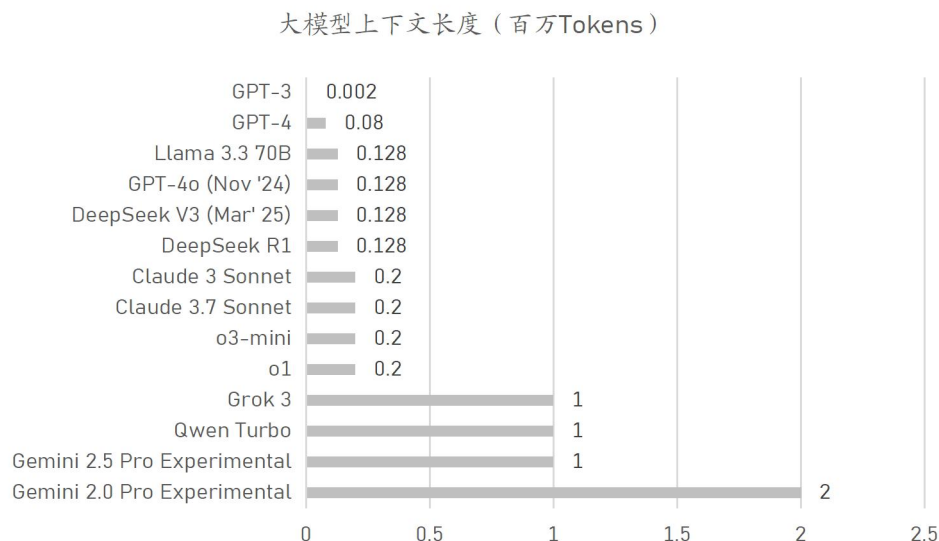
来源: spearheadt, 中泰证券研究所

■ 上下文长时记忆

- 长时记忆对 AI Agent 尤为重要, 减少了 AI Agent 频繁分段记忆导致的结果偏离。不支持足够长的文本输入, AI Agent 就较难建立起对目标任务完整的认知逻辑。例如 GPT-3 只能支持 2048 个 tokens 的输入, 很难完成逻辑链较长的任务, 发展到 GPT-4, 最大已经可以支持到 32K Tokens 上下文窗口, 再到 Gemini 甚至可以支持到 2M 的上下文 Tokens。此前受制于 Transformer 模型的自注意力机制, 计算开销随序列长度

指数增长，目前技术上通过更高效的稀疏注意力和分块处理实现上下文窗口的扩展。

图表 3: 大模型上下文窗口支持 Token 量发展趋势

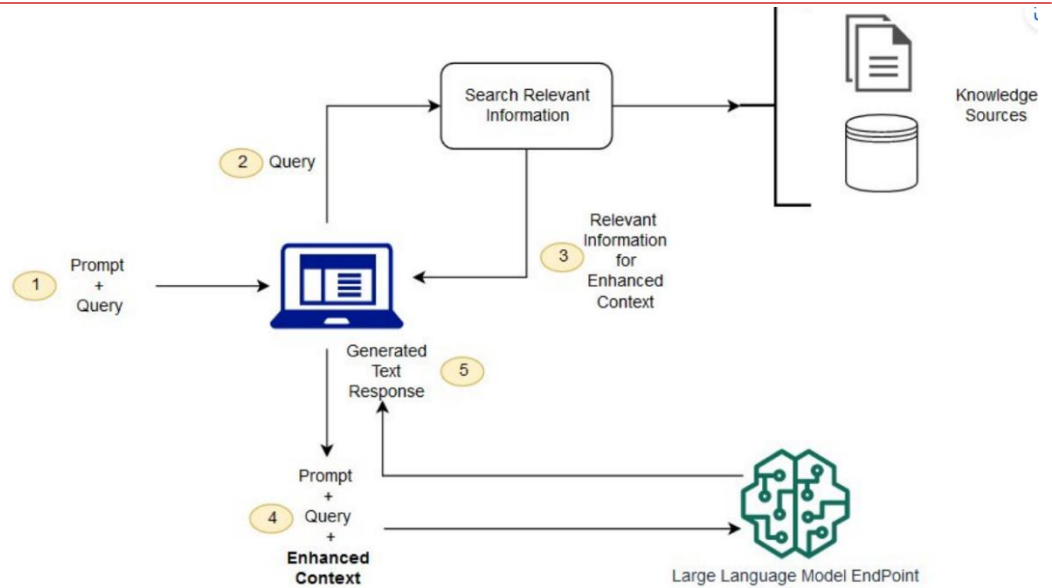


来源: artificialanalysis, 中泰证券研究所

可调用工具及协议不断完善 (RAG、Function calling、MCP)

- 大语言模型经历了从预训练回答历史知识，到实施调用工具回答实施问题的阶段，工具调用与协议的完善大幅拓展了大语言模型的能力。
- **RAG (检索增强生成) 技术的成熟:** 使 AI 大模型能够在生成响应之前引用训练数据来源之外的知识库，对模型在专业领域的知识补充、长时记忆、新知识补充等都起到至关重要的作用。提升了大模型输出的信息的及时性和质量、减少幻觉。在大模型本就已经很强的上下文记忆能力基础上，接上了一个随时可以扩展的向量数据库。

