AI 商业化加速推进，量子科技前景广阔

——2026 年计算机行业策略报告

推荐|维持

报告要点：

- **行业创新加速推进，市场表现位居中游**

2025 年，计算机（申万）指数上涨 18.24%，上证指数上涨 18.41%，沪深 300 上涨 17.66%，创业板指上涨 49.57%，科创 50 上涨 35.92%。计算机板块跑赢沪深 300，跑输创业板指、科创 50 指数、上证指数，涨幅位居申万行业第 14 位。从技术演进角度来看，AI 技术不断创新演进，DeepSeek 以远低于海外大模型厂商的成本实现了先进的性能，模型成本不断降低带动 AI 技术平权，并推动大模型技术在千行百业的实际应用，token 消耗量实现成倍增长；资本市场方面，摩尔线程、沐曦股份等国产 GPU 厂商成功上市，智谱、MiniMax 等国产头部大模型厂商 2026 年初在香港上市，AI 全栈的国产替代正在如火如荼地推进。此外，量子科技、商业航天等细分方向也迎来较好的投资机会。

- **AI 技术快速迭代，中美差距逐步缩小，商业化加速推进**

2025 年初以来，生成式人工智能技术加速迭代，模型能力的高度不断超出想象，在减少幻觉、提升能力等方面进展显著，逐步蜕变为稳定可靠的专家助手；与此同时，模型使用成本也在快速下降，加速 AI 在行业端的渗透和应用。海外，美国科技巨头持续扩大资本开支，根据 Artificial Analysis 发布的《Q3 2025 Artificial Analysis State of AI Highlights Report》，2025 年以来，Amazon、Google、Meta、Microsoft 和 Oracle 的资本开支总和逐季度快速增长；从商业化方面来看，随着模型性能的快速提升与推理成本的下降，OpenAI、Anthropic、xAI 等的年化收入呈现快速增长趋势。国内，智谱、DeepSeek、MiniMax、阿里巴巴等持续加大投入，在技术领域不断创新突破，同时，商业化进展也逐步加快，长期发展空间广阔。

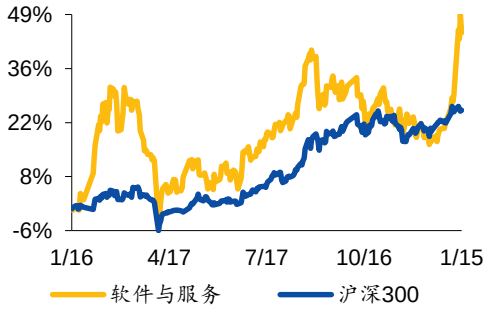
- **政策助力，量子科技前景广阔，商业化有望逐步开启**

2025 年 11 月，英特尔前首席执行官帕特·基辛格(Pat Gelsinger)接受英国《金融时报》采访：量子计算将成为未来计算体系中不可或缺的核心力量，并与经典计算、AI 计算共同构成推动科技变革的“三位一体(Holy Trinity)”，将在未来两年内进入主流市场，成为下一轮计算革命的关键。2025 年 10 月，《中共中央关于制定国民经济和社会发展第十五个五年规划的建议》中将量子科技纳入未来产业布局，进一步从国家战略层面明确量子科技产业的长期发展方向。海外，微软、谷歌、IBM、英伟达等具体持续加大投入，在量子软硬件领域持续突破，引领全球发展，商业化前景可期；国内，国仪量子、本源量子等企业紧跟技术趋势，不断突破，同时与下游客户紧密合作，商业化逐步开启，长期成长空间广阔。

- **风险提示**

人工智能大模型技术在软件及服务领域的落地节奏低于预期的风险；量子信息技术开发的风险；量子信息技术市场开拓与商业化进展不及预期的风险；行业竞争加剧的风险；宏观经济环境变动的风险；国际贸易与技术管制带来的风险。

过去一年市场行情



资料来源：Wind

相关研究报告

《国元证券行业研究-量子信息技术专题研究报告（二）：科技巨头加速布局，量子产业前景可期》2025.12.21

《国元证券行业研究-计算机行业周报：板块下跌幅度较大，阿里千问拓展 C 端市场》2025.11.17

报告作者

分析师 耿军军
执业证书编号 S0020519070002
电话 021-51097188-1856
邮箱 gengjunjun@gyzq.com.cn

分析师 王朗
执业证书编号 S0020525020001
邮箱 wanglang2@gyzq.com.cn

目 录

1. 行业创新加速推进，市场表现位居中游 4
1.1 2025 年行情回顾 4
1.2 行业财务表现概况 5
1.3 行业估值概述 6
2. AI 技术快速迭代，中美差距逐步缩小 8
2.1 海外巨头持续投入，引领人工智能发展 15
2.1.1 谷歌 17
2.1.2 OpenAI 19
2.2 国内企业成果显著，持续推动应用落地 21
2.2.1 智谱 21
2.2.2 DeepSeek 23
2.2.3 MiniMax 27
2.2.4 阿里巴巴 28
3. 量子科技前景广阔，政策助力产业落地 31
3.1 量子科技加速发展，引领未来产业变革 31
3.2 产业政策不断出台，未来发展动力强劲 32
3.3 国际巨头快速推进，引领量子产业发展 33
3.3.1 微软(Microsoft) 33
3.3.2 谷歌(Google) 34
3.3.3 国际商业机器公司(IBM) 35
3.3.4 英伟达(NVIDIA) 36
3.4 国内企业加速突破，夯实量子产业根基 37
3.4.1 国仪量子 37
3.4.2 本源量子 38
4. 风险提示 40

图表目录

图 1：2025 年申万各板块及主要指数表现情况 4
图 2：2015-2025 年前三季度计算机（申万）板块经营情况 5
图 3：2015-2025 年前三季度计算机（申万）板块利润水平（整体法） 6
图 4：2025 年 12 月 31 日申万各板块及主要指数 PE TTM（剔除负值） 6
图 5：过去十年间计算机（申万）指数 PE TTM（剔除负值） 7
图 6：2025 年前沿大模型持续快速迭代，智能水平不断提升 8
图 7：模型性能提升的三重 Scaling Laws 9
图 8：2025 年大模型上下文窗口突破百万 9
图 9：原生多模态架构与传统“拼接式”架构的区别 10
图 10：2025 年大模型推理价格不断降低 10
图 11：各类模型每周的 token 消耗量快速增长 11

图 12：AI Chatbots 产品网页流量增长趋势 12
图 13：2024-2029 年中国 AI 与生成式 AI 市场规模预测（单位：百万美元）13
图 14：2022-2030 年中国大语言模型市场规模及预测（单位：十亿元）..... 13
图 15：中美前沿模型差距逐步缩小 14
图 16：主流大模型的 AI 分析指数(Artificial Analysis Intelligence Index)排名 15
图 17：美国科技巨头持续加大 AI 投入 15
图 18：OpenAI 的算力版图 16
图 19：海外 AI 厂商年化收入快速增长 16
图 20：谷歌是极少数具备全栈 AI 能力的厂商 17
图 21：Gemini 3.0 Flash 以更低价格实现高性能 18
图 22：ChatGPT 上线群聊功能 20
图 23：GPT-5.2 的基准测试结果较上一代模型显著提升 21
图 24：智谱是中国领先的独立通用大模型开发商 22
图 25：智谱大模型及智能体 23
图 26：DeepSeek 模型发布时间线 23
图 27：DeepSeek-R1-0528 在各项评测集上均取得了优异表现 24
图 28：DeepSeek-V3.1 在编程智能体测评中的表现显著提升 25
图 29：引入 DSA 后模型长文本训练和推理效率大幅提升 25
图 30：DeepSeek-V3.2 在各类数学、代码与通用领域评测集中表现优异..... 26
图 31：MiniMax 发布了一系列多模态模型 27
图 32：MiniMax-M2 在编程、使用工具和深度搜索等基准测试上表现优秀 ... 28
图 33：通义千问 Qwen 模型家族 29
图 34：千问 APP 接入高德地图 30
图 35：《金融时报》对帕特·基辛格的采访 31
图 36：2025 年诺贝尔物理学奖 32
图 37：量子信息三大领域的发展定位和应用前景 32
图 38：微软首枚拓扑量子比特处理器 Majorana 1 34
图 39：Google 新一代 Willow 量子芯片 34
图 40：IBM 量子路线图 36
图 41：NVAQC 如何改变量子计算 37
图 42：许克标博士发布前沿量子传感器 38
图 43：“本源悟空”量子计算机模型 39

表 1：计算机（申万）板块 2025 年涨幅前 10 的个股 5
表 2：量子信息技术产业近期政策汇总 33

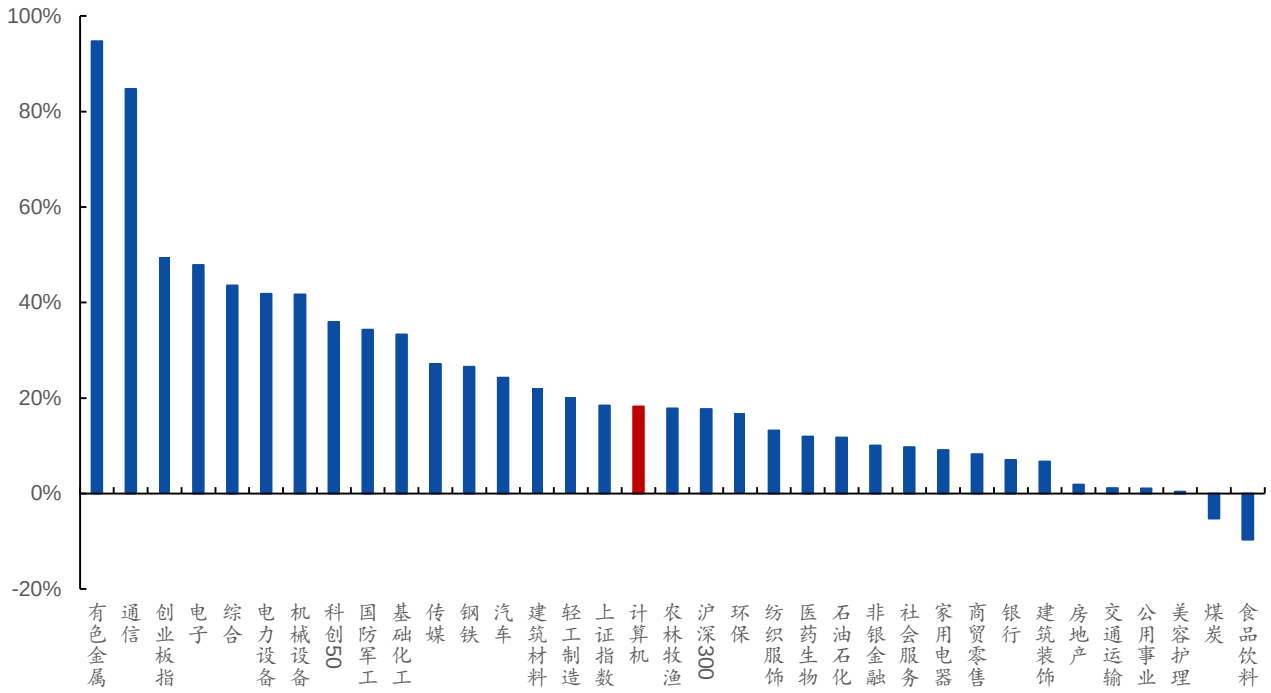
1. 行业创新加速推进，市场表现位居中游

1.1 2025 年行情回顾

2025 年，计算机（申万）指数上涨 18.24%，上证指数上涨 18.41%，沪深 300 上涨 17.66%，创业板指上涨 49.57%，科创 50 上涨 35.92%。计算机板块跑赢沪深 300，跑输创业板指、科创 50 指数、上证指数，涨幅位居申万行业第 14 位。

从技术演进的角度来看，AI 技术不断创新演进，DeepSeek 以远低于海外大模型厂商的成本实现了先进的性能，模型成本不断降低带动 AI 技术平权，并推动大模型技术在千行百业的实际应用，token 消耗量实现成倍增长；资本市场方面，摩尔线程、沐曦股份等国产 GPU 厂商成功上市，智谱、MiniMax 等国产头部大模型厂商 2026 年初在香港上市，AI 全栈的国产替代正在如火如荼地推进。此外，量子科技、商业航天等细分方向也迎来较好的投资机会。

图 1：2025 年申万各板块及主要指数表现情况



资料来源：同花顺 iFinD，国元证券研究所　　注：收盘价时间为 2025 年 12 月 31 日

2025 年，计算机（申万）板块涨幅排名前三的上市公司为：星图测控、品茗科技、海峡创新。

表 1：计算机（申万）板块 2025 年涨幅前 10 的个股

股票代码	公司名称	子行业	年初至今涨幅/相对 IPO 发行价涨幅
920116.BJ	星图测控	航天装备	1647.16%
688109.SH	品茗科技	工业软件	433.36%
300300.SZ	海峡创新	政务信息化	421.88%
688228.SH	开普云	基础及通用软件	338.22%
603516.SH	淳中科技	芯片与专业硬件	306.14%
300368.SZ	汇金股份	金融信息化	249.31%
301396.SZ	宏景科技	算力服务	231.97%
301638.SZ	南网数字	电力信息化	230.58%
688171.SH	纬德信息	信息安全	179.60%
300469.SZ	信息发展	交通信息化	162.98%

资料来源：同花顺 iFinD，国元证券研究所

注：收盘价时间为 2025 年 12 月 31 日，年内上市企业采用相对 IPO 发行价涨跌幅

1.2 行业财务表现概况

根据同花顺 iFinD 的数据，2025 年前三季度，计算机（申万）板块共实现营业收入 9386.14 亿元，同比增长 9.19%；实现净利润 244.14 亿元，同比增长 30.37%。

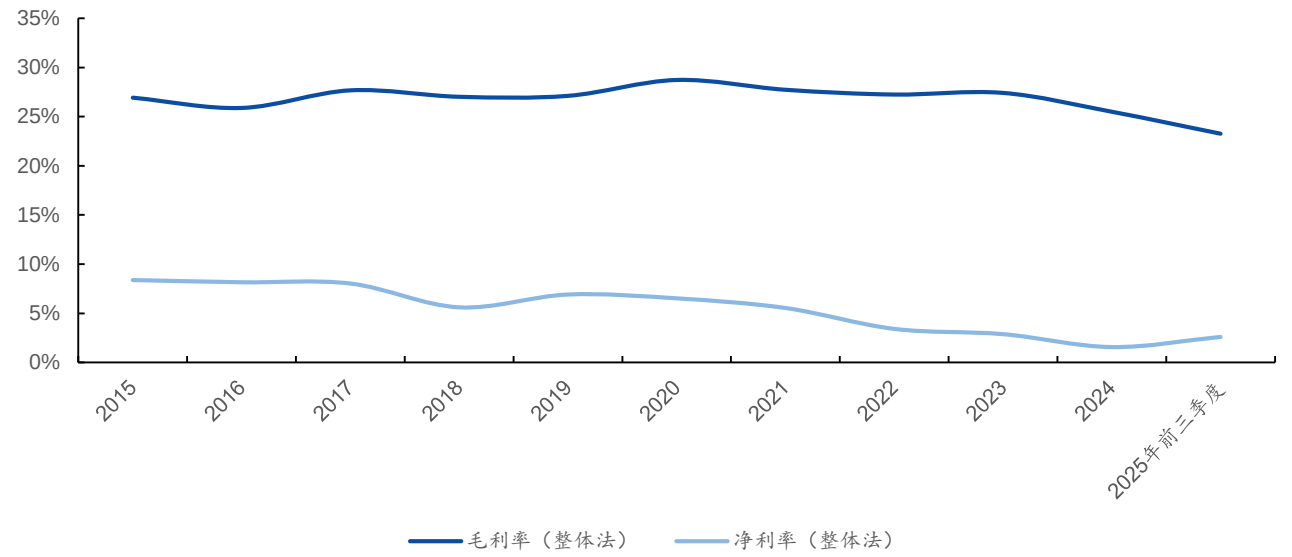
图 2：2015-2025 年前三季度计算机（申万）板块经营情况



资料来源：同花顺 iFinD，国元证券研究所

2025 年前三季度，计算机（申万）板块销售毛利率约为 23.26%，较 2024 年全年下降 2.23 个百分点；净利率为 2.60%，较 2024 年全年上升 1.03 个百分点。