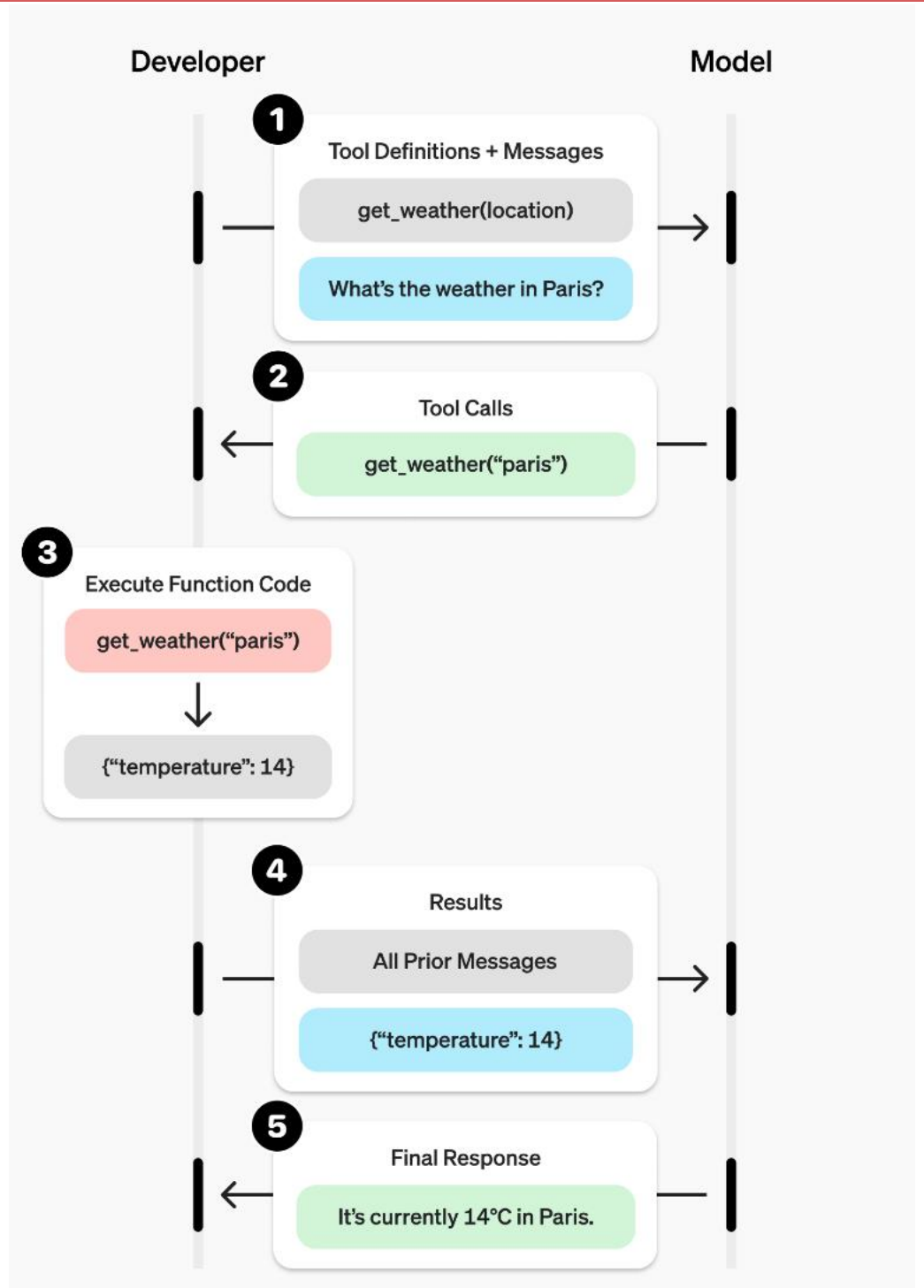
图表 4: RAG 与 LLM 配合使用的概念流程



来源: Amazon 官网, 中泰证券研究所

- **OpenAI 推出 Function calling 功能:** 2023 年 6 月, OpenAI 推出 Function calling 功能, 允许开发者在 Chat Completions API 中定义函数, 大模型可根据用户输入自动调用相应函数, 生成所需的 JSON 对象输出。结合了 Function calling 后, 大模型相当于拥有了可以调用函数的能力, 进而获得了与外部交互的能力。

图表 5: Function Calling 实现原理图

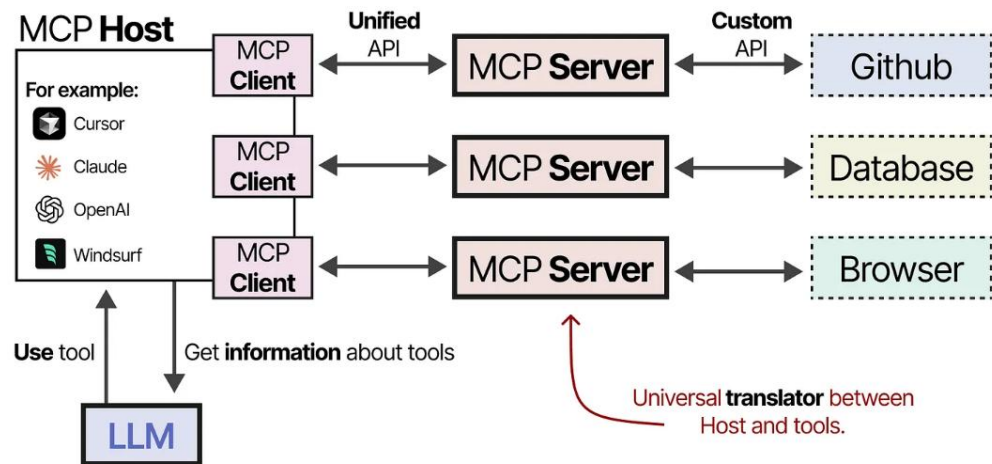


来源: OpenAI 官网, 中泰证券研究所

- **Anthropic 推出 MCP (Model Context Protocol) 协议:** 2024 年 11 月 25 日 Anthropic 正式推出开源 MCP 协议, 相当于在 Function Calling 之上提供了一套标准协议, 把 API 调用流程进一步抽象, 让 AI Agent 可以自动调用多个工具。Anthropic 已经预先预构建了一些 MCP 服务器, 支持连接 Google Drive、Slack、GitHub、Git、Postgres、Puppeteer 等企业级系统, 也可以由社区编写更多功能的 MCP 服务器。MCP 简化了 AI 连接外部数据的难度, 一次性构建兼容多个数据源的 AI 解决方案,

提高数据访问效率。有了 MCP，就可以将企业数据与 AI 大模型更好的链接起来，让大模型更高效更自主使用外部数据。

图表 6: MCP 实现原理



来源: maartengrootendorst, 中泰证券研究所

- **其他工具的完善。**包括: 1、搜索 (Google Search API 等), 让大模型更加方面从互联网获取信息; 2、文件系统 (File search), 让大模型可以调用本地或者云端文件辅助分析; 3、计算机调用 (Computer Use), 让大模型具备类人的计算机操作能力; 4、函数调用, 让大模型可以使用编写的函数完成自定义功能; 5、专业 API 调用, 例如金融系统、天气系统、电商系统等数据信息; 6、自动化工具, 例如 Google 邮箱、Discord 平台、Slack 平台等信息的同步与编写。

推理赋予 AI Agent 处理复杂问题的能力

- 推理与制定计划 (Planning) 能力是 AI Agent 能够完成复杂目标的核心: 将最终目标分解为可执行的具体步骤。大模型推理能力发展脉络:
 - 1、初代大模型只是一问一答, 没有推理和计划能力;
 - 2、COT (Chain of Thought, 思维链) 在 2022 年 1 月被提出, 一种专门用于提升大模型推理能力的提示方法, 能够逐步拆解问题;
 - 3、2022 年底, ChatGPT 推出后, Auto-GPT/BabyAGI 等引入了任务列表与思维链形式, 可以按照任务规划拆分行动分步执行, 大模型初步拥有了解决复杂任务的能力, 这是把 AI 从工具变成 Agent 的分水岭;
 - 4、2024 年 9 月 OpenAI 推出了 o1 模型, 强调其在回答问题前会进行“思考”, 具备更强解决复杂任务的能力, 而 Anthropic 在 2025 年 2 月发布的 Claude 3.7 Sonnet 是首个采用混合推理的大模型, 具备深度思考模式, X 在同月也发布了 Grok3, 具备深度思考能力的大模型。
 - 至此大模型进入到了推理时代, 模型对目标的规划和不断迭代尝试的能力进一步提升。例如询问 Grok3: “美股中, 受益于 ai 或者使用 ai 或者最容易接入 ai 的标的有哪些, 帮我详细梳理一下”, 可以看到其进行了 9 个步骤, 搜索了 18 个信息来源, 期间反复调整解决问题的思路, 最终形成结果列表。