图 3：2015-2025 年前三季度计算机（申万）板块利润水平（整体法）



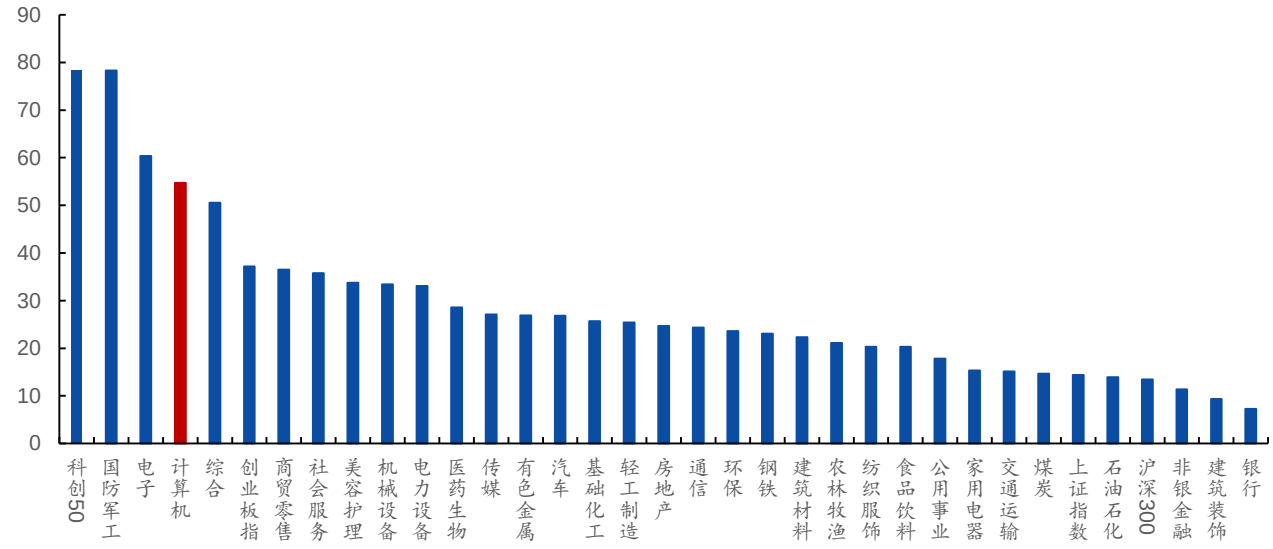
资料来源：同花顺 iFinD，国元证券研究所

1.3 行业估值概述

2025 年 12 月 31 日收盘，剔除负值后，计算机（申万）指数 PE TTM 为 54.70。申万各行业 PE TTM 排名前三的行业分别为：国防军工、电子、计算机，排名后三名的分别为：非银金融、建筑装饰、银行。

2025 年 12 月 31 日收盘，剔除负值后，上证指数的 PE TTM 为 14.36；沪深 300 的 PE TTM 为 13.45；创业板指的 PE TTM 为 37.18。

图 4：2025 年 12 月 31 日申万各板块及主要指数 PE TTM（剔除负值）



资料来源：同花顺 iFinD，国元证券研究所

目前，剔除负值后计算机（申万）行业的估值水平较年内高点有所回落。我们认为，计算机行业的成长属性赋予其较高的估值溢价，从历史估值水平来看，当前计算机板块估值水平处于合理区间，具备较好的长期投资价值。

图 5：过去十年间计算机（申万）指数 PE TTM（剔除负值）

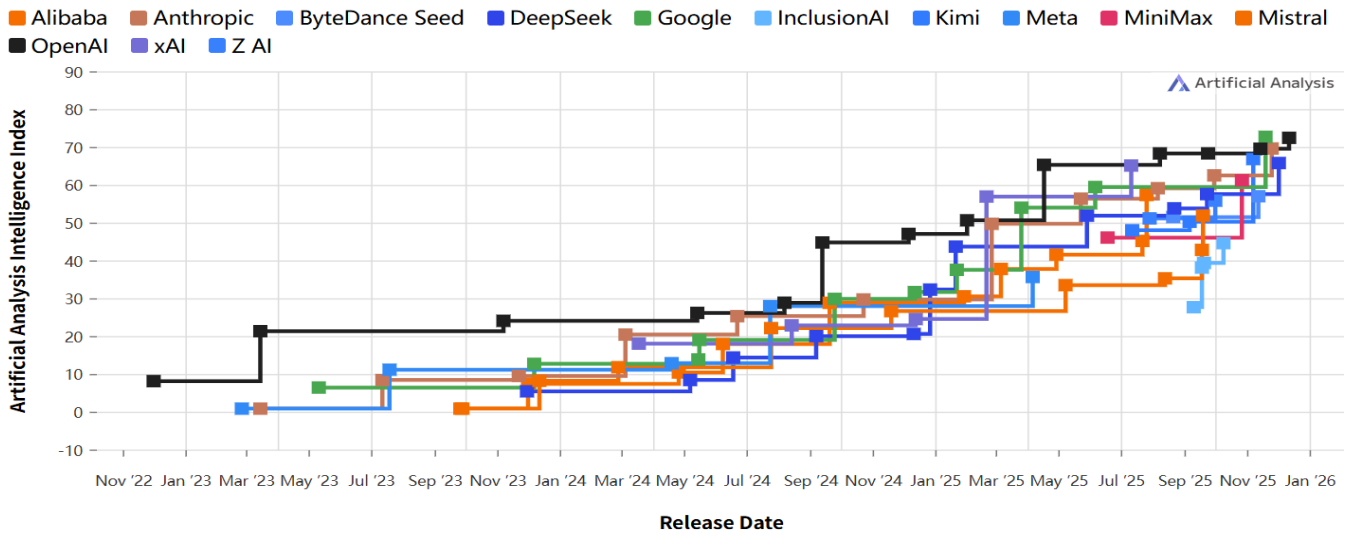


资料来源：同花顺 iFinD，国元证券研究所

2. AI 技术快速迭代，中美差距逐步缩小

2025 年初以来，人工智能技术迭代的步伐从未放缓，模型能力的高度不断超出想象，在减少幻觉、提升指令遵循能力等方面进展显著，并逐步蜕变为稳定可靠的专家助手；与此同时，模型使用成本也在快速下降，加速 AI 在行业端的渗透和应用，并反哺基础模型的不断优化，进一步推动了模型的应用和创新。

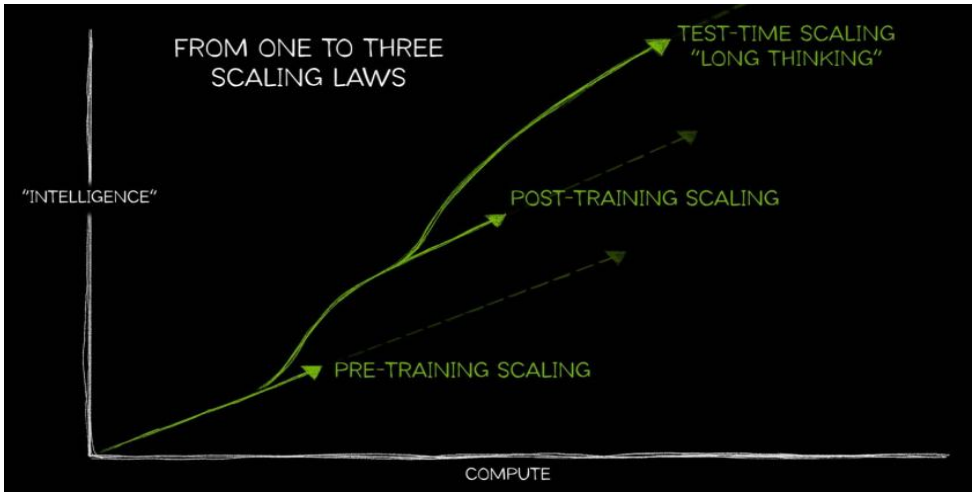
图 6：2025 年前沿大模型持续快速迭代，智能水平不断提升



资料来源：Artificial Analysis 官网，国元证券研究所

从模型性能提升的路径来看，Scaling Laws 持续有效，并从预训练阶段逐渐向后训练以及推理阶段延伸。预训练阶段的规模定律(Scaling Laws)持续有效，随着预训练阶段投入的数据量、计算资源以及模型参数的不断扩大，模型的性能呈现稳步提升的态势。但在近期，预训练阶段提高模型参数量带来的边际收益开始递减。为了继续提升模型解决长程问题的能力，在后训练(post-training)阶段使用强化学习等技术引导模型从被动应答转向主动求解，让模型的表现取得了突破。全球人工智能机构纷纷开始尝试挖掘强化学习在后训练阶段的巨大潜力，并推出推理模型，如 OpenAI o3、Gemini 2.5 Pro、DeepSeek R1、Qwen3 等。实践表明，在后训练阶段投入更多算力得到的推理，随着思考时间以及消耗 token 的增加，解决复杂问题的能力也随之提升。大模型将继续借助强化学习等技术持续提升思考能力，变得越来越聪明。当前，基于结果、过程和格式反馈的强化学习在数学、物理、编码等任务中取得了良好的效果，并表现出良好的泛化能力。

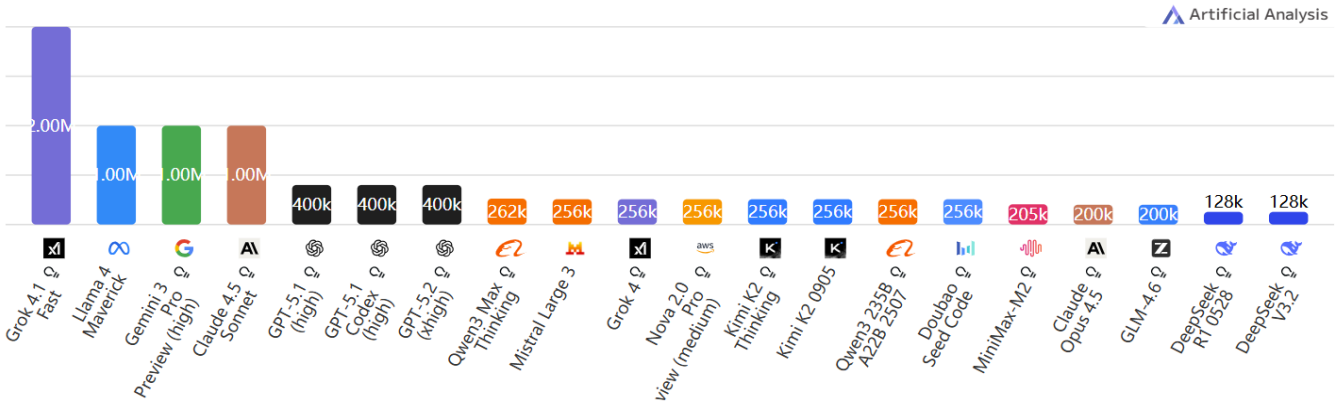
图 7：模型性能提升的三重 Scaling Laws



资料来源：AINLPer 公众号，国元证券研究所

顶尖大模型 context window 突破百万 token，这与 RAG（信息检索增强）LLM 工作量密切相关，有助于实现模型的高质量输出。根据 Artificial Analysis 的数据，Grok4.1 的上下文窗口可达 200 万 token，Gemini3 Pro 与 Claude 4.5 Sonnet 也达到了 100 万 token。

图 8：2025 年大模型上下文窗口突破百万

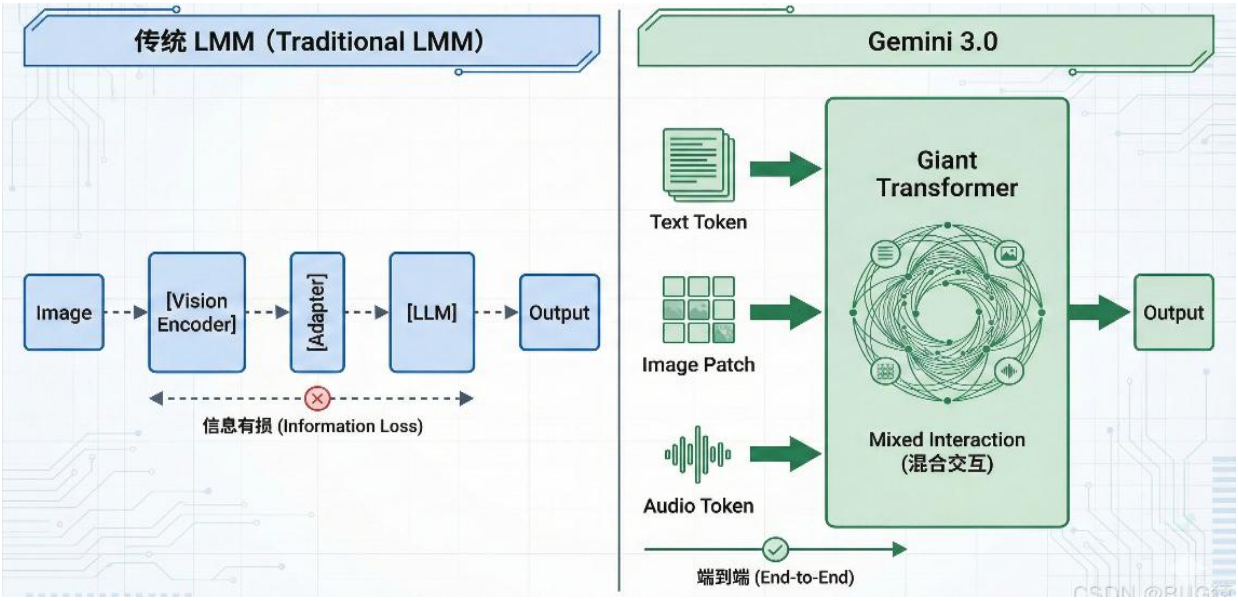


资料来源：Artificial Analysis 官网，国元证券研究所

多模态技术不断突破，原生多模态架构成为标配，成为 AI 理解真实世界的基石。2025 年多模态技术的核心突破，在于架构层面实现了从“拼接式多模态”(Stitched Multimodality)向“原生多模态”(Native Multimodality)的根本性演进。CLIP 和 DALL-E 的早期版本通常采用多个独立的、针对特定模态的编码器(Encoder)。例如，使用一个预训练好的视觉模型（如 ViT）来编码图像，使用一个语言模型（如 BERT）来编码文本，然后通过一个轻量级的“连接层”(Projection Layer)将它们的特征向量映射到同一个语义空间进行对齐和融合。这种方式虽然在当时取得了不错的效果，但存在跨模态理解不够精细、扩展性差等技术缺陷。以 Google Gemini 系列、OpenAI GPT-5 以及国内的通义千问 Qwen2.5-VL 为代表的新一代原生多模态模型，在架构层面就实现了根本性的统一。它们采用统一的 Transformer 架构和共享的向量空间

来处理所有模态的数据。其核心思想是“万物皆可 Token 化”，无论是文本、图像、声音还是视频，都会被一个统一的“分词器”(Tokenizer)或多个协同工作的分词器，转换成一系列离散的“语义令牌”(Semantic Tokens)，并采用端到端深度融合训练，在统一的 Transformer 架构中，来自不同模态的 Token 通过自注意力机制(Self-Attention)进行无差别的、深度的交互和融合。

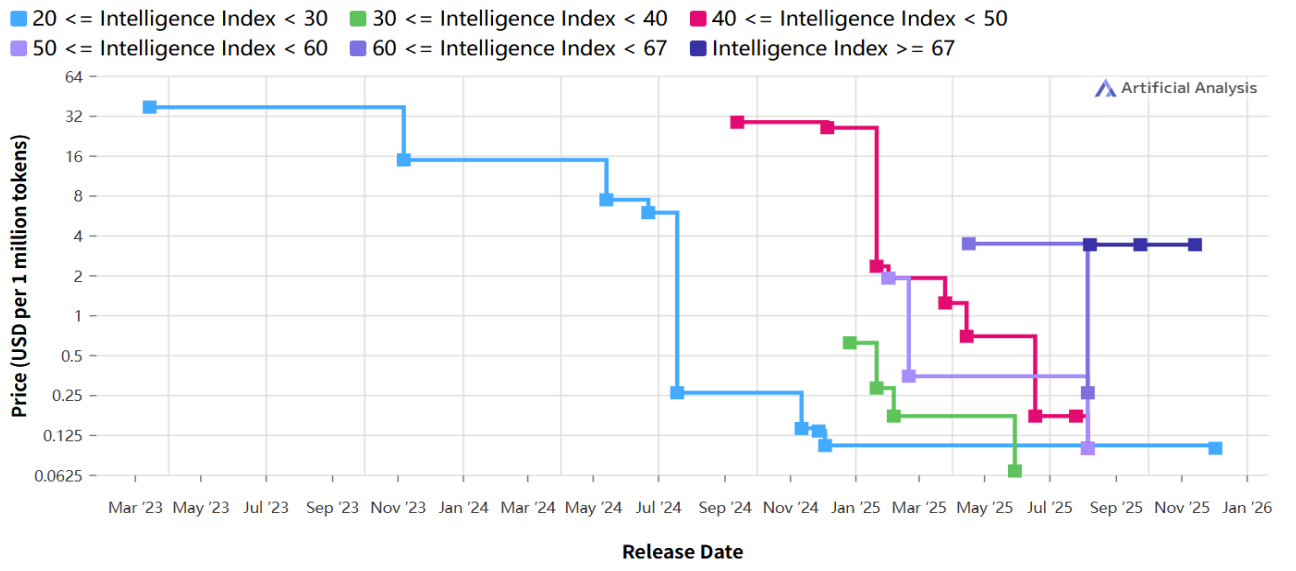
图 9：原生多模态架构与传统“拼接式”架构的区别



资料来源：技术栈官网，国元证券研究所

根据 Artificial Analysis 的数据，2025 年以来，除了最新发布的几款前沿指数模型以外，其余的大模型推理成本出现成倍数下降。当推理变得便宜时，用户开始执行更为复杂的任务，带来 token 消耗数量的倍数增长。

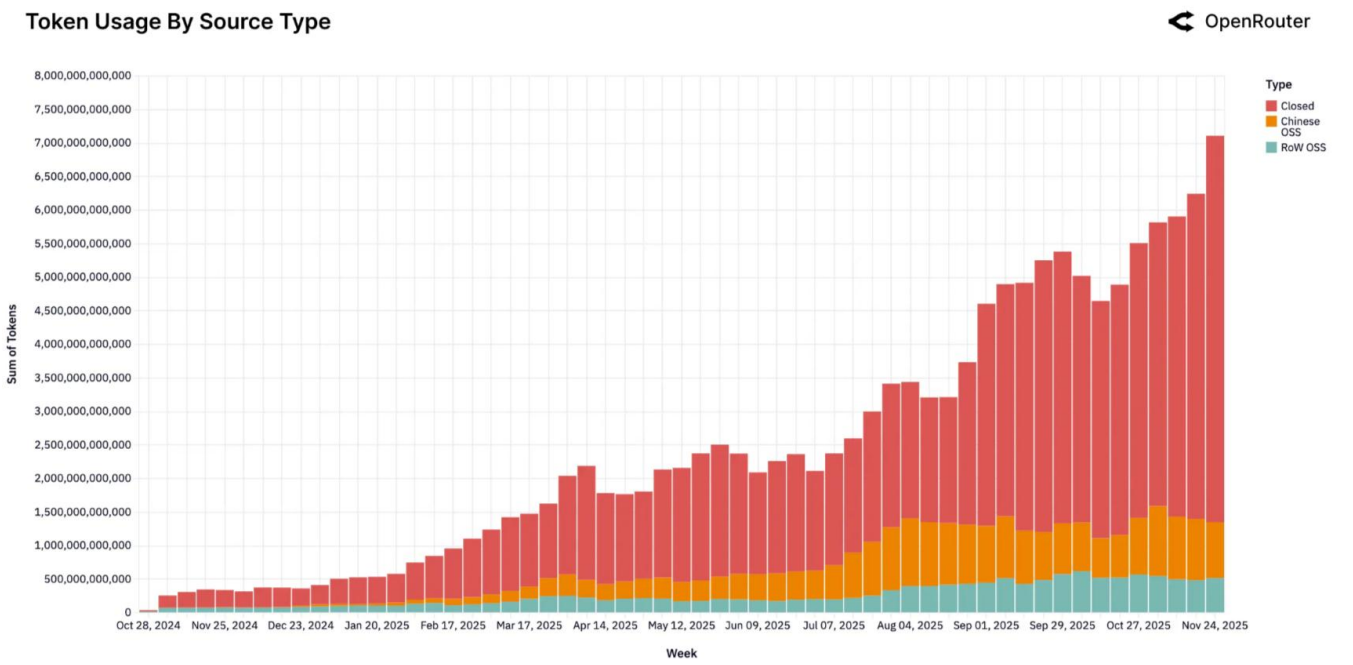
图 10：2025 年大模型推理价格不断降低



资料来源：Artificial Analysis 官网，国元证券研究所

从 token 消耗量来看，AI 模型需求端依旧旺盛。根据知名第三方模型 API 提供商 OpenRouter 的统计数据，通过该平台调用各类模型的 token 消耗量在 2025 年增长约 14 倍，近三个月实现翻倍增长。

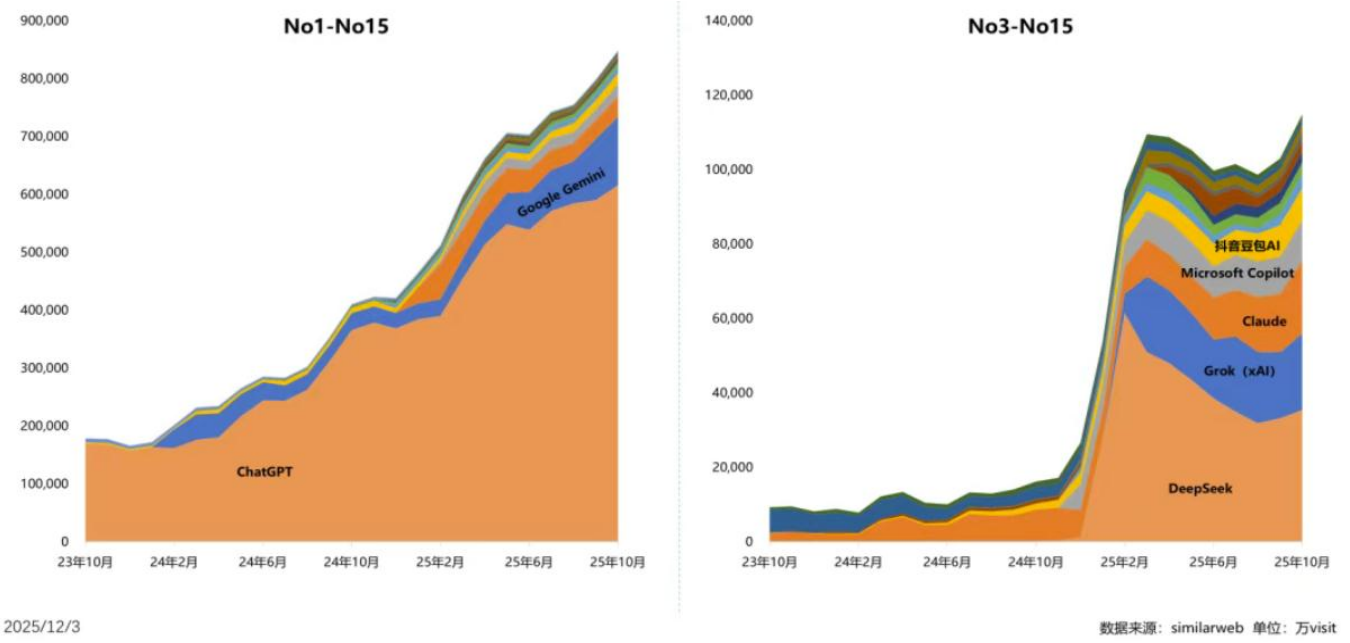
图 11：各类模型每周的 token 消耗量快速增长



资料来源：OpenRouter 官网，国元证券研究所

从整体访问流量来看，AI 赛道增长斜率依然陡峭。2025 年 10 月，AI Chatbots 赛道前十五名产品网页总访问量达 84.9 亿次，全领域总访问量为 85.8 亿次，ChatGPT 以 61.6 亿次访问量稳居第一，Google Gemini 以 11.8 亿次访问量排名第二，中国的 DeepSeek 以 3.55 亿次访问量位列第三。此外，还有五款中国产品进入前十五名，分别是抖音豆包 AI（第 7 名）、Qwen（第 9 名）、Kimi.com（第 10 名）、阿里通义千问（第 11 名）和腾讯元宝（第 13 名）。

图 12：AI Chatbots 产品网页流量增长趋势

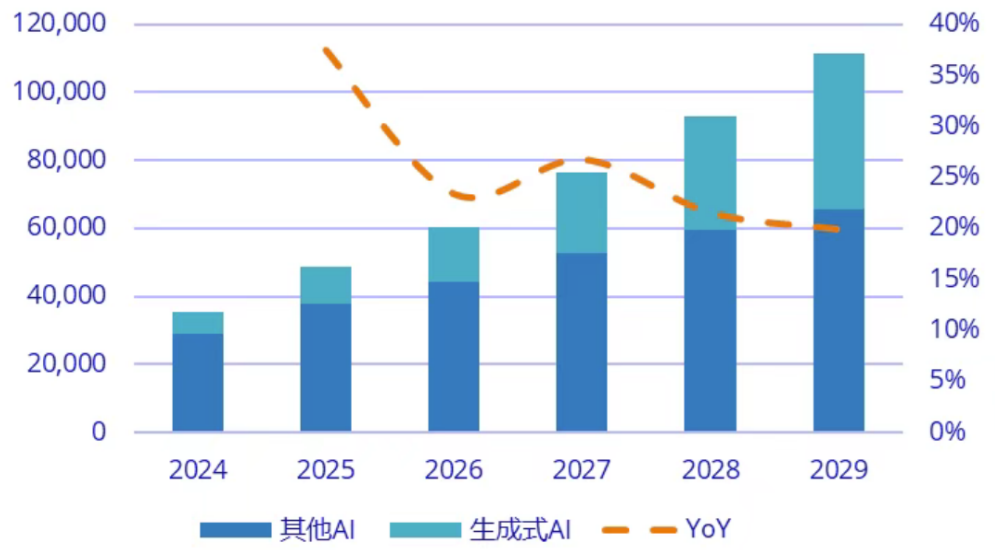


资料来源：similarweb，郎瀚威 Will 公众号，国元证券研究所

根据 IDC 发布的《全球人工智能和生成式人工智能支出指南》，2024 年全球人工智能 IT 总投资规模为 3159 亿美元，并有望在 2029 年增至 12619 亿美元，五年复合增长率为 31.9%。生成式 AI 的增长尤为迅猛，成为引领本轮 AI 浪潮的绝对主力，IDC 预测，全球生成式 AI 市场五年复合增长率或达 56.3%，到 2029 年全球生成式 AI 市场规模将达 6071 亿美元，占 AI 市场投资总规模的 48.1%。

从地区分布来看，中美两国作为全球 AI 产业的两大主导力量，市场规模合计占比已接近七成。根据中国信息通信研究院的数据，截至 2025 年 9 月，中国 AI 核心产业规模已突破 9000 亿元人民币，约占全球核心产业规模的 10%，相关企业数量超过 5300 家。IDC 预测，到 2029 年，中国在 AI 领域的总投资规模将达到 1114 亿美元，五年复合增长率为 25.7%，增速持续领先全球主要经济体。具体而言，2024 年中国生成式 AI 占 AI 市场投资总规模的 18.2%。随着生成式 AI 技术的快速发展，2029 年生成式 AI 投资占比将达到 41.1%，投资规模超 450 亿美元，五年复合增长率为 48.0%。

图 13：2024-2029 年中国 AI 与生成式 AI 市场规模预测（单位：百万美元）

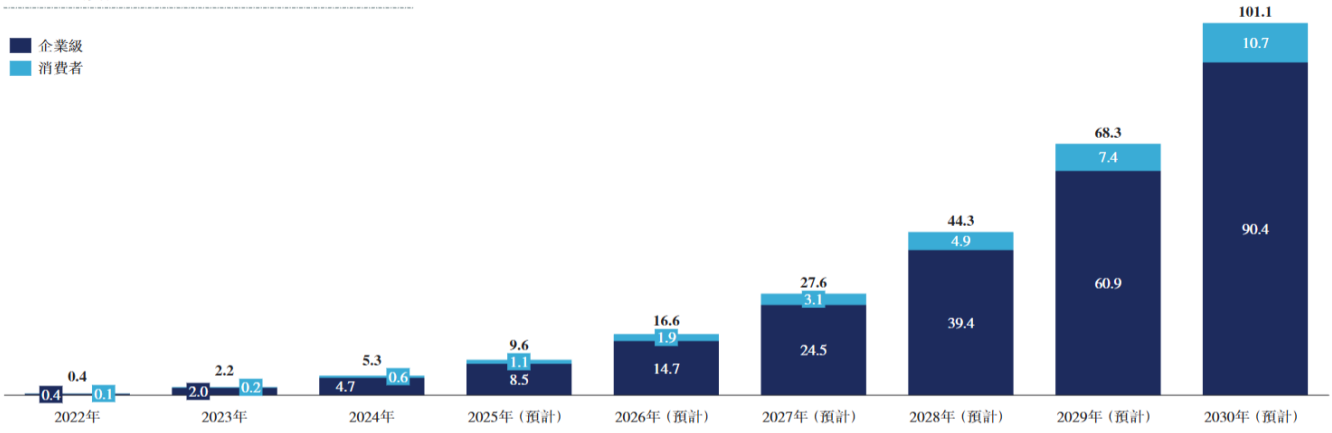


资料来源：IDC 公众号，国元证券研究所

根据智谱港股招股书，2024 年中国大语言模型市场规模（以收入计）达 53 亿元，其中机构客户贡献 47 亿元，个人客户贡献 6 亿元；2025 年，中国大语言模型市场规模将达 96 亿元，较上年增长约 81%。随着大语言模型技术进步与用户需求增长，预计 2030 年市场规模将增至 1011 亿元，2024-2030 年复合年增长率为 63.5%。机构客户仍是核心驱动力，预计 2030 年中国企业级大语言模型市场规模将达 904 亿元，2024-2030 年复合年增长率为 63.7%。

图 14：2022-2030 年中国大语言模型市场规模及预测（单位：十亿元）

複合年增長率	2022年至2024年	2024年至2030年（預計）
總計	264.0%	63.5%
企業級	261.3%	63.7%
消費者	192.8%	61.6%

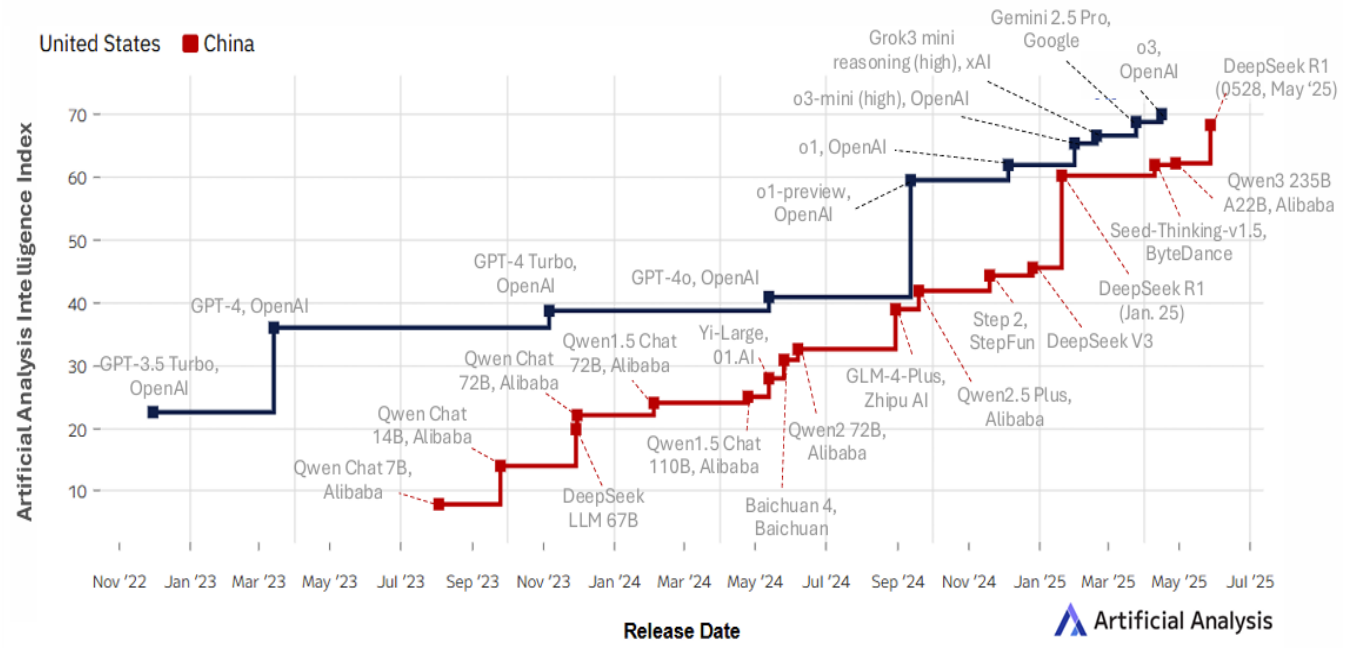


资料来源：智谱港股招股书，国元证券研究所

美国顶尖大模型依然领先，但随着 DeepSeek R1、Qwen3、Kimi K2 等一批国产优秀模型的发布，中美 AI 的差距从几年缩短至几个月。根据海外第三方人工智能研究

机构 Artificial Analysis 编制的 Artificial Analysis Intelligence Index（包含 MMLU-Pro, GPQA Diamond, Humanity's Last Exam, LiveCodeBench, SciCode, AIME, MATH-500 等七项模型基准），2025 年 5 月发布的 DeepSeek R1 虽晚于 OpenAI o3，但模型能力已经接近。

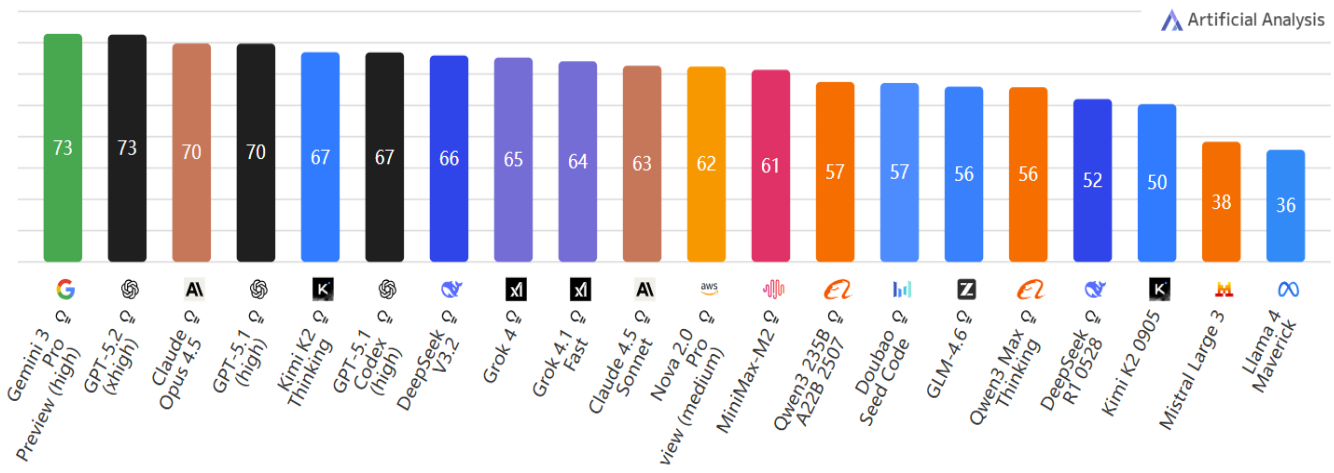
图 15：中美前沿模型差距逐步缩小



资料来源：《Artificial Analysis State of AI China Q2 2025 Highlights》，国元证券研究所

与此同时，前沿大模型竞争趋向白热化，OpenAI 的领先地位正受到来自多方的挑战，Google、Anthropic、XAI 等头部玩家不断推出更强大的模型。海外第三方人工智能研究机构 Artificial Analysis 根据 MMLU-Pro、GPQA Diamond、Humanity's Last Exam、LiveCodeBench、SciCode、AIME 2025、IFBench、AA-LCR、Terminal-Bench Hard、τ²-Bench 等多项模型基准，编制了 AI 分析指数(Artificial Analysis Intelligence Index)，Gemini 3 pro 与 GPT 5.2 并列第一，Claude Opus 4.5 与 GPT5.1 紧随其后，得分都为 70，国产模型 Kimi K2 Thinking 排名第五，DeepSeek v3.2、Minimax M2、Qwen3 235B 等模型也排名靠前。