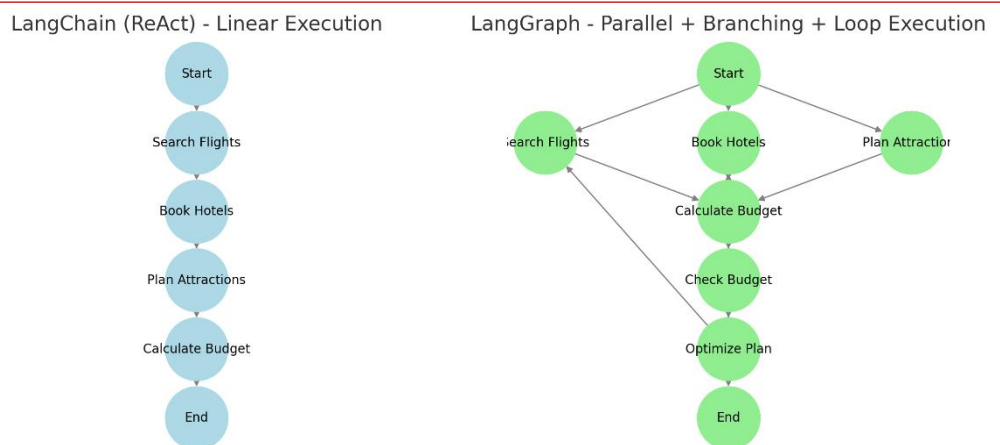
图表 7: 利用 Grok3 进行 DeeperSearch 处理问题

<p>🔍 DeepSearch</p> <p>深度搜索</p> <p>18 个来源</p> <ul style="list-style-type: none"> 👉 理解用户问题 👉 探索更多 AI 股票 👉 分析 AI 公司列表 👉 验证更多公司 👉 整理公司分类 👉 优化列表完整性 👉 比较市场价值 👉 确认更多 AI 公司 👉 整理最终列表 	<p>理解用户问题</p> <ul style="list-style-type: none"> • 用户的问题是：在美国股市中，哪些股票受益于 AI、使用 AI 或最容易整合 AI？让我来详细列出它们。 • 看起来用户想知道与 AI 相关的公司，特别是那些在产品或服务中受益于 AI 的公司。 • 我在考虑如何找到符合这些类别的股票，比如提供 AI 技术或服务的公司。 • 比如 Nvidia 的 AI 芯片，或在医疗和金融中使用 AI 的公司，如药物发现或算法交易。 • 当前时间是 2025 年 3 月 26 日星期三下午 5:28，+08 时区，这很重要，因为信息需要与这个时间相关。 • 我觉得最好的开始方式是通过网络搜索找到一些相关信息，比如“2025 年值得购买的顶级 AI 股票”。 • 网络搜索显示了一些 2025 年 AI 相关股票的列表，包括 Nvidia 和微软 (NBC News)。 • 等等，用户的实际问题是“beauty stock in AI”，可能是个笔误，我需要确认是否是“AI 相关美容股票”。 • 我在思考是否需要澄清用户的意图，特别是“beauty stock”是否是个拼写错误。 • 目前正在探索如何处理可能的笔误，继续寻找与 AI 相关的股票信息。 • 我在考虑用户的目标可能是投资 AI 相关股票，所以需要提供更准确的信息。 • 网络搜索还提到了其他 AI 相关股票，如 Alphabet 和 Amazon，涉及 AI 技术或服务的公司 (Forbes)。 <p>探索更多 AI 股票</p>
---	---

来源: Grok3, 中泰证券研究所

- LangGraph 框架简化 AI Agent 的开发，引入复杂任务循环概念。LangGraph 是由 LangChain 团队开发的一个开源框架，旨在帮助开发者构建具有状态管理能力的多 Agent 或单 Agent AI 应用。LangGraph 引入了循环图 (Cyclical Graph) 概念，弥补了 LangChain 在处理动态、复杂任务时的局限性。简单而言，就是用户在调用每个 Node (节点) 功能前，都会有一个 Edge (边) 根据 state (当前状态) 判断是否需要再进行下一个 Node (节点) 调用，或者是 END (结束) 流程，这样可以使整个处理流程自主循环直到解决问题。

图表 8: LangGraph 与 LangChain 实现路径对比 (模拟制定旅行规划)



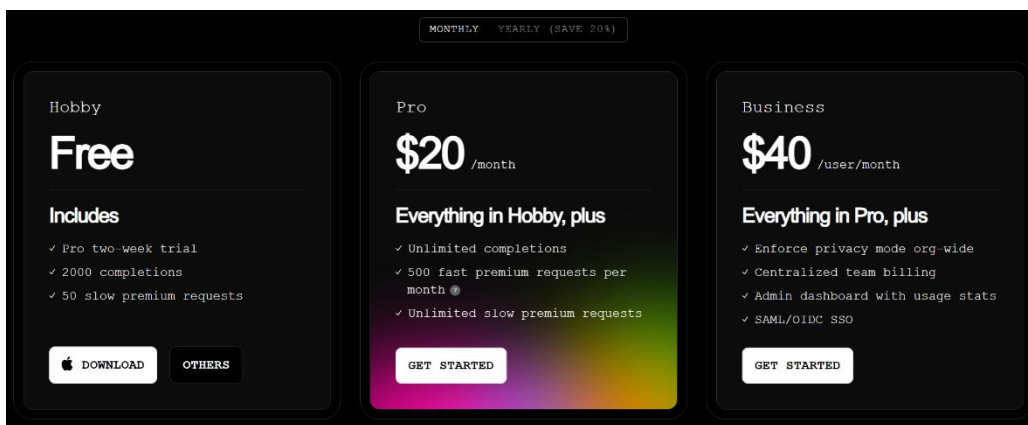
来源: 中泰证券研究所

- 这一类的开发框架的完善，也推动了整个 AI Agent 领域的发展。

Coding (代码) 就是大模型的自然语言

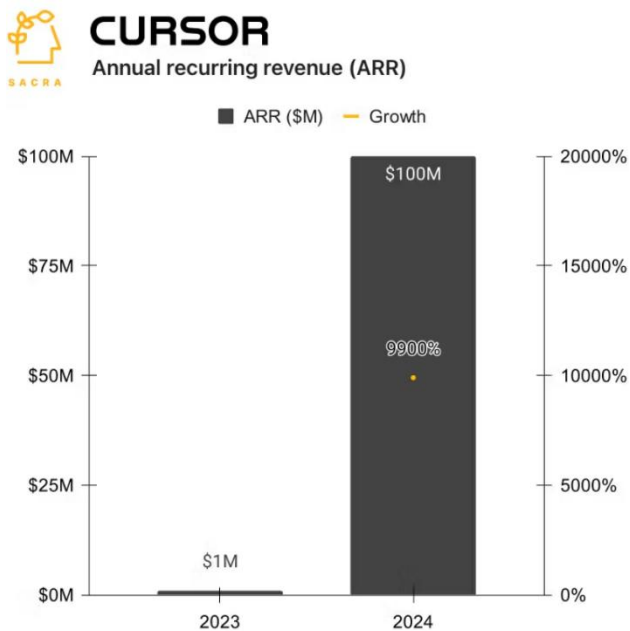
- 虽然当前的 AI Agent 还很难完成较为复杂、逻辑要求缜密、精度与准确度较高的任务,但我们第一章提及的上下文长时记忆+Tool Use+基础大模型更加智能且推理能力不断增强,这几方面的发展非常迅速,再结合当前例如 Anthropic claude sonnet 3.7 或 Gemini 2.5pro 等的强大 Coding 能力,使得大模型在拥有更强智慧大脑的同时,用机器能听懂的命令,完成大模型的思考到执行的过程。
- 随着 AI 大模型代码能力的不断增强,我们认为本质上, 代码就是大模型和机器交互的自然语言, 在没有大模型和 AI Agent 之前,人类需要通过学习编程,获得与机器交流的能力,让机器懂得如何去按照我们的要求完成工作,因而演化出了各类软件,演化出了 SaaS、PaaS、IaaS 等平台,以及各类的 APP,大大提高了生产效率。
- 如今, AI Agent 就是架设在人类与机器之间的中间层,即可以理解人类的自然语言,也可以编写机器看的懂的自然语言(代码),这就完成了全人类的效率提升。因为写代码的能力不是人人具备,而且现成的工具也不一定能满足所有人个性化需求,并且工具很多基于部分人的智慧,实际很难集成大量的知识库做出最优决策并结合实时情况更新解决方案和工具,但有了一个拥有极强 Coding 能力的 AI Agent 之后,这一切都将成为可能。
- **举例说明:** 由于大模型 Coding 能力的不断增强,基于大模型推出了不少编码工具,包括 Claude Code、Cursor 和 Cline,他们分别专注于终端、代码编辑器和 VS Code 集成。国内字节跳动的 Trae 也是一个支持免费 AI IDE,通过接入 LLM 技术提升开发代码开发效率。
 - Claude Code 是 Anthropic 开发的一个命令行 AI 编码助手,帮助开发者通过自然语言命令执行代码搜索、编辑、测试和 Git 操作。
 - Cursor 是基于 Visual Studio Code 的 AI 增强代码编辑器,提供代码生成、智能建议和任务完成代理模式,它集成了多种 AI 模型,适合个人和团队提高生产力,根据 Sacra 估计, Cursor 是 ARR 从 100 万美元增长到 1 亿美元有史以来最快的 SaaS 公司, Cursor 订阅的定价分为免费、20 美元/月、40 美元/月。2025 年 1 月, Cursor 宣布完成 1.05 亿美元的 B 轮融资 (Thrive Capital 融资), 估值为 25 亿美元。