图 16：主流大模型的 AI 分析指数(Artificial Analysis Intelligence Index)排名

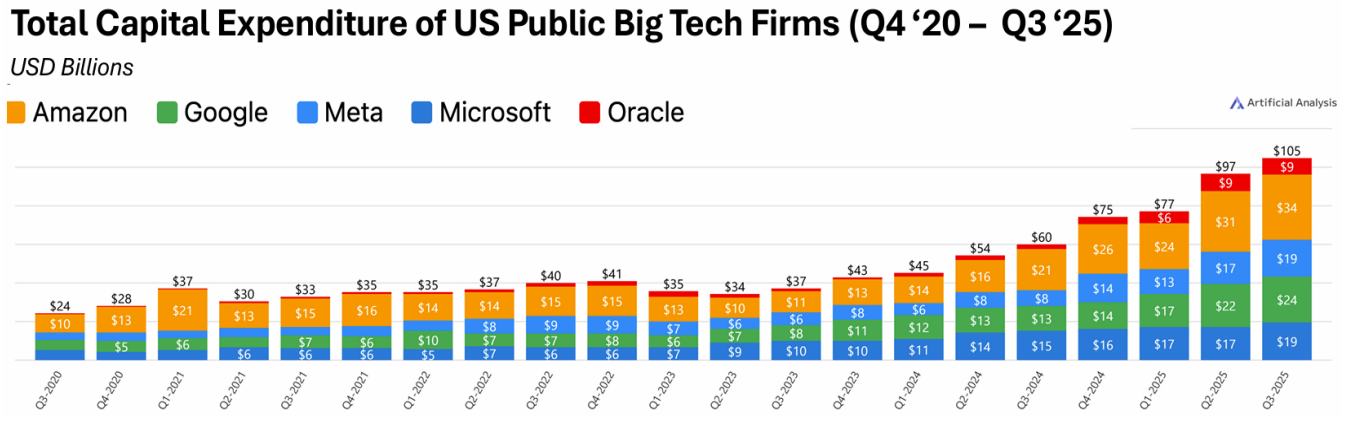


资料来源：Artificial Analysis，国元证券研究所

2.1 海外巨头持续投入，引领人工智能发展

美国科技巨头持续扩大资本开支，押注人工智能的发展。根据 Artificial Analysis 发布的《Q3 2025 Artificial Analysis State of AI Highlights Report》，2025 年以来，Amazon、Google、Meta、Microsoft 和 Oracle 的资本开支总和逐季度快速增长。

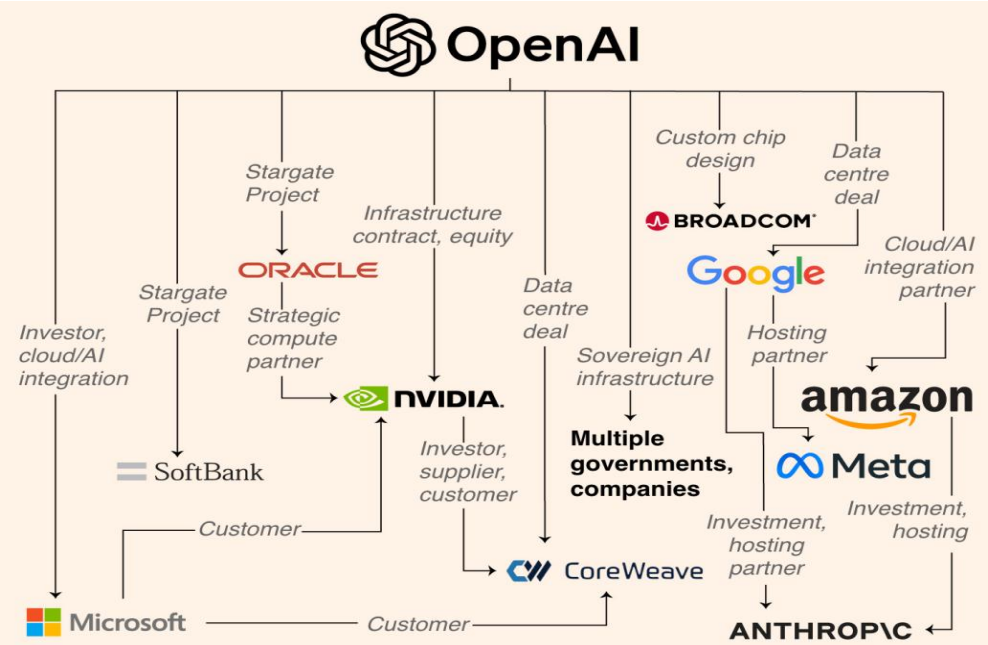
图 17：美国科技巨头持续加大 AI 投入



资料来源：《Q3 2025 Artificial Analysis State of AI Highlights Report》，国元证券研究所

此外，XAI、OpenAI 等非上市科技巨头也在快速扩大资本开支。OpenAI 迄今已承诺投入约 1.4 万亿美元用于基础设施建设，相当于约 30 吉瓦的数据中心容量，包括此前已宣布的 OpenAI 与 AMD、博通、英伟达、甲骨文等合作伙伴的交易。其中，甲骨文 OCI 合约总价 3000 亿美元，从 2027 年开始为期 5 年，年均 600 亿美元；微软 Azure 合约总价 2500 亿美元，从 2026 年中开始为期 7 年，年均 360 亿美元；谷歌 GCP 合约总价 400 亿美元，为期 7 年，亚马逊 AWS 提供 380 亿美元、为期 7 年的合约。

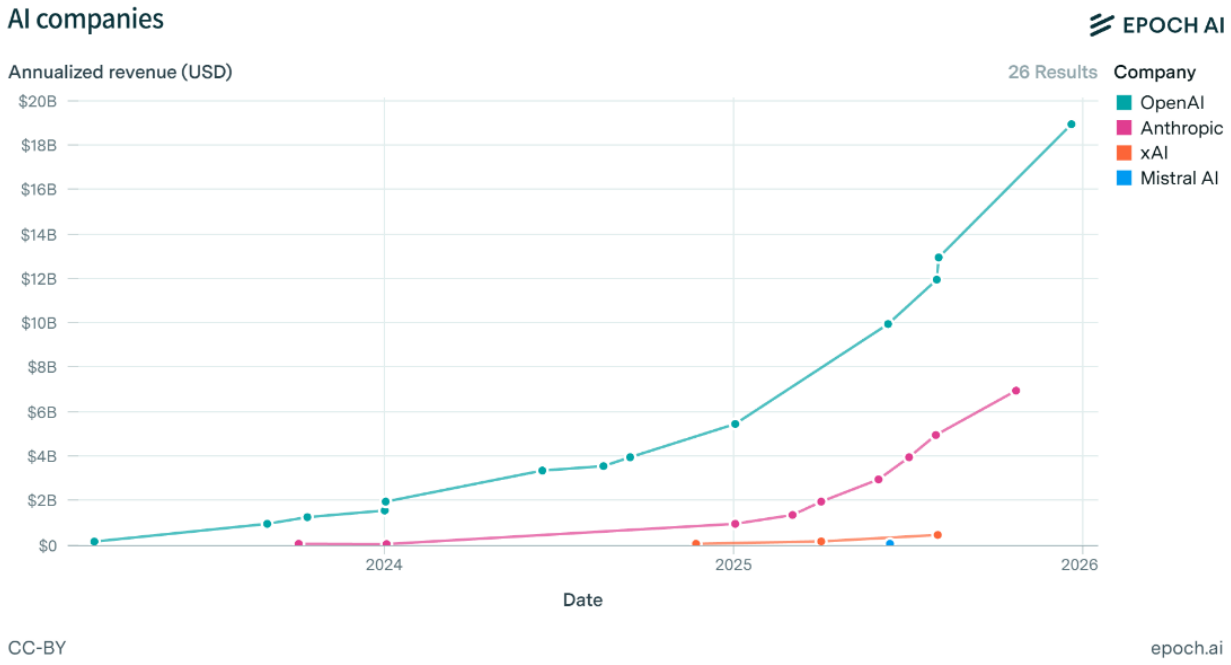
图 18：OpenAI 的算力版图



资料来源：APPSO 公众号，国元证券研究所

从商业化方面来看，随着模型性能的快速提升与推理成本的下降，OpenAI、Anthropic、xAI 等海外 AI 巨头年化收入呈现快速增长趋势。根据 epoch.ai 的数据，OpenAI 年化收入即将突破 200 亿美元，较年初增长约 4 倍，Anthropic 年化收入约 80 亿美元，较年初增长约 7 倍。

图 19：海外 AI 厂商年化收入快速增长



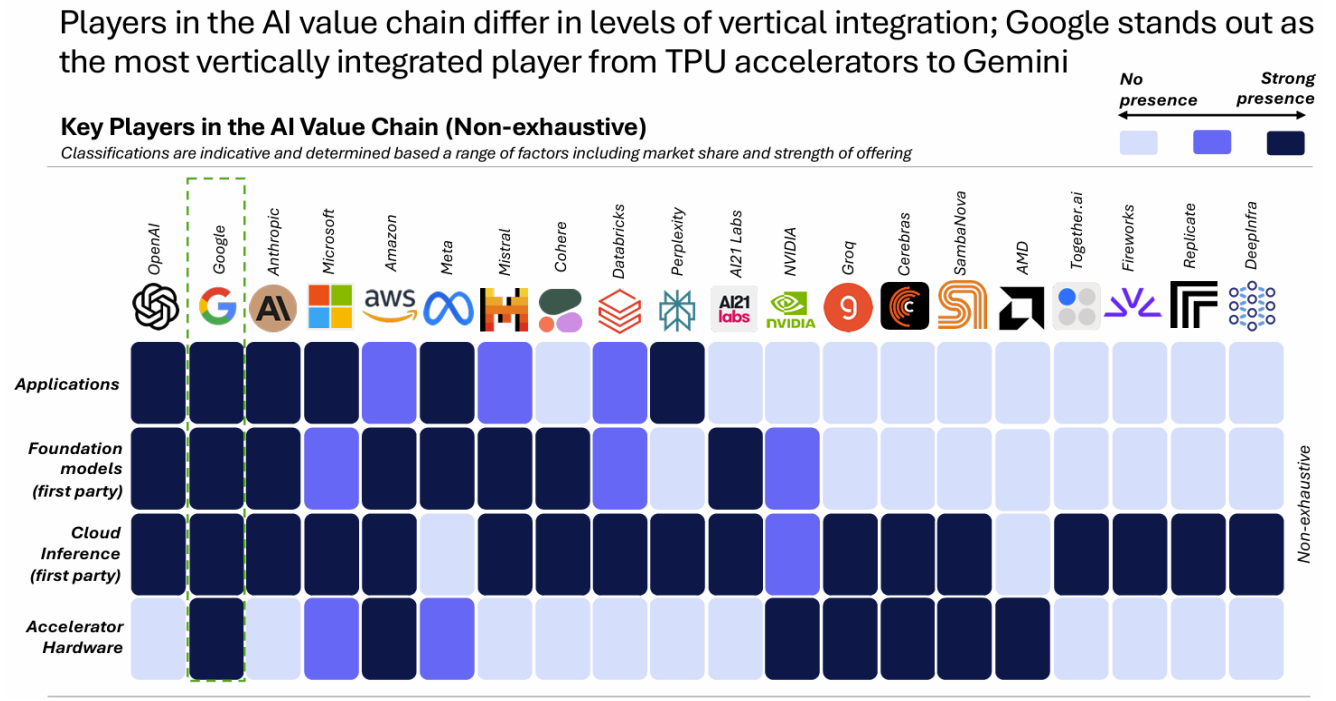
资料来源：epoch.ai 官网，国元证券研究所

2.1.1 谷歌

谷歌以 Gemini 系列大模型为核心，持续迭代推进纯文本、多模态及代码等能力，打造全栈 AI 战略。谷歌 CEO Sundar Pichai 在 Gemini 3 发布时表示，谷歌在 AI 创新上有一套与众不同的全栈做法——从领先的基础设施，到世界一流的研究、模型和工具，再到触达全球数十亿人的产品——以前所未有的速度，把先进的能力带给全世界。进入 2025 年，谷歌连续发布了 Gemini 2.0 Pro、Gemini 2.5 Pro、Nanobanana Pro、Veo3.1、Gemini 3 Flash 等一系列模型，在模型参数、实战能力、用户反馈上都实现了跃迁，重新建立起在 AI 领域的领导地位。

谷歌是极少数具备从底层芯片(TPU)、到数据中心基础设施、再到中间层模型(Gemini)、以及最上层应用(Search、YouTube、Waymo)的高度一体化与可控性的 AI 厂商。

图 20：谷歌是极少数具备全栈 AI 能力的厂商



资料来源：《Artificial Analysis AI Review 2024 Highlights》，国元证券研究所

在芯片层，谷歌自研的 TPU 芯片支撑了 Gemini 2.0 多模态模型的训练和推理，TPU 的高效能使谷歌能够以相对低成本训练大规模模型，使模型迭代周期更短、成本更低。2025 年发布的 TPU v7 Ironwood 作为新一代产品，在浮点运算性能、内存及带宽这三项核心指标上，几乎已完全追平同期英伟达的旗舰级 GPU，相较于 GB200，TPUv7 的浮点运算性能与内存带宽仅存在小幅差距，二者的内存容量处于同一水平，均搭载 8 层高带宽内存第三代增强版(8-Hi HBM3E)。

12 月 18 日，谷歌最新发布高速、低成本的推理模型 Gemini 3 Flash，旨在用更低的成本提供前沿的智能水平。具体来看，其输出每百万 token 的价格仅 Claude Sonnet 4.5 的 20%、GPT-5.2 的 21%，却能在基准测试上达到甚至超越这些旗舰级模型的

水平，例如在多模态推理基准 MMMU Pro 上，Gemini 3 Flash 以 81.2%的高分达到了当前最先进水平；与 Gemini 3 Pro 相比，其价格仅为 Pro 的 25%，却在 MMMU-Pro、SWE-bench Verified 这些核心基准测试上更强。

Gemini 3 Flash 的设计以极高的效率为目标，能够在保证准确性的同时，以更低的成本和更高的效率完成日常任务。在最高思考等级下运行时，Gemini 3 Flash 能够动态调节自身的思考深度：面对更复杂的使用场景，它会投入更长时间进行推理；而在处理日常任务时，则能以更高性能完成目标，同时在典型业务流量下，平均使用的 token 数量比 Gemini 2.5 Pro 减少约 30%。目前，Gemini 3 Flash 已取代 2.5 Flash 成为 Gemini 应用的默认模型，全球 Gemini 用户都已能免费体验 Gemini 3，从而大幅提升日常任务的处理效率。

图 21：Gemini 3.0 Flash 以更低价格实现高性能

Benchmark	Description		Gemini 3 Flash Thinking	Gemini 3 Pro Thinking	Gemini 2.5 Flash Thinking	Gemini 2.5 Pro Thinking	Claude Sonnet 4.5 Thinking	GPT-5.2 Extra high	Grok 4.1 Fast Reasoning
Input price		$/1M tokens	$0.50	$2.00 $4.00 > 200k tokens	$0.30	$1.25 $2.50 > 200k tokens	$3.00 $6.00 / MTok > 200k tokens	$1.75	$0.20
Output price		$/1M tokens	$3.00	$12.00 $18.00 > 200k tokens	$2.50	$10.00 $15.00 > 200k tokens	$15.00 $22.50 > 200k tokens	$14.00	$0.50
Humanity's Last Exam	Academic reasoning (full set, text + MM)	No tools With search and code execution	33.7% 43.5%	**37.5%** **45.8%**	11.0% —	21.6% —	13.7% —	34.5% 45.5%	17.6% —
ARC-AGI-2	Visual reasoning puzzles	ARC Prize Verified	33.6%	31.1%	2.5%	4.9%	13.6%	**52.9%**	—
GPQA Diamond	Scientific knowledge	No tools	90.4%	91.9%	82.8%	86.4%	83.4%	**92.4%**	84.3%
AIME 2025	Mathematics	No tools With code execution	95.2% 99.7%	95.0% **100%**	72.0% 75.7%	88.0% —	87.0% **100%**	**100%** —	91.9% —
MMMU-Pro	Multimodal understanding and reasoning		**81.2%**	81.0%	66.7%	68.0%	68.0%	79.5%	63.0%
ScreenSpot-Pro	Screen understanding	No tools unless specified	69.1%	72.7%	3.9%	11.4%	36.2%	**86.3%** with python	—
CharXiv Reasoning	Information synthesis from complex charts	No tools	80.3%	81.4%	63.7%	69.6%	68.5%	**82.1%**	—
OmniDocBench 1.5	OCR	Overall Edit Distance, lower is better	0.121	**0.115**	0.154	0.145	0.145	0.143	—
Video-MMMU	Knowledge acquisition from videos		86.9%	**87.6%**	79.2%	83.6%	77.8%	85.9%	—
LiveCodeBench Pro	Competitive coding problems from Codeforces, ICPC, and IOI	Elo rating, higher is better	2316	**2439**	1143	1775	1418	2393	—
Terminal-bench 2.0	Agentic terminal coding	Terminus-2 harness	47.6%	**54.2%**	16.9%	32.6%	42.8%	—	—
SWE-bench Verified	Agentic coding	Single attempt	78.0%	76.2%	60.4%	59.6%	77.2%	**80.0%**	50.6%
τ2-bench	Agentic tool use		90.2%	**90.7%**	79.5%	77.8%	87.2%	—	—
Toolathlon	Long horizon real-world software tasks		**49.4%**	36.4%	3.7%	10.5%	38.9%	46.3%	—
MCP Atlas	Multi-step workflows using MCP		57.4%	54.1%	3.4%	8.8%	43.8%	**60.6%**	—
Vending-Bench 2	Agentic long term coherence	Net worth (mean), higher is better	$3,635	**$5,478**	$549	$574	$3,839	$3,952	$1,107
FACTS Benchmark Suite	Factuality benchmark across grounding, parametric, search, and MM		61.9%	**70.5%**	50.4%	63.4%	48.9%	61.4%	42.1%
SimpleQA Verified	Parametric knowledge		68.7%	**72.1%**	28.1%	54.5%	29.3%	38.0%	19.5%
MMMLU	Multilingual Q&A		**91.8%**	**91.8%**	86.6%	89.5%	89.1%	89.6%	86.8%
Global PIQA	Commonsense reasoning across 100 Languages and Cultures		92.8%	**93.4%**	90.2%	91.5%	90.1%	91.2%	85.6%

资料来源：Google 官网，国元证券研究所

在应用层，谷歌正在将 AI 能力深度整合到 Search、Gmail 等多个平台中，Gemini 不再只是一个独立的聊天机器人或 API，而是成为贯穿整个生态的基础设施，从而充分发挥其庞大的用户规模优势，打造商业闭环。在 C 端，搜索中的 AI 模式和 AI Overview 已逐步切换至由 Gemini 3 提供支持；Gemini 应用以及 Android 系统中的部分智能功能，也以这一代模型为主。在 B 端，Workspace 中的写作助手、文档摘要、演示文稿生成等功能，统一基于 Gemini 模型系列。在云端，Vertex AI 和 Gemini API 面向开发者开放同一组模型，既服务于聊天应用，也支持代码、搜索和数据处理等多种场景。根据谷歌 2025Q3 财报，Gemini APP 月活已达 6.5 亿，较二季度的 4.5 亿大幅增长。

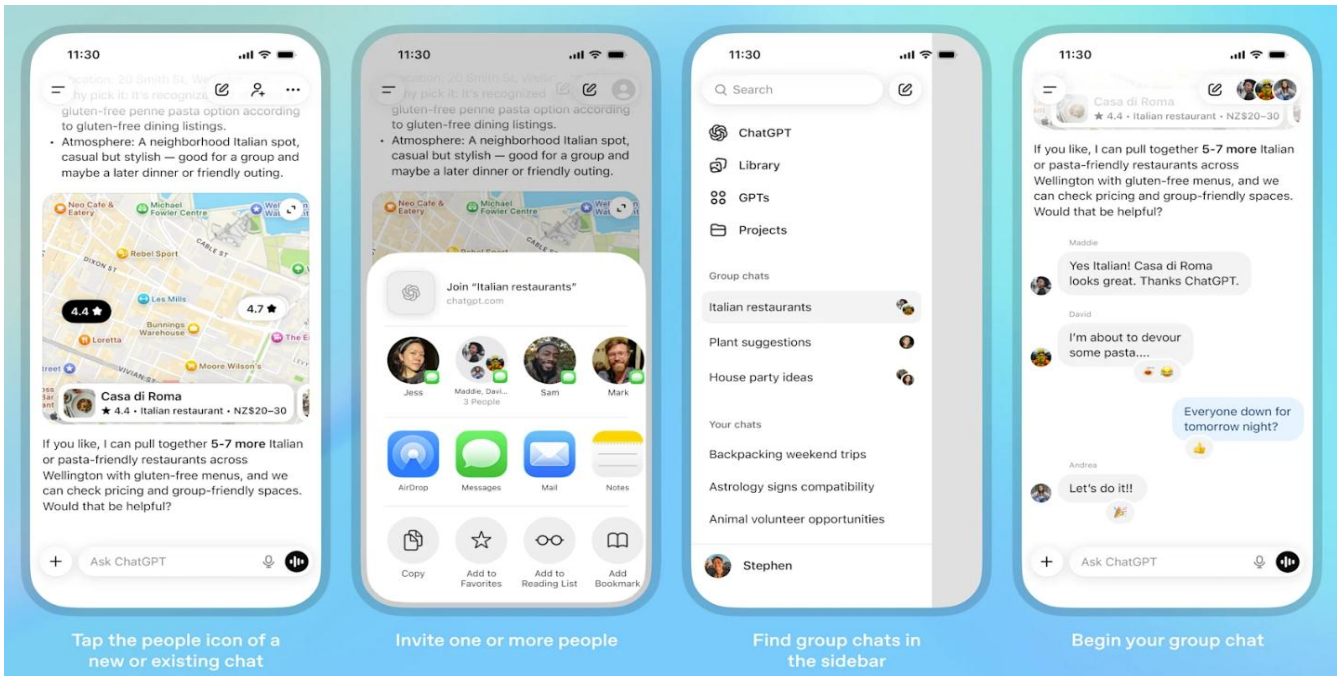
2.1.2 OpenAI

进入 2025 年，OpenAI 推出了 GPT-5、GPT-5.1、GPT-5.2 和 Sora 2 等一系列先进模型，并发布了 Operator、Deep Research 和 ChatGPT Agent 等 Agent 系统，推动 AI 向生产力场景的渗透转型。

OpenAI CEO Sam Altman 在 2025 年 10 月阐述了 OpenAI 的战略转型：过去的产品形态主要是 ChatGPT 这样的 AI 超级助手，现在正朝着更宏大的目标演进：成为一个平台，一个其他人可以在其上构建应用和服务的“AI 云”。为了支撑这个平台，OpenAI 正构建一整套 AI 交互系统，并逐步开放接口：网页、浏览器 Atlas、移动设备、应用市场、插件生态、企业平台。企业可以通过 API 接入 OpenAI 技术，构建自己的应用，开发者可以在 ChatGPT 插件平台里创建“应用型智能体”，未来还会有全新的硬件设备形态，让 AI 不再只活在网页里，而是随时随地为用户服务。

在 C 端领域，OpenAI 推出群组聊天、ChatGPT Pulse、即时结账功能等多项个性化服务功能，从工具助手转向“all in one”的生态平台。例如，ChatGPT 于 11 月上线群聊功能，让用户们和 ChatGPT 能够聚在同一个对话里互动协作，当用户正在与朋友规划周末旅行，创建一个群组聊天，这样 ChatGPT 就可以协助比较目的地、制定行程，并创建一个行李清单。此外，ChatGPT 推出“即时结账”(Instant Checkout)的新功能，用户可在与 AI 的对话中直接下单，享受跨支付平台的灵活性，表明 OpenAI 正为下一代 AI 驱动的商业做好准备。此项功能由 OpenAI 和 Stripe 联合构建的开放标准——代理商业协议(Agentic Commerce Protocol, ACP)提供支持，让商家和开发者能够简单地与消费者建立连接，已支持 Etsy 等平台的商家。

图 22：ChatGPT 上线群聊功能



资料来源：OpenAI 官网，国元证券研究所

8 月 8 日，OpenAI 发布 GPT-5 系列模型，将原有的 GPT 和 o 系列模型纳入统一系统，从单一、庞大的模型转向一个更加动态、异构和响应迅速的生态系统。GPT-5 系统包含一个能够解答大多数问题的智能快速模型、一个能够解决更复杂问题的深度推理模型，以及一个实时路由器，可以根据对话类型、复杂性、工具需求和明确意图快速决定使用哪个模型，这样路由框架掌握了“谁来答、答到什么深度、何时调用工具”的主动权，能够在体验与成本之间做实时最优分配。

12 月 12 日，OpenAI 发布最新系列模型 GPT-5.2，包含 GPT-5.2 Instant、GPT-5.2 Thinking 和 GPT-5.2 Pro，GPT-5.2 Thinking 主打专业工作，具备最先进的长上下文推理能力，GPT-5.2 Instant 专为日常学习和工作设计，GPT-5.2 Pro 则是最聪明强大的版本，在编程等复杂领域表现更强。GPT-5.2 的设计初衷在于“创造更大的经济价值”。相较前代，它在电子表格处理、演示文稿制作、代码编写、图像感知、长文本理解及复杂多步项目执行等方面，均实现了全面性能跃升。为了验证其在真实业务环境中的价值，OpenAI 采用了 GDPval 基准测试，该测试覆盖了 9 大行业、44 类职业的 1320 个真实业务场景，GPT-5.2 Pro 在高达 74.1%的任务中表现超越或持平人类专家。

在智能体任务方面，GPT-5.2 Thinking 在 Tau2-bench Telecom 多轮客服任务的工具调用测试中取得 98.7%完成度，该评测覆盖订票、退款、延误补偿、物品遗失与跨系统调度等场景，反映其可承担更高自治度的流程任务。OpenAI 表示，GPT-5.2 的“工具决策粒度”更加稳定，推理链条更可控，在连续 20~40 步的任务中不易出现跳步、误调用或不必要调用等问题，使其更适合作为长流程自动化智能体(Autonomous Agent)的执行核心。

在涉及编程、数学、复杂推理等领域的基准测试中，GPT-5.2 Thinking 较上一代模型显著提升。例如，AIME 2025（竞赛数学）：GPT-5.2 Thinking 达 100%满分，Claude Opus 4.5 为 92.8%，Gemini 3 Pro 为 95.0%。与此同时，GPT-5.2 Thinking 事实准确率显著提升，在面对模糊或信息不完整的查询时，会更主动给出依据来源或使用结构化推理路径，以降低误导性回答的概率。价格方面，GPT-5.2 较上一代有所增加，GPT-5.2 Pro 的输入价格为 21 美元/百万 tokens，输出价格为 168 美元/百万 tokens。

图 23：GPT-5.2 的基准测试结果较上一代模型显著提升

	OpenAI	Run with maximum available reasoning effort.	Anthropic	Google
	GPT-5.2 Thinking	GPT-5.1 Thinking	Claude Opus 4.5	Gemini 3 Pro
SWE-Bench Pro Software engineering	• 55.6%	50.8%	52.0%	43.3%
GPQA Diamond Science questions (No tools)	• 92.4%	88.1%	87.0%	91.9%
CharXiv Reasoning Scientific figure questions (No tools)	• 82.1%	67.0%	—	81.4%
FrontierMath Advanced mathematics (Tier 1 - 3, Tier 4)	• 40.3% 14.6%	31.0% 12.5%	— —	37.6% • 18.8%
AIME 2025 Competition math (No tools)	• 100.0%	94.0%	92.8%	95.0%
ARC-AGI-1 Abstract reasoning	• 86.2%	72.8%	80.0%	75.0%
ARC-AGI-2 Abstract reasoning	• 52.9%	17.6%	37.6%	31.1%
GDPval Knowledge work tasks	• 70.9%	38.8% GPT-5	59.6%	53.5%

资料来源：腾讯科技公众号，国元证券研究所

2.2 国内企业成果显著，持续推动应用落地

2.2.1 智谱

智谱成立于 2019 年，作为中国领先的人工智能公司，致力于开发先进的通用大模型。公司在全方位的人工智能研究中扎实交付先进技术，并稳步扩大其商业应用，以实现收入的快速增长。2021 年，公司发布了中国首个专有预训练大模型框架 GLM 框架，并推出了模型即服务(MaaS)产品开发及商业化平台，通过该平台提供大模型服务。2022 年，公司开源首个 1000 亿规模的模型(GLM-130B)。公司运营的大语言模型(LLM)市场为更广泛 AI 市场的一个细分领域，为机构客户（包括私营企业及公共行业实体）及个人用户（包括个人终端用户及独立开发者）提供通用大模型服务。截至 2025 年 6 月 30 日，公司的模型已为逾八千家机构客户提供支持。根据公司港股招股书，2024 年公司实现营业收入 3.12 亿元，在中国独立通用大模型开发商中位列第一，在所有通用大模型开发商中位列第二。

图 24：智谱是中国领先的独立通用大模型开发商



资料来源：智谱港股招股书，国元证券研究所

智谱的模型在单体模型能力方面同样实现了行业性能领先，部分模型达到全球最先进(SOTA)水平：

1）语言模型——高级智能的可靠基础。智谱的旗舰模型 GLM-4.5 的功能可与全球最先进模型比肩。根据弗若斯特沙利文的资料，依据 2025 年 7 月进行的 12 项业界标准基准测试的评估结果，GLM-4.5 在全球排名第三、在中国跃居第一及全球开源模型位居榜首。首发后仅 48 小时内，GLM-4.5 便在 Hugging Face（全球最大开源模型平台）的热门榜上跃居全球榜首。

2）智能体模型——自动执行复杂任务和设备控制。智谱的智能体模型旨在为通用智能体提供支持，该等智能体能够代表人类理解、思考和执行复杂的任务。此性能基于两个核心模型构建：GLM-Z1-Rumination，作为智能体进行深度推理和自主规划的"大脑"；及 AutoGLM，提供感知和操作用户界面的"双手"。AutoGLM 在 AgentBench（斯坦福大学《2024 年 AI 指数》报告认证的智能体 AI 基准）中实现 SOTA 性能。

3）多模态模型——重新定义内容创作。在文生图生成方面，CogView4 不仅在 DPG-Bench 基准测试中排名第一，也是全球首款能够准确理解并生成中文文本的开源模型。就视频生成而言，CogVideoX 实现了顶级性能，在全面的 SuperCLUE-I2V 基准测试中名列前茅。GLM-4V 是中国首个开源双语多模态对话模型，GLM-4-Voice 是中国首个端到端超拟人语音模型，二者进一步巩固了公司全面的多模态领导地位，展示了公司在所有主要模态方面的开创性能力。

4）代码模型——转换编码体验。公司于 2023 年发布并持续迭代的 CodeGeeX 代码模型性能卓越，被程序员群体广泛使用。截至 2025 年 6 月 30 日，CodeGeeX 平均每天生成逾 100 百万行代码。

图 25：智谱大模型及智能体

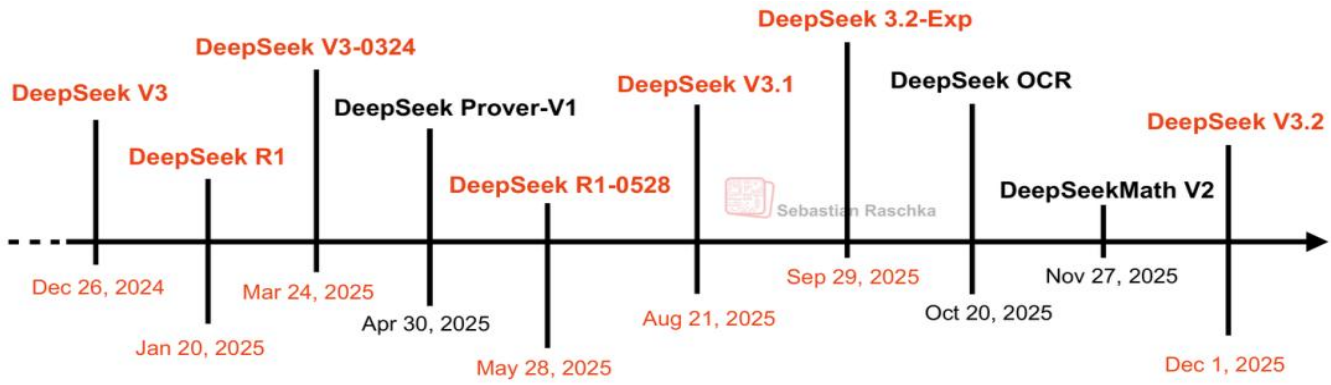


资料来源：智谱港股招股书，国元证券研究所

2.2.2 DeepSeek

从年初的推理模型 DeepSeek-R1 演进到年末的 DeepSeek-V3.2 系列，DeepSeek 的系列模型不仅在推理能力上比肩国际顶尖模型，更在智能体和推理成本控制上实现了跨越式的突破。

图 26：DeepSeek 模型发布时间线



资料来源：机器之心公众号，国元证券研究所

1 月 20 日，DeepSeek-R1 正式发布，在后训练阶段大规模使用了强化学习技术，在

仅有极少标注数据的情况下，极大提升了模型推理能力，在数学、代码、自然语言推理等任务上，性能比肩 OpenAI o1 正式版。该模型最重要的技术突破是用纯深度学习的方法让 AI 自发涌现出推理能力，对模型推理训练后续的范式产生深刻影响。