图表 9: Cursor 的订阅价格



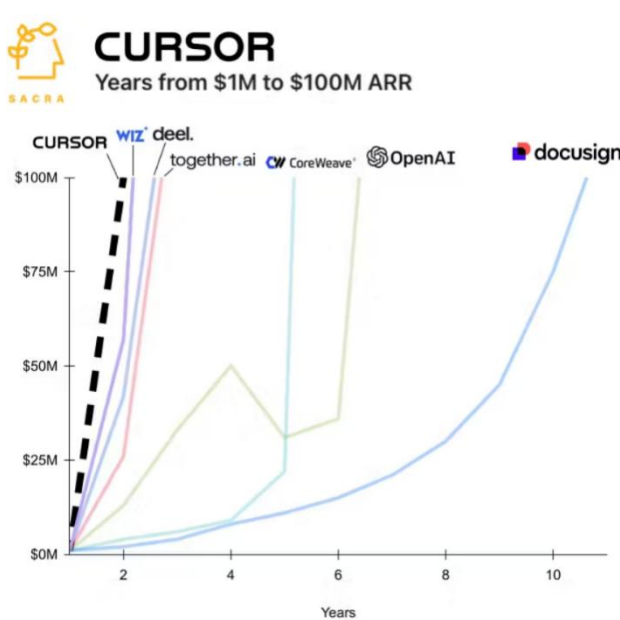
来源: Cursor 官网, 中泰证券研究所

图表 10: Cursor 年度经常性收入 (百万美元)



来源: sacra, 中泰证券研究所

图表 11: Cursor 完成 1 亿美元 ARR 的时间



来源: sacra, 中泰证券研究所

- Cline 是一个开源的 AI 自主编码代理, 集成到 VS Code 中, 可创建/编辑文件、执行命令、使用浏览器等。

- 大模型在 Coding 领域的突破, 预计会带动 LLM 和工具的进一步紧密结合, 无限拓展 LLM 能力边界, 向实现 AGI 迈进。

AI Agent 实现路径探寻

模型自治 + 外部协同

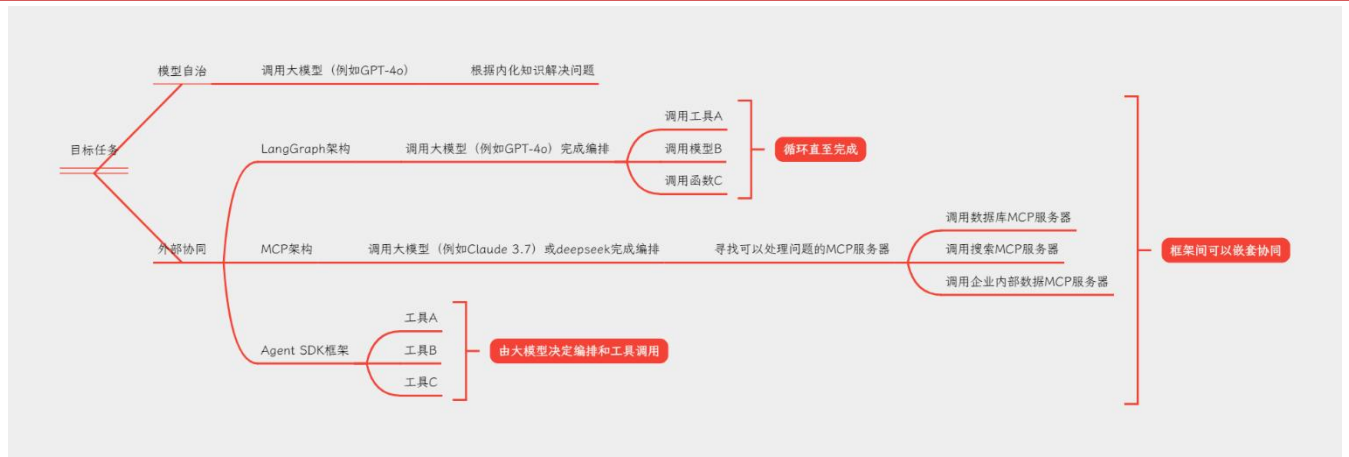
- AI Agent 的发展, 总结看来朝两个方向不断推进:
 - 大模型能力内化: 可以简称“内脑”, 通过监督学习+强化学习 (包括对齐人类指令的 RLHF 以及适合自主学习和探索性的 SELF-RL) 方式, 不断增强模型的逻辑能力, 目的是提升模型自主推理的能力, 甚至部分

内嵌功能，例如搜索、规划等功能，我们实际在调用 AI 大模型的时候，模型就自主开始搜索互联网信息，无需外接 API 工具实现。

● 外化框架的协同：

- 1、**规范工具调用**：上文提到的 LangGraph 架构，包括 Anthropic 的 MCP 协议，以及 OpenAI 开源的 Agent SDK 框架，本质上都是通过预先设定的逻辑关系，使得 AI 大模型在解决目标问题时，遵循一定规则和框架协议，可以更加高效的调用外部工具，实现跨模型跨平台统一调用，提升处理问题的能力。
- 2、**多模型协同**：不同的 AI 大模型特长不同，可以集各家所长，更好应对用户泛化的需求，使 AI Agent 可以处理更多复杂问题。例如基础大模型负责综合调度，而擅长图片处理、视频分析、数据整理等的微调垂模型可以承接不同环节，多模型共同协作完成目标任务。特别是在特定领域，需要专业知识的情况下，往往通用大模型是无法很好胜任，例如金融、医药等领域，要求模型具备更加专业的能力，这是就可以通过通用模型+专精垂模型配合的模式协同完成。

图表 12: AI Agent 实现路径探寻



来源：中泰证券研究所

■ 当前各路径存在的问题：

- **模型能力内化**：模型的参数会变得非常庞大，处理速度可能降低，训练+推理算力消耗会更大，并且应对不断变化的环境，模型需要不断进行新的训练，无法做到实时更新，且训练过程中的不可控因素较大。
- **外化框架协同**：存在多工具调用协同时 token 消耗大幅增加的问题，且不同工具之间的切换，需要大模型能够有较长记忆及综合处理能力，处理效果或许没有内化到大模型中的效果好。同时多模型之间协同，频繁的模型调用对基础设施要求高，模型间的通信信息传递损失，延迟积累等问题也需要处理。