5 月，DeepSeek R1 完成小版本升级为 DeepSeek-R1-0528，深度思考能力强化。该模型仍然使用 2024 年 12 月所发布的 DeepSeek V3 Base 模型作为基座，但在后训练过程中投入了更多算力，显著提升了模型的思维深度与推理能力。更新后的 R1 模型在数学、编程与通用逻辑等多个基准测评中取得了当前国内所有模型中首屈一指的优异成绩，并且在整体表现上已接近其他国际顶尖模型，如 o3 与 Gemini-2.5-Pro。

图 27：DeepSeek-R1-0528 在各项评测集上均取得了优异表现

Benchmarks	DeepSeek-R1-0528	OpenAI-o3	Gemini-2.5-Pro-0506	Qwen3-235B	DeepSeek-R1
AIME 2024 数学竞赛 pass@1	91.4	91.6	90.8	85.7	79.8
AIME 2025 数学竞赛 pass@1	87.5	88.9	83.0	81.5	70.0
GPQA Diamond 科学测试 pass@1	81.0	83.3	83.0	71.1	71.5
LiveCodeBench 代码生成 pass@1	73.3	77.3	71.8	66.5	63.5
Aider 代码编辑 pass@1	71.6	79.6	76.9	65.0	57.0
Humanity's Last Exam 推理与百科知识 pass@1	17.7	20.6	18.4	11.75	8.5

资料来源：DeepSeek 公众号，国元证券研究所

8 月，DeepSeek-V3.1 正式发布，模型升级主要在于，1）混合推理架构：一个模型同时支持思考模式与非思考模式；2）更高的思考效率：相比 DeepSeek-R1-0528，DeepSeek-V3.1-Think 能在更短时间内给出答案；3）更强的 Agent 能力：通过 Post-Training 优化，新模型在工具使用与智能体任务中的表现有较大提升。

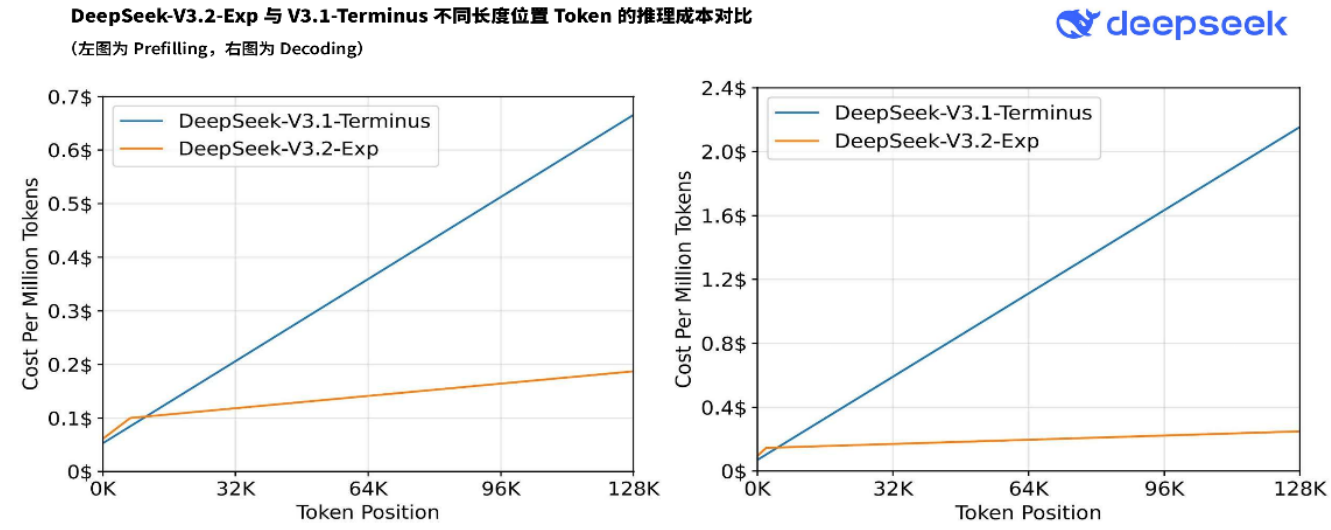
图 28：DeepSeek-V3.1 在编程智能体测评中的表现显著提升

Benchmarks	DeepSeek-V3.1	DeepSeek-V3-0324	DeepSeek-R1-0528
SWE-bench Verified	66.0	45.4	44.6
SWE-bench Multilingual	54.5	29.3	30.5
Terminal-Bench	31.3	13.3	5.7

资料来源：DeepSeek 公众号，国元证券研究所

9 月 29 日，DeepSeek-V3.2-Exp 正式发布，这是一个实验性(Experimental)的版本。作为迈向新一代架构的中间步骤，V3.2-Exp 在 V3.1-Terminus 的基础上引入了 DeepSeek Sparse Attention（一种稀疏注意力机制），针对长文本的训练和推理效率进行了探索性的优化和验证。DSA 的核心逻辑在于，通过引入了一种极其高效的筛选机制在于将计算复杂度从平方级降低到了线性级，令模型不必在每一个步骤都关注上下文中的所有信息。这种机制并不盲目地丢弃信息，而是像人类阅读长文时那样，学会了速读与精读的结合，由两个精密配合的组件构成：闪电索引器(Lightning Indexer)与细粒度 Token 选择机制(Fine-grained Token Selection Mechanism)。

图 29：引入 DSA 后模型长文本训练和推理效率大幅提升



资料来源：DeepSeek 公众号，国元证券研究所

12 月 1 日，DeepSeek-V3.2 和 DeepSeek-V3.2-Speciale 正式发布。

DeepSeek-V3.2 的目标是平衡推理能力与输出长度，适合日常使用，例如问答场景和通用 Agent 任务场景。在公开的推理类 Benchmark 测试中，DeepSeek-V3.2 达到了 GPT-5 的水平，仅略低于 Gemini-3.0-Pro；相比 Kimi-K2-Thinking，V3.2 的输

出长度大幅降低，显著减少了计算开销与用户等待时间。DeepSeek-V3.2-Speciale 的目标是将开源模型的推理能力推向极致，探索模型能力的边界。

DeepSeek V3.2-Speciale 是 DeepSeek-V3.2 的长思考增强版，同时结合了 DeepSeek-Math-V2 的定理证明能力。该模型具备出色的指令跟随、严谨的数学证明与逻辑验证能力，在主流推理基准测试上的性能表现媲美 Gemini-3.0-Pro。更令人瞩目的是，V3.2-Speciale 模型成功斩获 IMO 2025（国际数学奥林匹克）、CMO 2025（中国数学奥林匹克）、ICPC World Finals 2025（国际大学生程序设计竞赛全球总决赛）及 IOI 2025（国际信息学奥林匹克）金牌。其中，ICPC 与 IOI 成绩分别达到了人类选手第二名与第十名的水平。

DeepSeek-V3.2 首次将思考融入工具使用，改进了以往版本在思考模式下无法调用工具的缺陷，并且同时支持思考模式与非思考模式的工具调用。DeepSeek 提出了一种大规模 Agent 训练数据合成方法，构造了大量“难解答，易验证”的强化学习任务（1800+环境，85000+复杂指令），大幅提高了模型的泛化能力。DeepSeek-V3.2 模型在智能体评测中达到了当前开源模型的最高水平，大幅缩小了开源模型与闭源模型的差距。值得说明的是，V3.2 并没有针对这些测试集的工具进行特殊训练，足以证明其在真实应用场景中能够展现出较强的泛化性。

图 30：DeepSeek-V3.2 在各类数学、代码与通用领域评测集中表现优异

Benchmark	GPT-5 High	Gemini-3.0 Pro	Kimi-K2 Thinking	DeepSeek-V3.2 Thinking	DeepSeek-V3.2 Speciale
AIME 2025 美国数学邀请赛	94.6(13k)	95.0(15k)	94.5(24k)	93.1(16k)	96.0(23k)
HMMT Feb 2025 哈佛 MIT 数学竞赛	88.3(16k)	97.5(16k)	89.4(31k)	92.5(19k)	99.2(27k)
HMMT Nov 2025 哈佛 MIT 数学竞赛	89.2(20k)	93.3(15k)	89.2(29k)	90.2(18k)	94.4(25k)
IMOAnswerBench 国际数学奥林匹克竞赛	76.0(31k)	83.3(18k)	78.6(37k)	78.3(27k)	84.5(45k)
LiveCodeBench 世界级编程竞赛	84.5(13k)	90.7(13k)	82.6(29k)	83.3(16k)	88.7(27k)
CodeForces 世界级编程竞赛	2537(29k)	2708(22k)	-	2386(42k)	2701(77k)
GPQA Diamond 理工科博士生测试	85.7(8k)	91.9(8k)	84.5(12k)	82.4(7k)	85.7(16k)
HLE 人类全学科前沿难题测试	26.3(15k)	37.7(15k)	23.9(24k)	25.1(21k)	30.6(35k)

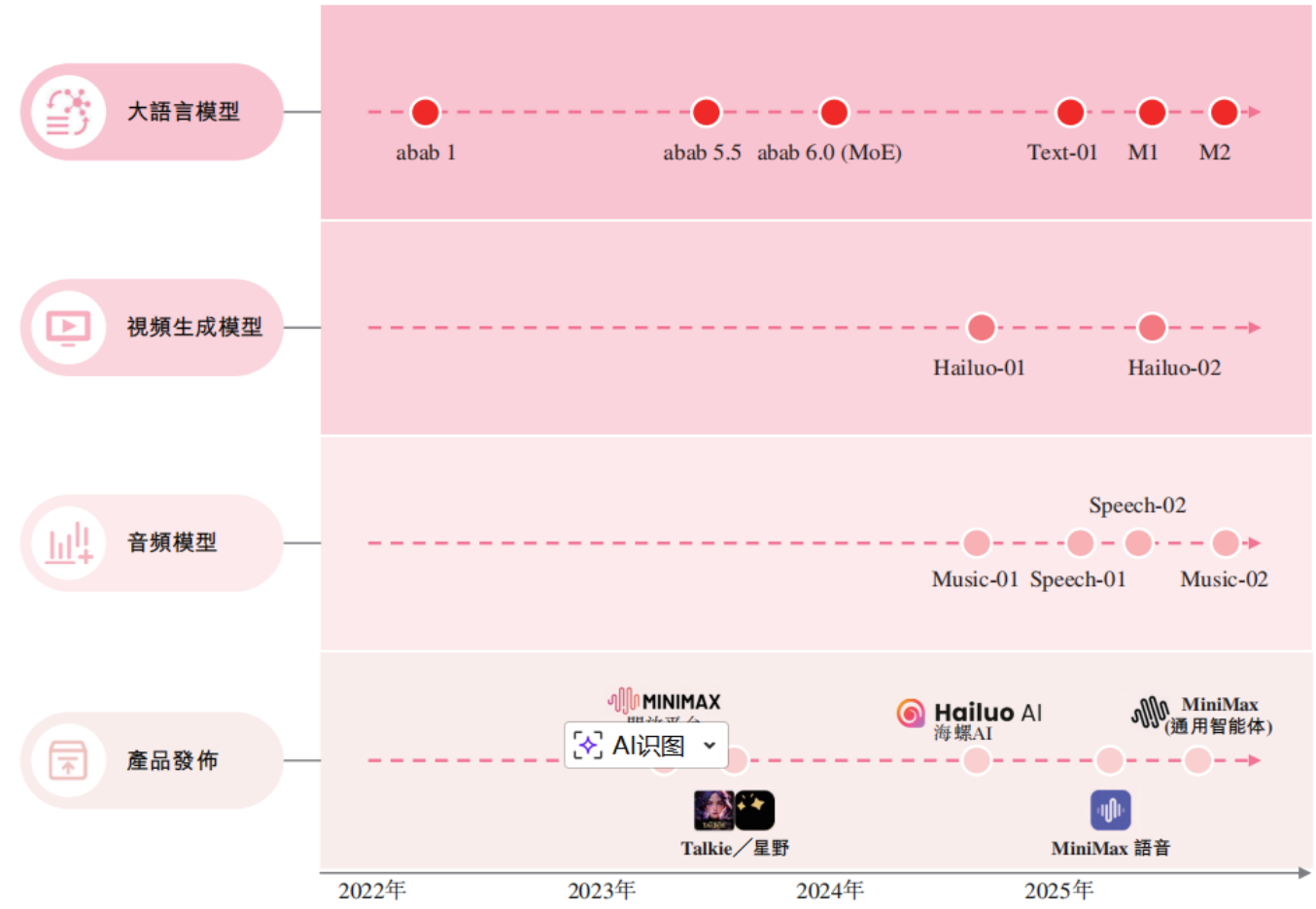
资料来源：DeepSeek 公众号，国元证券研究所

2.2.3 MiniMax

MiniMax 是全球领先的通用人工智能科技公司。自 2022 年初成立以来，公司以“与所有人共创智能”为使命，致力于推动人工智能科技前沿发展，实现通用人工智能。

MiniMax 自主研发了一系列多模态通用大模型，包括文本模型 MiniMax M2、视频模型 Hailuo 2.3、语音合成 Speech 2.6 和音乐模型 Music 2.0，具备强大的代码和 Agent 能力，以及超长上下文处理能力，能够理解、生成并整合包括文本、音频、图像、视频和音乐在内的多种模态。

图 31：MiniMax 发布了一系列多模态模型

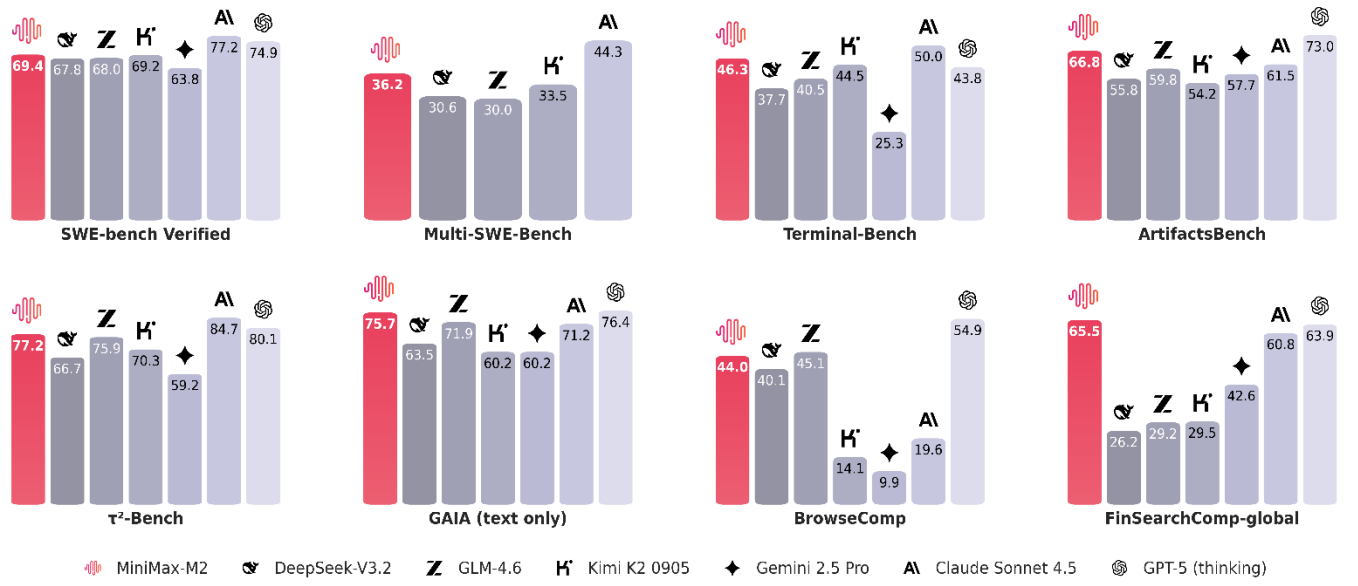


资料来源：MiniMax 港股招股书，国元证券研究所

10 月 27 日，MiniMax M2 正式开源并发布，模型具备：

- 顶级代码能力：专为端到端开发工作流打造，在 Claude Code、Cursor、Cline、Kilo Code、Droid 等多种应用中表现卓越；
- 强大 Agentic 表现：出色规划并稳定执行复杂长链条工具调用任务，协同调用 Shell、Browser、Python 代码执行器和各种 MCP 工具；
- 极致性价比&速度：通过高效的激活参数设计，实现智能、速度与成本的最佳平衡。

图 32：MiniMax-M2 在编程、使用工具和深度搜索等基准测试上表现优秀



资料来源：MiniMax 官网，国元证券研究所

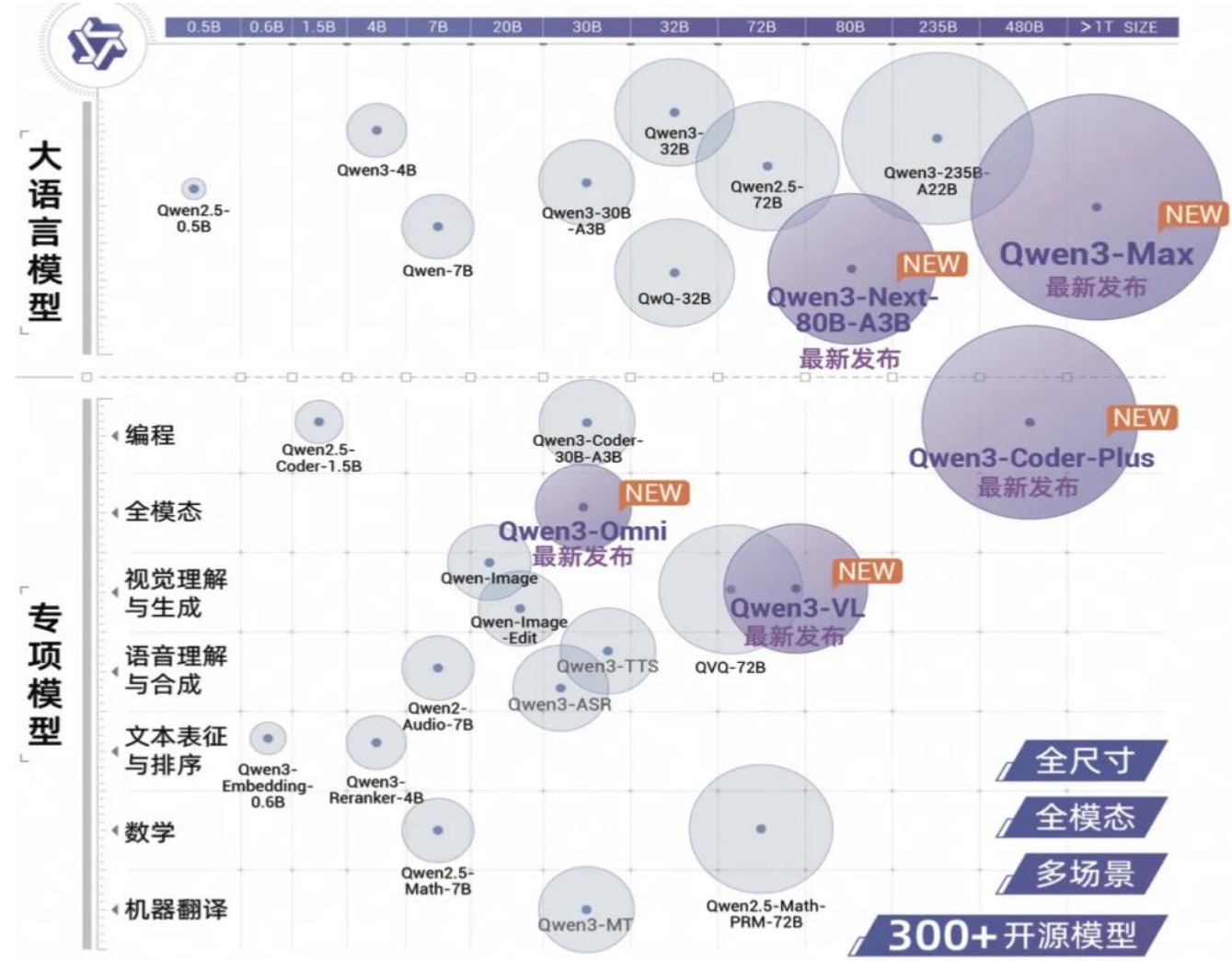
基于这些自研模型，MiniMax 面向全球推出一系列 AI 原生产品，包括 MiniMax Agent、海螺视频 HailuoAI、MiniMax Audio、星野 Takie 等，以及面向企业和开发者的开放平台。2024 年 9 月，公司上线“AI 创作音乐”和“AI 生成视频”功能，形成海螺音乐、海螺视频两款应用，2025 年 1 月全球同步上线海螺语音（后更名为 MiniMax Audio）。海螺音乐能够生成英语、中文、日语、阿拉伯语等多种语言的歌曲，支持全球化 AI 音乐创作，AI 会自动调整旋律与发音，确保不同语种歌曲的自然流畅。海螺视频基于 Minimax 自研视频模型，可将文本指令或图像转化为高质量视频内容，应用采用 Diffusion Transformer 架构，模拟现实物理规律，处理快速运动、表情变化等复杂动作更流畅。同时依托较高的提示词转化率、电影级运镜算法和像素级动态演算引擎，海螺视频在海内外用户中获得较高评价。海螺语音核心功能基于 T2A-01 系列模型实现，支持中文、粤语、英语、法语、俄语等 17 种语言及上百种预置音色，具备情感理解能力，语音输出效果更加真实。

自研模型的飞速进展，有力支撑了 MiniMax 的全球业务拓展。MiniMax 的自研多模态模型及 AI 原生应用已累计为来自超过 200 个国家及地区的逾 2.12 亿名个人用户，以及来自超过 100 个国家及地区的 100000 余名企业客户以及开发者提供服务。

2.2.4 阿里巴巴

阿里巴巴是全球少数大模型与云计算都全栈自研且技术领先的全栈人工智能提供商。在模型层面，截至 2025 年 9 月，通义大模型开源 300 余个模型，覆盖不同大小的“全尺寸”及语言、图像、语音、视频等“全模态”，全球下载量突破 6 亿次，全球衍生模型 17 万个，稳居全球第一。超 100 万家客户接入通义大模型，国际调研机构沙利文报告显示，2025 年上半年，中国企业级大模型调用市场中通义大模型占比第一。

图 33：通义千问 Qwen 模型家族



资料来源：阿里云公众号，国元证券研究所

旗舰模型 Qwen3-Max 是能力最强、规模最大的“全能型选手”，尤其擅长处理复杂实际问题。其预训练数据量高达 36T tokens，总参数超过万亿；性能在全球权威评测中已超越 GPT-5、Claude Opus 4 等顶尖模型，跻身世界前三；在代表 Agent 核心能力的工具调用能力和 Coding 能力评测中均位列全球第一梯队。

下一代基础模型架构 Qwen3-Next 及系列模型通过一系列前沿的工程技术，在保持卓越性能的同时，大幅提升了训练和推理的效率。该模型总参数 80B，但仅需激活 3B 参数，性能便可媲美 235B 的旗舰模型。通过混合注意力、高稀疏度 MoE 等创新技术，其训练成本相较 Qwen3-32B 锐减超过 90%，长文本推理吞吐量提升 10 倍以上。

全模态模型 Qwen3-Omni 则成功解决了以往多模态模型中常见的“模态性能权衡”难题，首次实现音、视、文多模态混合训练而各项能力不降反升，能像人类一样“听说写”，为车载、智能穿戴等设备提供高质量、低延时的交互。多模态生成模型通义万相 Wan2.5 突破性实现了原生音画同步的视频生成，可生成匹配人声、音效和背景

音乐的视频，时长提升至 10 秒，进一步降低了高质量视频创作的门槛。

在 AI 基础设施层面，阿里实现从服务器、网络、存储到智算集群的 AI Infra 布局，打造支撑万亿参数模型与复杂智能体系统的算力引擎。阿里云推出了新一代磐久超节点服务器，在一个机柜中可容纳高达 128 颗 AI 芯片。这背后是一整套针对高密度算力的系统级工程创新：350 千瓦的单柜功率、精细化的液冷系统，以及高可靠供电设计，确保在高热负载、高并发计算环境下依然保持稳定运行。在硬件兼容性上，磐久服务器支持产业主流的开放架构，可灵活兼容不同厂商的 AI 芯片与 Scale-up 协议，实现从单节点到超节点的高效组装。这种“模块化+开放式”的设计，使 AI 算力的部署与扩展更加灵活，为后续的模型训练、推理和强化学习提供了坚实的算力基础。随着模型规模迈向万亿参数、集群规模扩展到“万卡”甚至“十万卡”级别，网络性能成为 AI 集群效率的生命线。阿里云此次推出的 HPN 8.0 高性能网络架构，是全球首批实现“训推一体”设计的 AI 网络。

在应用层面，千问 APP 正式上线，高德地图、淘宝众多 C 端业务场景正加速融入 AI 生态体系。11 月 17 日，阿里巴巴正式发布了基于 Qwen 系列模型的“千问 App”，并将其战略定位清晰地锚定为“AI 超级入口”。12 月 10 日，千问仅用 23 天，月活跃用户数突破 3000 万，跻身全球增长最快的 AI 应用行列。与此同时，AI PPT、AI 写作、AI 文库、AI 讲题四项新功能同步开放。12 月 18 日，千问 APP 已接入高德庞大的实时地理数据系统——每日超千亿次实时调用的时空引擎、覆盖全球超 2 亿 POI（兴趣点）。千问 APP 能够调用高德地图、扫街榜等底层服务，即时生成包含推荐餐厅、地理位置及出行方案的可视化卡片，点击即可跳转导航或打车。

图 34：千问 APP 接入高德地图

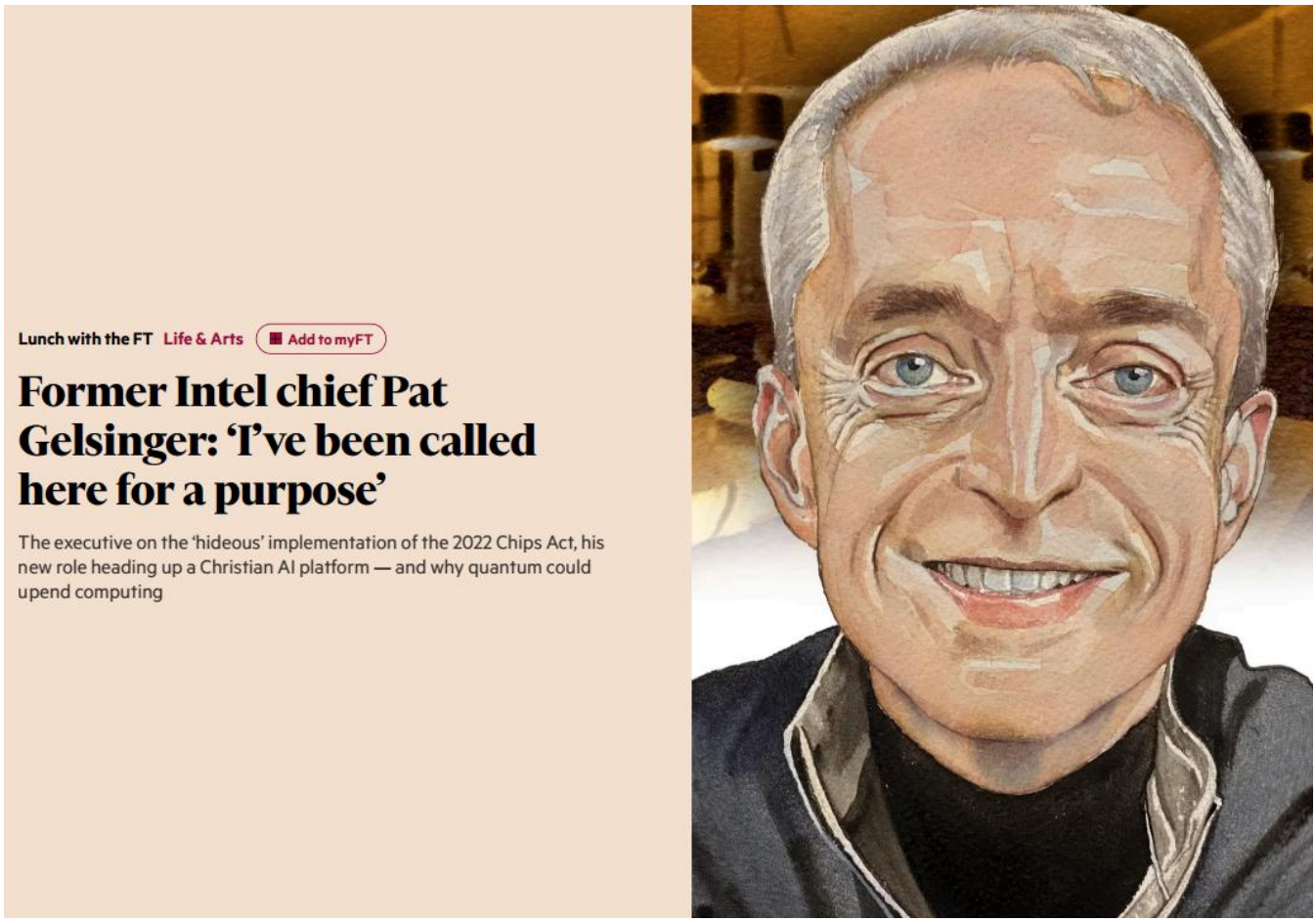


资料来源：财联社公众号，国元证券研究所

3. 量子科技前景广阔，政策助力产业落地

2025 年 11 月，英特尔前首席执行官帕特·基辛格(Pat Gelsinger)接受英国《金融时报》采访，围绕量子科技作出罕见而大胆的技术预言，引发产业界与投资界的广泛关注。基辛格指出：量子计算将成为未来计算体系中不可或缺的核心力量，并与经典计算、AI 计算共同构成推动科技变革的“三位一体(Holy Trinity)”。在他看来，量子计算的商业化速度将远超主流预期，不需要二十年，而是将在未来两年内进入主流市场，成为下一轮计算革命的关键。

图 35:《金融时报》对帕特·基辛格的采访



资料来源：《金融时报》官网，国元证券研究所

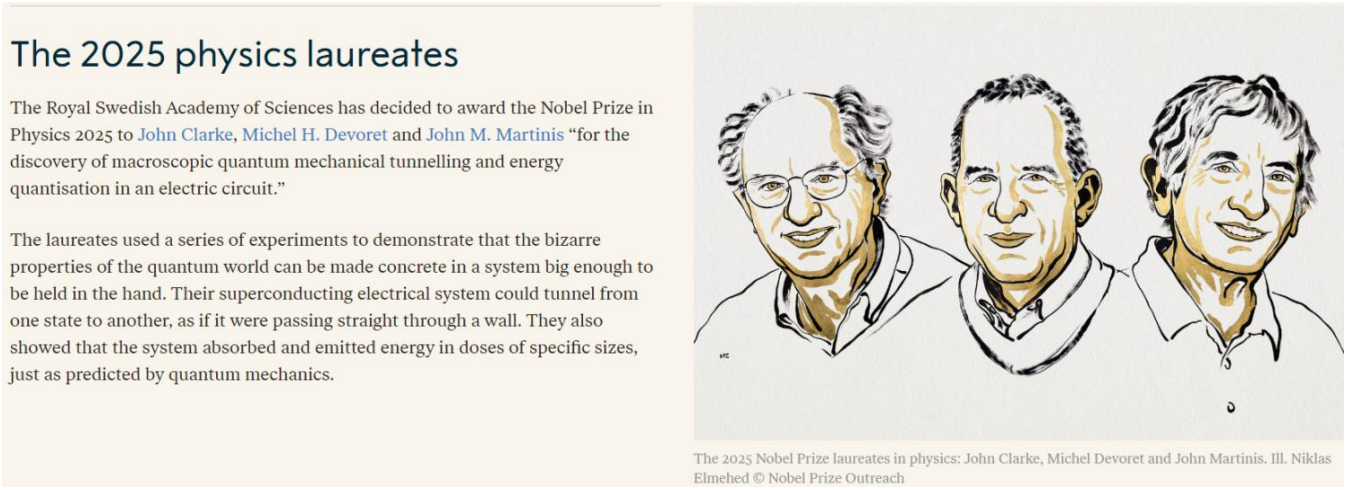
从行业视角看，基辛格的判断意味着量子科技已从“前沿概念”跃升为“战略高地”。当量子计算与 AI 计算深度融合，经典架构无法解决的高复杂度计算、仿真、材料设计、密码学、安全、能源系统等瓶颈将被打开。对于产业研究和资本布局而言，量子科技不再只是长期愿景，而是真实且迫近的结构性机会窗口。

3.1 量子科技加速发展，引领未来产业变革

在量子力学诞生 100 周年之际，2025 年诺贝尔物理学奖颁发给了约翰·克拉克(John

Clarke)、米歇尔·H·德沃雷特(Michel H. Devoret)和约翰·M·马蒂尼斯(John M. Martinis)，以表彰他们在宏观系统量子隧道效应及能级量子化的开创性实验工作。

图 36：2025 年诺贝尔物理学奖



资料来源：诺贝尔奖官网，国元证券研究所

目前，量子信息技术主要分为量子计算、量子通信和量子测量三大支柱领域。

图 37：量子信息三大领域的发展定位和应用前景

	量子计算	量子通信	量子测量
发展定位	①为计算困难问题提供高效解决方案，实现突破经典计算极限的算力飞跃 ②量子计算与经典计算长期并存，相辅相成	①连接量子信息处理节点构成量子信息网络;量子密钥分发服务于经典通信加密 ②量子通信与经典通信应用场景不同	①实现物理量测量和信息获取的精度、分辨率、稳定度等性能指标进一步提升 ②经典测量到量子测量是发展必然趋势
应用前景	①5年:基于含噪声中等规模量子处理器(NISQ)和云平台探索具备实用化价值的应用算例 ②远期:大规模可编程容错量子计算机及其应用	①5年:量子信息网络关键技术突破，实验网络和标准体系建设;量子保密通信商用化探索 ②远期:量子通信与量子计算融合形成量子信息网络	①5年:新一代定位、导航和授时系统，微弱磁场和重力场测量系统，高灵敏度成像系统 ②远期:小型化和商用化量子测量系统和量子传感器

资料来源：前瞻产业研究院《2024-2029 年中国量子通信行业市场前瞻与投资策略分析报告》，国元证券研究所

3.2 产业政策不断出台，未来发展动力强劲

近年来，我国政府不断出台政策，中央和地方不断加大投入力度，加快量子信息产业发展。从总体趋势看，我国量子信息产业政策正从顶层设计逐步走向关键技术攻关与

应用场景落地。2025 年 3 月，第十四届全国人大三次会议审议《2025 年政府工作报告》，将量子科技纳入未来产业发展的重点领域，强调扩大科研投入，推动量子科技与生物科技、脑机智能等前沿技术协同演进，加快量子技术从科研突破走向产业化。2025 年 6 月，市场监管总局与工业和信息化部联合发布《计量支撑产业创新生产力发展行动方案（2025-2030 年）》，聚焦量子计量和量子传感关键技术，推动量子传感器在温度、磁场、重力等方向的标准化、小型化与产业化应用，强调量子精密测量在工业、能源和国防领域的支撑作用。2025 年 10 月，《中共中央关于制定国民经济和社会发展第十五个五年规划的建议》中将量子科技纳入未来产业布局，进一步从国家战略层面明确量子信息产业的长期发展方向。总体来看，政策内容从科研任务向市场应用逐步延展，覆盖关键技术突破、核心器件体系构建、产业链完善和应用推广，为我国量子信息产业形成系统化、全链条的发展路径提供了坚实支撑。