■ 上述提及的各类问题，很多是短期工程的优化问题，我们预计随着辅助框架的不断完善和 AI 大模型自身能力的提升，将逐步得到解决。且预计两种路径会在不同的需求场景下不断推进，例如：

- **B 端商业场景**：强调数据的安全性，考虑 LangGraph 与 MCP 协议结合的方式，使用开源的大模型+垂模型+外化工具的组合，完成目标任务，这需要企业自有算力部署，且内部 API 开发完善。
- **C 端需求场景**：优先选成熟框架（OpenAI SDK、LangChain）快速实现功能，把更多决定权交由大模型。

- **小结：**外化+内化结合的 AI 大模型，犹如拥有强运动天赋的运动员配备上了专业的设备。举例：当前大模型推出的类 Deep Research 功能，就是相当于把搜索、计划等能力内化在 AI 大模型中，同时又支持工具调用，如数据库查阅或者用户上传文件辅助研究，内外集合，大幅提升处理效果。

AI Agent 应用算力消耗假设

一个全球爆款 AI Agent 应用需要多少张 H100 支持？

- 做一个假设测算，希望能大概估算，一个爆款的 AI Agent 应用，需要多少块 NVIDIA H100 GPU 算力支持，基于假设的测算思路为：
 - 1、通过 OpenAI 对 Plus 用户月订阅价格倒推用户每日一次交互的 token 消耗量。
 - 2、以计算出来的每日每次交互的 token 消耗量为基准，假设 AI Agent 应用的日活以及每日交互次数，计算出其一天算力的消耗量。
 - 3、以 NVIDIA H100 SXM GPU 的 FP8 算力为基准，计算在不同算力利用率情况下的不同日活的 AI Agent 应用对应的 GPU 需求数。

图表 13: H100 SXM GPU 参数一览

Technical Specifications		
	H100 SXM	H100 NVL
FP64	34 teraFLOPS	30 teraFLOPS
FP64 Tensor Core	67 teraFLOPS	60 teraFLOPS
FP32	67 teraFLOPS	60 teraFLOPS
TF32 Tensor Core*	989 teraFLOPS	835 teraFLOPS
BFLOAT16 Tensor Core*	1,979 teraFLOPS	1,671 teraFLOPS
FP16 Tensor Core*	1,979 teraFLOPS	1,671 teraFLOPS
FP8 Tensor Core*	3,958 teraFLOPS	3,341 teraFLOPS
INT8 Tensor Core*	3,958 TOPS	3,341 TOPS
GPU Memory	80GB	94GB
GPU Memory Bandwidth	3.35TB/s	3.9TB/s
Decoders	7 NVDEC 7 JPEG	7 NVDEC 7 JPEG
Max Thermal Design Power (TDP)	Up to 700W (configurable)	350-400W (configurable)

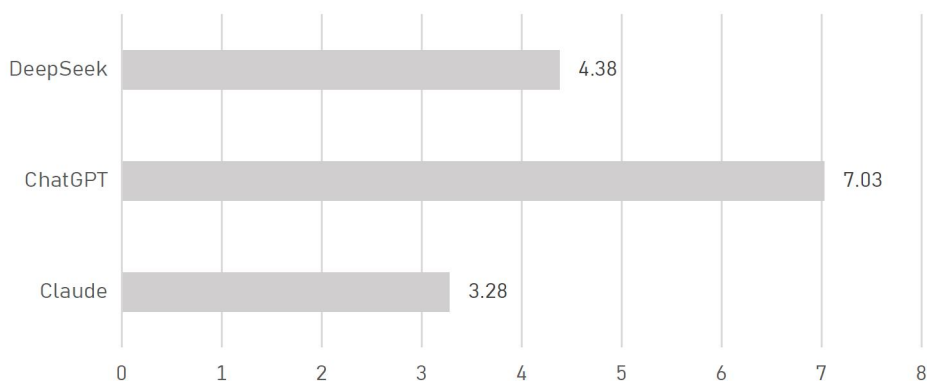
来源：NVIDIA 官网，中泰证券研究所

- 详细步骤：
 - 步骤 1：
 - 月订阅价格倒推 token 消耗量:这部分计算的依据的假设为 OpenAI 在设定订阅价格时，考虑了每个 Plus 用户每天的 tokens 消耗量，以此为基准，设

定了合理的订阅价格。但是由于免费用户也可以无限制使用例如 GPT-3.5、GPT-4o mini 以及有限制使用 GPT-4o、OpenAI o3-mini 等，因此考虑 Plus 用户订阅价格设定时，OpenAI 应需要覆盖免费用户的成本开支，加之还有数据中心、电力、硬件折旧等开支。

- 因此我们用敏感性分析，假设一定的百分比作为实际 Plus 用户开支覆盖其自身 Token 消耗部分。根据 SensorTower 数据显示，截止 2025 年 2 月底，ChatGPT APP 每日平均交互次数为 7.03 次，DeepSeek 为 4.38 次，Claude 为 3.28 次，因此我们把日交互次数也做一个敏感性分析，得出下表：

图表 14: DeepSeek、ChatGPT、Claude 日均交互次数 (次)



来源: SensorTower, 中泰证券研究所

图表 15: Token 消耗量敏感性分析

订阅价格折扣%/ 日交互次数 (次)	每日每次交互 tokens 消耗量 (个)				
	20%	40%	60%	80%	100%
3	5926	11852	17778	23704	29630
4	4444	8889	13333	17778	22222
5	3556	7111	10667	14222	17778
6	2963	5926	8889	11852	14815
7	2540	5079	7619	10159	12698
8	2222	4444	6667	8889	11111

来源: 中泰证券研究所

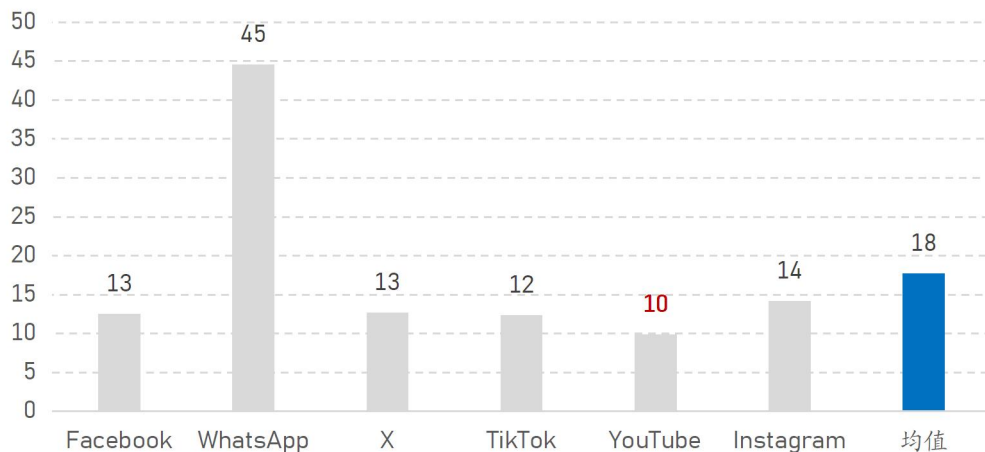
- 计算结果显示，最高的每日每次消耗 token 量为 29630 个，最低为 2222 个，以日交互 5 次与 6 次/订阅价格折价 60%这两个中间水平条件对应的 2 个数值为基准，计算平均消耗量为每日每次交互 9777.8 个。

■ 步骤 2:

- 做第二层的敏感性分析，由于 OpenAI Plus 用户系统默认采用 GPT-4o 大模型，因此我们以 GPT-4o 的参数作为计算基准，根据微软与华盛顿大学发表的论文《MEDEC: A Benchmark for Medical Error Detection and Correction in Clinical Notes》披露显示 GPT-4o 参数量为 2000 亿，我们对每次激活的参数量做敏感性分析，对应假设这个超级 AI Agent APP 每日的打开次数在 3~10 次范围内(基于当前全球超级 APP 中交互频次相对低一

些的 Youtube 和 AI 大模型应用中较低一些的 Claude 作为基准范围), 以激活参数+日交互次数为两个计算参数, 基于上述的 Token 消耗结果平均值, 做二层敏感性推导, 计算每日的算力总消耗, 得出下表,

图表 16: 代表性超级 APP 的日均交互次数 (次)



来源: SensorTower, 中泰证券研究所

图表 17: 每日算力消耗量敏感性分析

激活参数比%/ 每日交互次数 (次)	每日的算力总消耗 (FLOPs/day)				
	5%	25%	50%	75%	100%
3	7.3E+14	3.7E+15	7.3E+15	1.1E+16	1.5E+16
4	9.8E+14	4.9E+15	9.8E+15	1.5E+16	2.0E+16
5	1.2E+15	6.1E+15	1.2E+16	1.8E+16	2.4E+16
6	1.5E+15	7.3E+15	1.5E+16	2.2E+16	2.9E+16
7	1.7E+15	8.6E+15	1.7E+16	2.6E+16	3.4E+16
8	2.0E+15	9.8E+15	2.0E+16	2.9E+16	3.9E+16
9	2.2E+15	1.1E+16	2.2E+16	3.3E+16	4.4E+16
10	2.4E+15	1.2E+16	2.4E+16	3.7E+16	4.9E+16

来源: 中泰证券研究所

■ 计算结果显示, 最高的每日消耗算力量为 4.9E+16 FLOPs, 最低为 7.3E+14 FLOPs, 若依旧考虑中位线水平, 则每日交互 6/7 次+50%的参数激活对应的 1.6E+16 FLOPs 的总算力需求。

■ **步骤三:**

■ 我们做第三层的敏感性分析, 考虑到 NVIDIA H100 SXM GPU 在实际推理的时候算力的利用率是有限制的, 并不能完全 100%的跑满理论算力, 因此我们将 GPU 的算力利用率按一定比例做敏感性分析, 同时超级 AI Agent APP 需要有一个日活的门槛线, 我们对于日活也做了一定的选择范围, 然后以 GPU 的算力利用率+日活为两个计算参数, 最终计算 NVIDIA H100 SXM GPU 卡的需求量, 得出下表,

图表 18: H100 SXM GPU 需求量敏感性分析

英伟达 H100 GPU 需求张数 (万)					
日活 (亿) / 算力利用率%	1	5	10	15	20
10%	7.07	35.37	70.73	106.10	141.46
30%	2.36	11.79	23.58	35.37	47.15
50%	1.41	7.07	14.15	21.22	28.29
70%	1.01	5.05	10.10	15.16	20.21
90%	0.79	3.93	7.86	11.79	15.72

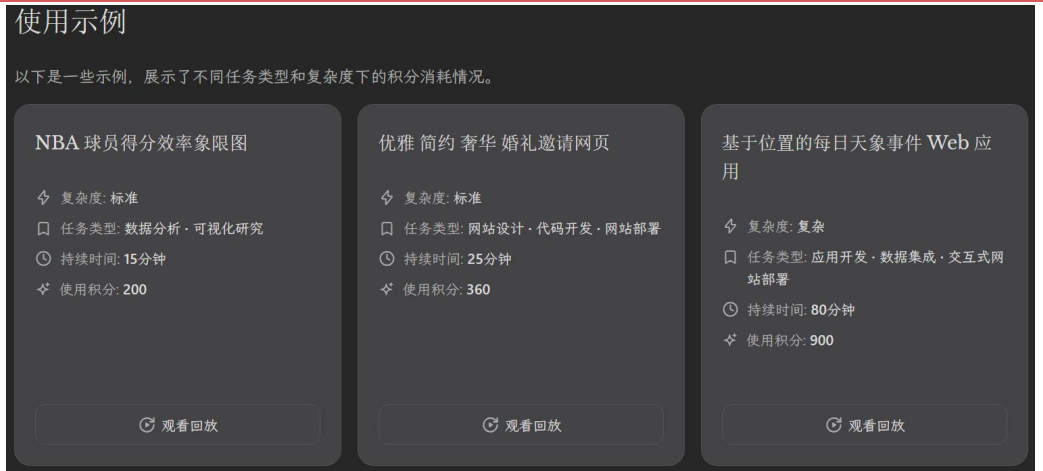
来源: 中泰证券研究所

- 计算结果显示, 取敏感性分析的中位值, 对应 50%的算力利用率+10 亿日活情况下, 一个超级 AI Agent APP 运行一天所消耗的算力需要 14.15 万张 NVIDIA H100 SXM GPU 支撑。
- **总结:** 以上测算基于 GPT-4o 模型, 但 GPT-4o 模型为基础大模型, 若考虑推理能力更强的 OpenAI o1 大模型, 或者在 GPT-4o 基础上使用 OpenAI 的 Deep Research 功能, 那实际的 token 消耗量会更大, 且当前 GPT-4o 已经开放图片模态的输出, 算力消耗预计会较文字更大, 且未来文字+图片+语音+视频的多模态输入和输出预计将成为标配, 那算力的消耗将比仅图文会更大。同时考虑到 AI Agent 本身可能基于多重的思维链推理, 调用外部工具, 特别在执行复杂的任务时, 存在多重的并行调用和循环执行, 那 Token 的消耗量将更加惊人。
- 因此以上的计算仅在当前环境背景下, 做了诸多的假设, 作为算力需求的一个概览。

案例分享: Manus 与 Cline + MCP + claude sonnet 3.7

- **1、Manus:**
- 就此前火爆的定义为通用 Agent 的应用 Manus, 公布了其订阅的价格, 分为 Starter/Pro 两档, 分别对应 39 美元/月与 199 美元/月, 其中 Starter 用户的权益包括每月 3900 积分以及并行运行两个任务, Pro 用户可以拥有 19900 积分, 并同时运行 5 个任务。同时 manus 也给了相应的案例, 复杂度为标准的“NBA 球员得分效率象限图”预计持续时间 15 分钟, 需要消耗 200 积分, 对应单次任务 2 美元, 时间上预估 0.133 美元/分钟, 复杂度同样为标准的“优雅 简约 奢华 婚礼邀请网页”设计需求需消耗 360 积分, 预计持续时间 25 分钟, 对应单次任务 3.6 美元, 时间上预估 0.144 美元/分钟, 而复杂度为复杂的“基于位置的每日天象事件 WEB 应用”则需要消耗 900 积分, 预估时间 80 分钟, 对应单次任务 9 美元, 时间上 0.113 美元/分钟。

图表 19: Manus 不同复杂度任务积分消耗



来源: Manus 官网, 中泰证券研究所

- 以此看，单次完成任务消耗成本至少在 2 美元以上，换算成调用 GPT-4o 的 API 价格（按输出 10 美元/百万 Token），对应 Token 消耗量为 20 万个，若为复杂的任务，单次任务 9 美元，则对应 Token 消耗量为 90 万个。

■ **2、Cline + MCP + Claude Sonnet 3.7:**

- 上文提到了 MCP 协议，目前已经看到包括 Unity、Blender 等软件社区开发了 MCP 服务器提供接入，地图相关 Google Maps 与百度地图、高德地区均有相应的 MCP 服务器，还有网络搜索、文件处理、数据库分析、思维规划等应用均接了 MCP 服务器。

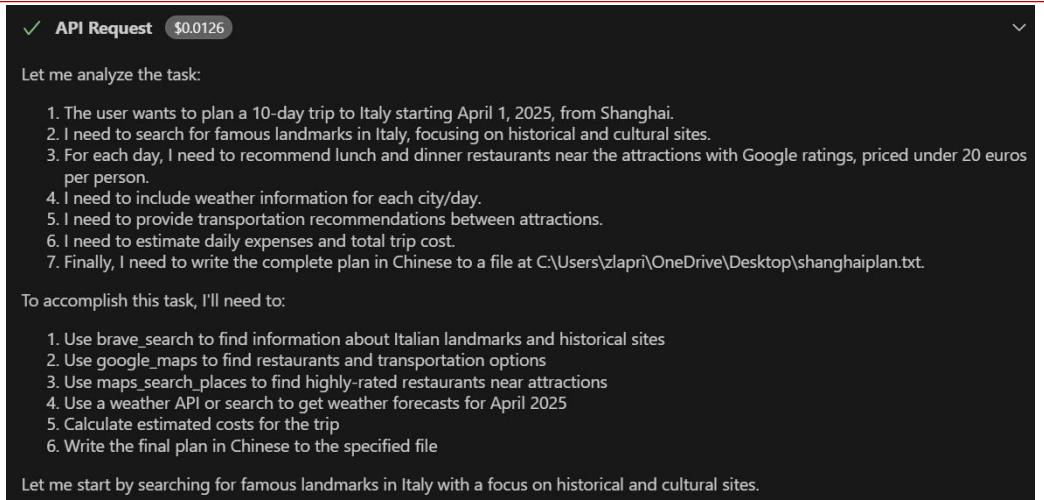
- 目前在 MCP 集中 Smithery 平台上，已经有 3579 个相关服务应用，预计后续会有越来越多的应用接入 MCP，开发 MCP 的服务器，大模型进入应用的场景会越来越多，包括 OpenAI 的 Agent SDK 也扩展支持了 MCP 协议的调用。可以想见的是，越来越多应用将集成大模型，提升用户使用体验，进而也不断增大了 token 的消耗和算力的需求，试想一下，若每个软件都接入大模型作为基础服务层。

- 我们做了一个旅行的案例：集成了 Brave 浏览器 MCP、Google Maps MCP、anthropic 官方 Filesystem MCP，通过 Cline 客户端，接入 Claude Sonnet 3.7 大模型，给予 prompt 为：“我想去意大利度假，计划 2025 年 4 月 1 日出发。搜索下意大利有什么著名的地标景点，你从中筛选偏人文历史的景点。按规划的时间，在中午和晚上就餐时间，所在游玩景点旁边推荐一家 google 评分高的餐馆，价格在 20 欧元一人以内。计划差不多 10 天的时间，每个城市当日的天气也需要标注，包括是否下雨或者气温，给我一个行程安排，我从上海出发，给我具体的景点间转换乘坐建议。按照你的规划和交通费用的预计，给出每天的支出预算，最后总结整个旅行的支出总计。最后在 shanghaiplan.txt 中写好中文计划，流程用中文展示”，整体步骤分为：

- 1、调用大模型进行任务规划：claude 将任务逐步拆解，分成了 7 个需求点。
- 2、生成了 6 部的处理方案，包括先调用 Brave 的 MCP 进行搜索知名的城市及景点，然后使用 Google Maps MCP 进行交通和餐馆的发掘，从

景点附近找出高分餐馆，再使用天气 API 搜索当地天气，最后预估旅行费用及输出最终形成规划文档。

图表 20：大模型进行任务规划示意图



来源：Cline，中泰证券研究所

- 整个流程总共向大模型输入了 9K Tokens，同时大模型输出了 933.4K Tokens，合计整个需求消耗了 942.4K Tokens，cline 统计对应的调用资金消耗为 0.5904 美元。这意味着一个简单的旅行规划，一次请求，就消耗了接近百万的 Tokens，那如果整个流程中加入酒店、餐馆、机票等预订，再加入各个景点的历史人文背景介绍，再添加相应的循环调整机制，加入知识的搜索，例如搜索人的喜好习惯记忆，历史上旅行的偏好，身体的状况能接受怎么样强度的旅行等等，则消耗会更加巨大。

图表 21：大模型进行任务规划结果



来源：Cline，中泰证券研究所

AI Agent 应用场景展望与投资建议

现有应用场景分析

- 根据通用 Agent Manus 官网所列示案例，可以窥见一些 AI Agent 的使用场景，案例大致可以分为：1、金融与投资（12 个）；2、市场与品牌传播（8 个）；3、科技与互联网（6 个）；4、教育与培训（6 个）；5、消费零售与电商（3 个）；6、招聘与 HR（2 个）；7、房地产与居住服务（2 个）；8、保险

与医疗（2个）；9、办公效率与文档工具（2个）；10、旅游与生活（1个）；11、家庭与人文（1个）；12、文化创意与多媒体（1个）；13、法律与合规（1个）。

■ 综合各案例，可以总结衔接端 AI Agent 应用几大特征；

1) **非标内容类**:以内容输出为主，涉及到产业研究、文化创意、课程设计或生活需求，这类需求没有标准化输出限制，对内容的精确度没有强制要求，满足诸多非标个性化需求；

2) **标准内容类**:有明确数据或内容支撑，例如电商运营分析、房地产数据分析、医疗数据分析等，该类内容在外部数据支持下，大幅降低模型幻觉，输出更加可靠；

3) **模板化分析类**:内容有相对固定格式、固定模板、固定规则，可批量化操作，可复用且流程较为繁杂，例如利用 AI Agent 校对文本、分析合同、语音转录翻译、格式转换等。

图表 22: Manus 的应用分类

行业分类	项目名称	简要说明	行业分类	项目名称	简要说明
#金融与投资	深度分析特斯拉股票	提供股票财务摘要与可视化分析报告	#消费零售与电商	在线商店运营分析	针对亚马逊商店的销售数据分析
	NVIDIA财报分析	解读NVIDIA财报数据并优化摘要		制定商店销售改进策略	社区人口结构与消费数据分析
	财务报表审查与核实	审阅公司财务数据与核对准确性		电子商务产品评价分析	用户评论数据挖掘和偏好建模
	化妆品文献研究	研究护肤产品中的市场趋势与用户反馈	#房地产与居住服务	在纽约买房	综合家庭情况进行房产推荐
	AWS市场份额研究	研究AWS云服务市场份额		租赁合同分析	将合同要素结构化展示
	Github项目研究	分析Github开源项目走向	#法律与合规	合同审查	分析合同内容合法性与漏洞
	B2B供应商采购研究	企业采购平台趋势调研	#保险与医疗	保险政策比较分析	不同保险政策条款对比
	YC公司列表研究	YC孵化公司趋势分析		区域病患数量估算	用POI与统计数据推算患者数量
	收集公众观点研究	舆论数据整合与分析	#教育与培训	动量定理互动课程	初中物理交互式课程
	AR/AI眼镜研究	可穿戴设备产业报告汇总		交互式Transformer课程	用网页动画讲解AI模型结构
过去十年美国AI研究	追踪AI研究论文发展趋势	探索宇宙的规模		宇宙测量规模交互课	
服装行业AI产品研究	AI在服装领域的应用调查	活动介绍图		教育活动的可视化呈现	
四月日本之旅	旅行定制计划与地图整合	学习资源收集与整理		强化学习资源的组织与整合	
#旅游与生活方式	家族关系图	可视化呈现家谱或家族网络	将讲座录音转为笔记	将语音内容结构化为重点笔记	
#家庭与人文	个性化冥想音频	制作定制音频内容	候选人面试时间表	高效安排候选人面试流程	
#文化创意与多媒体	音频工程	自然音素与音调结合设计音频	#招聘与HR	筛选强化学习工程师简历	自动评估与分析简历内容
	播客编辑	自动剪辑播客片段与生成摘要		参与Kaggle机器学习竞赛	编写高性能机器学习模型
#市场与品牌传播	名片设计	数字与印刷名片的个性化设计	#科技与互联网	数字体验分析	网站功能评估与用户体验报告
	面试视频推理	多模态分析面试内容与表现		网站SEO优化	提供网站的SEO结构分析和优化建议
	撰写Adobe公司传记	基于公开信息写作公司介绍		YC公司列表	初创公司分析，偏向互联网创业方向
	剧本创作	支持创意产业剧本生成与分析		社交媒体舆情研究	Twitter、YouTube等平台数据舆情分析
	品牌形象设计	品牌视觉系统与图标设计		收集公众观点	舆情数据整合与分析（多为线上舆情）
	专业提词器	专业演讲辅助工具	#办公效率与文档工具	校对长文本	优化文案结构与可读性
				将 API.xlsx 转 Word	将结构化Excel转为API文档格式

来源: Manus 官网, 中泰证券研究所

■ 以上是 AI 大模型发展至今，AI Agent 能力所能触达的领域，但当前对于精度要求高、需求更加复杂的情况，例如复杂的研究需要上百步的推理及工具调用，模拟真实情况下的各种异常环境等，AI Agent 还较难胜任。

投资建议

■ 当前在 AI 领域的投资，可分为几类：1)硬件支持；2)底层基础模型；3)中间层架构辅助；4)应用。

■ 标的梳理：

● 1、硬件与基础设施

AI 硬件和基础设施包括半导体制造商和云服务提供商，它们为 AI 技术提供计算基础。主要公司：NVIDIA（GPU 领导者）、AMD（GPU 和 CPU 竞争者）、Intel（AI 芯片开发）、TSMC（芯片制造）、中芯国际

(芯片制造)、Amazon (AWS, AI 云服务)、Microsoft (Azure, AI 云平台)、Alphabet (Google Cloud Platform-GCP, TensorFlow 支持)、CoreWeave (AI HPC 云服务提供商)、阿里巴巴 (阿里云)。

● 2、模型开发与研究

聚焦于开发基础 AI 模型。主要公司：Alphabet (通过 Google AI 和 DeepMind 研究)、Microsoft (与 OpenAI 合作)、Meta (Meta Llama AI 模型)、阿里巴巴 (通义千问 Qwen 大模型)、腾讯 (混元大模型)、科大讯飞 (星火大模型)、Baidu (文心一言大模型)。

● 3、平台与工具

提供 AI 开发平台、工具和中间件的公司，帮助开发者构建和管理 AI 应用。主要公司：Hugging Face (提供开源 AI 模型库和工具的平台，未上市)、Brevian (为企业提供无代码平台定制 AI Agent，未上市)、Anysphere (提供 AI 驱动的开发环境，未上市)、Coze (节跳动推出 AI Agent 开发平台，未上市)、Scale AI (为机器学习和人工智能模型提供高质量的训练数据和数据标注服务，未上市)、Turing (提供人类专家用于 AI 训练和微调，未上市)。

● 4、应用

- AI 驱动的行业应用，公司覆盖医疗、金融、零售、媒体、驾驶、教育等多领域。主要公司：特斯拉 (自动驾驶+机器人)、Salesforce (Einstein AI 平台，增强其 CRM 软件功能)、Adobe (Adobe Sensei 集成 AI，增强创意和营销软件的功能)、Palantir (提供 AI 驱动的数据分析平台，服务政府和商业客户)、快手 (可灵，文本提示生成高质量视频)、阿里巴巴 (夸克，AI 超级助手，集成 Qwen 模型高级推理能力)、腾讯 (元宝，提供文档分析、文本和图像生成以及交互式问答，利用混元大模型)、Boss 直聘 (利用 AI 促进求职者和雇主之间的高效匹配)、美图 (AI 驱动的照片和视频编辑应用)、多邻国 (AI 提供个性化的语言学习体验)。

■ 对应 A 股传媒板块：

- AI 教育：南方传媒 (粤教翔云数字教材平台上线 AI 助教、AI 助学功能，新华乐育推出校园 AI 智能体，以 AI 赋能基础教育，成立广东新华人工智能与教育研究院)、豆神教育 (推出 AI 超拟人多对一直播课及豆神 AI 等产品，与智谱合作成立合资公司专注于 AI 教育产品的技术研发及销售)、荣信文化 (AI 儿童教育践行者，已积累适用于 AI 应用的数据资产，公司开发 AI 智能点读笔结合行业大模型技术，提高智能及个性化阅读体验)、世纪天鸿 (推出 AI 驱动的教育工具“小鸿助教”，参股一笔两划，开发作文 AI 辅导应用)。
- AI 陪伴：奥飞娱乐 (推出 AI 毛绒陪伴玩具喜羊羊超能铃铛娃娃)、上海电影 (推动旗下 IP 与 AI 的结合，战略合作即梦 AI 与悦然创新)、汤姆猫 (推出汤姆猫 IP 系列 AI 陪伴产品)。
- AI 工具：果麦文化 (推出 AI 校对王，布局 AI 漫画等领域)、昆仑万维 (“天工”系列构建起 AI 模型和应用矩阵)、易点天下 (布局 AIGC，赋能营销、数据分析、智能化管理)。
- AI 娱乐：恺英网络 (投资 AI 陪伴公司自然选择)、巨人网络 (原始征途、太空杀等多款游戏接入 AI 玩法，业内首个基于多智能体大模型打造 AI 原生游戏玩法，推出 AI 美术生产平台-iImagine)、盛天网络 (AI 社交产品“带带”，AI 音乐产品“给麦”)。

风险提示

1、AI 大模型发展进度不及预期风险

- 当前基础大模型的发展迭代速度较快，从单一语言交互到多模态再到工具调用和推理，本报告基于当前大模型发展进度做出的诸多预估，可能存在大模型发展遭遇技术瓶颈进度低于预期的风险。

2、AI 应用发展进度不及预期风险

- 即使在大模型不断强大的当下，超级应用的落地场景依旧不明晰，存在应用落地进度明显低于预期的风险。

3、因技术与环境变化导致测算依据失效的风险

- 报告中部分数据测算基于多重假设，敏感性测算是基于当前市场和技术的发展阶段为依据，存在技术迭代速度远超预期，导致现阶段设定的计算参数失真的风险。

4、报告中数据信息更新不及时风险

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。