表 2：量子信息技术产业近期政策汇总

日期	政策	颁布单位	内容
2025.10	《中共中央关于制定国民经济和社会发展第十五个五年规划的建议》	中国共产党第二十届中央委员会第四次全体会议	前瞻布局未来产业，探索多元技术路线、典型应用场景、可行商业模式、市场监管规则，推动量子科技、生物制造、氢能和核聚变能、脑机接口、具身智能、第六代移动通信等成为新的经济增长点。
2025.06	《计量支撑产业新质生产力发展行动方案（2025-2030 年）》	市场监管总局、工业和信息化部	面向量子通信、量子计算、量子精密测量高速发展的需求，开展量子传感、量子效应和量子调控关键技术研究，围绕时间频率、温度、磁场、电场、力学等物理量，攻克新一代计量基准量子化、计量标准小型化、量值传递扁平化等量子精密测量关键技术，研发量子计量基标准核心器件，提升深低温、强磁场、超高压等极端条件和复杂环境下的精密测量能力，选取典型领域开展示范应用。
2025.03	《2025 年政府工作报告》	第十四届全国人民代表大会第三次会议	建立未来产业投入增长机制，培育生物制造、量子科技、具身智能、6G 等未来产业。
2025.01	《关于组织开展 2025 年未来产业创新任务揭榜挂帅工作的通知》	工业和信息化部办公厅	围绕量子计算、量子通信、量子精密测量 3 大方向，拟部署 17 项揭榜任务，加强产业共性关键技术攻关，研制核心器件与设备，提升产业公共服务能力，推动量子技术在医疗、交通、能源、金融等领域应用落地。

资料来源：中国政府网，国元证券研究所

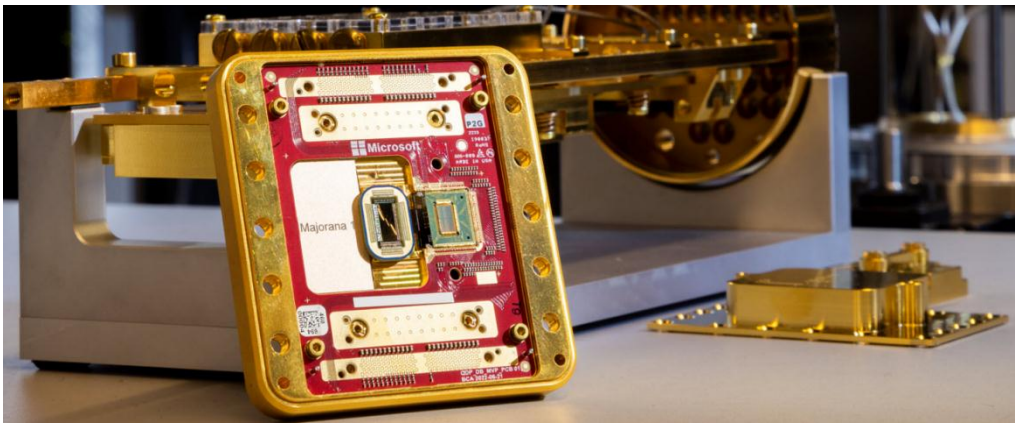
3.3 国际巨头快速推进，引领量子产业发展

3.3.1 微软(Microsoft)

2025 年，微软在量子科技领域加速布局，一方面通过 Azure Quantum 进一步强化量子-经典混合计算能力，另一方面正式推出基于拓扑量子比特的 Majorana 1 原型处理器，同时在全球范围扩建量子实验室以支持硬件工程化验证。

2025 年 2 月，微软通过官方 Azure Blog 正式宣布 Majorana 1 问世，这是其首款基于拓扑量子比特(topological qubits)构建的量子处理器原型。微软在公告中介绍，Majorana 1 使用一种新型 topoconductor 材料，通过超导-半导体异质结构构造出 H 形纳米线(H-shaped nanowire)，并在其中产生 Majorana 零模(Majorana zero modes)，作为拓扑量子比特的核心物理机制。微软认为拓扑比特路线有潜力提高量子比特的抗噪声能力，为未来构建大规模容错量子计算奠定基础。

图 38：微软首枚拓扑量子比特处理器 Majorana 1



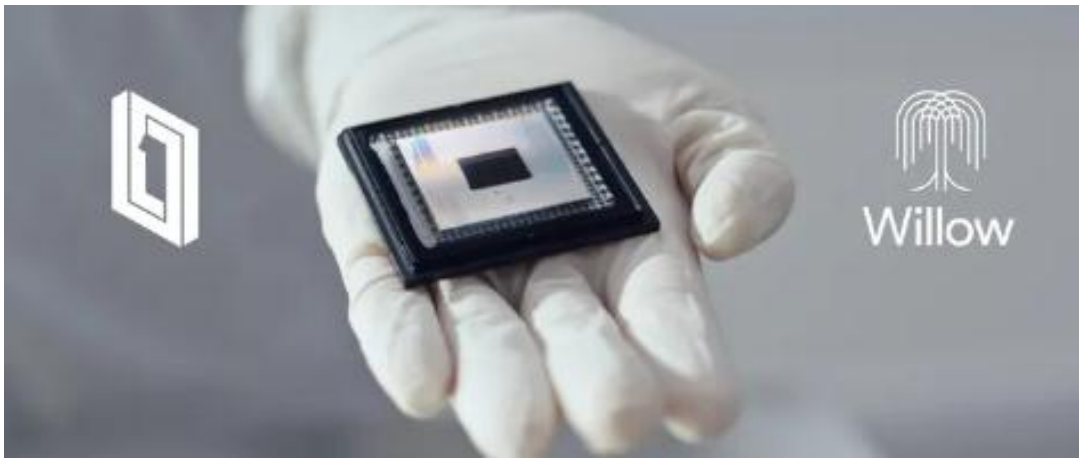
资料来源：微软官网，国元证券研究所

3.3.2 谷歌(Google)

2025 年，Google 在量子科技领域的布局持续推进：通过 Willow 量子芯片在量子纠错和超越经典算力上取得关键突破；依托 Willow 推出 Quantum Echoes 算法，首次在硬件上实现可验证的量子优势，将量子计算进一步推近真实应用场景；公司管理层与量子团队明确提出 5 年内实现商业化量子应用的时间表，并将重点锁定在材料科学等具体产业方向，为量子技术商业落地设定了清晰路线。

在硬件领域，Google 通过新一代 Willow 量子芯片在量子误差校正和随机电路采样(Random Circuit Sampling, RCS)基准上取得重要进展。Willow 采用 105 个超导量子比特，在表面码(surface code)框架下，通过将编码网格从 3×3 扩展到 5×5、7×7，实现了随着物理比特数增加、逻辑误差率反而指数级下降，解决了量子误差校正领域追踪近 30 年的一项关键难题。同时，在 RCS 基准上，Willow 在不到五分钟内完成的计算，相当于当今最快超算需要约 10^{25} 年才能完成，标志着其在“超越经典(beyond-classical)”算力测试中显著拉大了与经典计算的差距。

图 39：Google 新一代 Willow 量子芯片



资料来源：Google 官网，国元证券研究所

在算法领域，Google 依托 Willow 推出的 Quantum Echoes 算法被视为迈向真实应用的重要里程碑。2025 年 10 月，Google 在官方博客“Our Quantum Echoes algorithm is a big step toward real-world applications for quantum computing”中披露，其团队在 Willow 芯片上实现了名为 Quantum Echoes 的 out-of-order time correlator(OTOC)算法，并在 Nature 上发表论文，宣称这是首次在实际硬件上实现“可验证的量子优势(verifiable quantum advantage)”。该算法用于模拟分子结构等量子多体系统，实验表明，在 65 个量子比特规模下，Quantum Echoes 的运行速度约为当前最佳经典对手（运行在 Frontier 超算上的算法）的 13000 倍，同时计算结果可通过另一台同等级量子设备交叉验证，从而兼顾“超越经典算力”和“结果可验证性”。

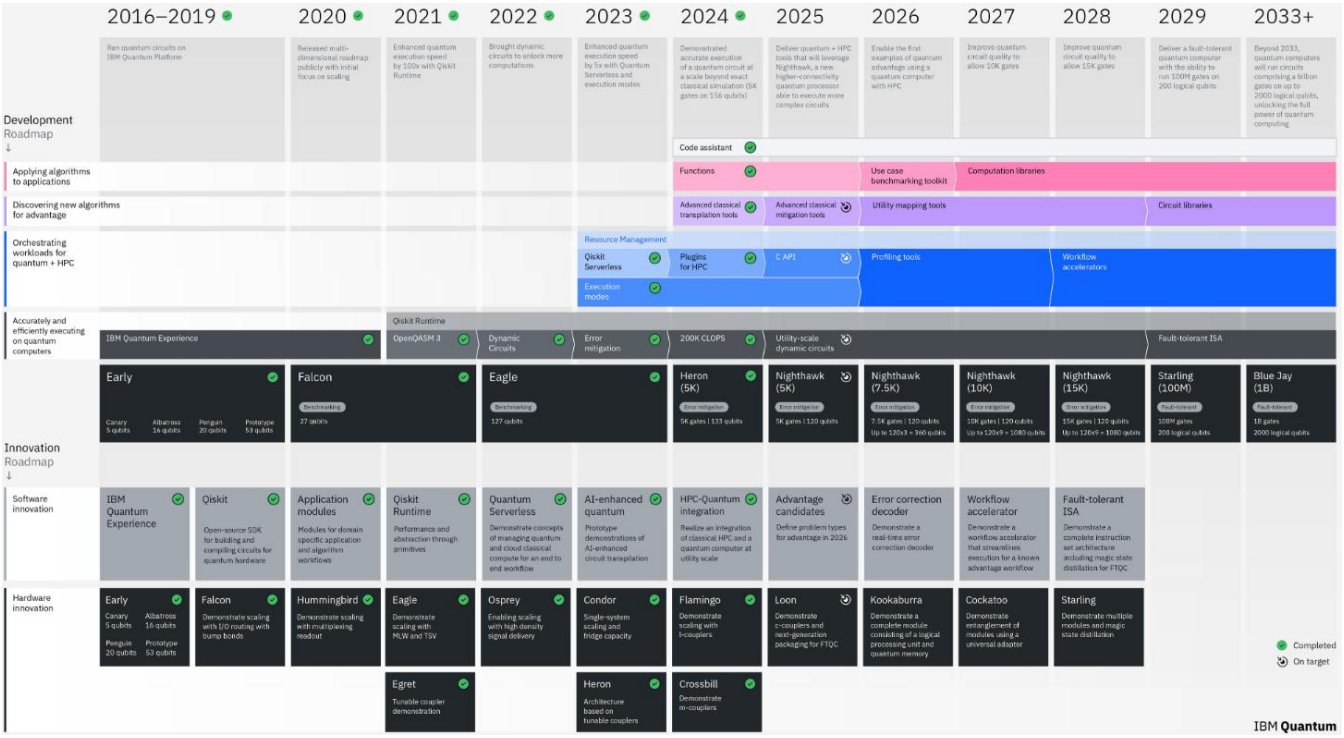
3.3.3 国际商业机器公司(IBM)

2025 年 11 月，IBM 在纽约召开的量子开发者大会上连续发布多项关键成果：全新 Nighthawk 量子处理器、实验性量子芯片 Loon 架构、更新后的 Qiskit 软件技术栈以及首个社区量子性能追踪器。这一系列举措勾勒出 IBM 向“验证性量子优势”迈进的清晰路径，并明确其计划在 2026 年前实现量子优势、在 2029 年前实现容错量子计算。

迄今最先进的 Nighthawk 量子处理器是 IBM 量子路线图中第一个明确为“应用规模扩展”而生的芯片，它采用方形量子比特拓扑结构，由 120 个量子比特组成、配备 218 个可调耦合器。这使其比前代 Heron 增加 20%连接度，可执行量子线路复杂度提升了 30%。这种重新设计并非单纯追求“量子数”，而是着眼于系统互连结构的优化——减少逻辑门中冗余的 SWAP 操作，降低噪声耦合比，提升电路运行效率。官方目标是让 Nighthawk 系列在 2025 年底可运行 5000 个双量子门操作，并计划在未来三年逐步扩展至 15000 个。IBM 计划于 2025 年底让首批 Nighthawk 体系对用户开放。

IBM 同步发布了 Qiskit v2.2，该版本新增 C API 接口，能与 C++、Fortran 等高性能计算语言无缝协作，意味着量子工作负载首次可直接嵌入传统 HPC 系统。在混合计算环境中，这一接口结构将误差校正开销降低 100 倍。与此同时，Qiskit 开放了动态电路(Dynamic Circuits)，能在电路运行过程中利用实时测量进行条件控制。在 100 以上量子比特的规模下，演示结果显示精度提高 24%，门操作量减少 58%。软件层面的重构并非“附属功能”；它对应的是量子领域的另一种核心诉求——控制与验证。这意味着量子算法不再孤立运行，而能作为 HPC 的协处理层直接插入工业工作流程。例如，化学模拟、材料优化、和金融风险模型等算法可在经典系统的高并行环境内调用量子核心计算。此外，IBM 宣布将在 2027 年前建立量子应用库(Qiskit Function Library)，涵盖哈密顿模拟、优化、机器学习与微分方程四大领域。这一步让量子生态从“实验算法”转向“可复用的软件堆栈”，与当年 CUDA 推动 GPU 计算普及的历程如出一辙。

图 40：IBM 量子路线图



资料来源：IBM 官网，国元证券研究所

3.3.4 英伟达(NVIDIA)

2025 年 10 月，NVIDIA 宣布推出 NVIDIA NVQLink™，这是一种开放式系统架构，可将 GPU 计算的极致性能与量子处理器紧密结合，以构建加速的量子超级计算机。NVIDIA 创始人兼首席执行官黄仁勋表示，“NVQLink 是连接量子与经典超级计算机的罗塞塔石碑，它将量子和经典超级计算机整合成一个统一的、连贯的系统，这标志着量子 GPU 计算时代的来临。”

NVQLink 将量子处理器和控制硬件系统的多种方法直接连接到 AI 超级计算，为量子从业人员克服在扩展硬件时面临的主要集成难题提供了统一的整体解决方案。

在超级计算中心、量子硬件制造商和量子控制系统提供商的大力支持下，NVQLink 为运行实用量子应用所需的在控制、校准、量子纠错和混合应用开发方面实现突破奠定了基础。

图 41：NVAQC 如何改变量子计算



资料来源：NVIDIA 官网，国元证券研究所

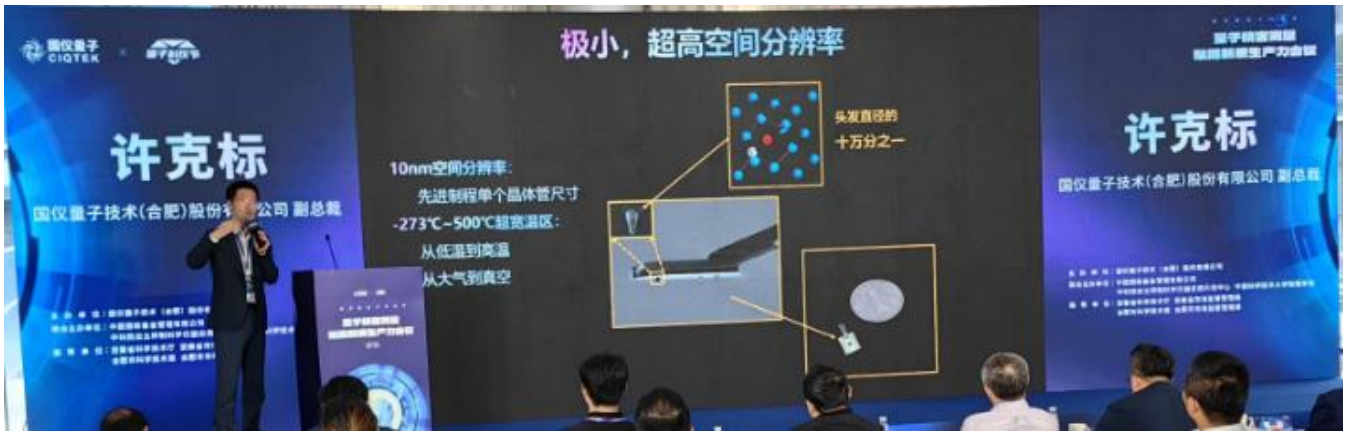
3.4 国内企业加速突破，夯实量子产业根基

3.4.1 国仪量子

国仪量子技术（合肥）股份有限公司的核心技术是以量子精密测量为代表的先进测量技术，为全球范围内企业、政府、研究机构提供以增强型量子传感器为代表的核心关键器件、用于分析测试的科学仪器装备、赋能行业应用的核心技术解决方案等优质的产品和服务。公司源自具有国际声誉的中国科学技术大学，承袭其创新基因与探索精神，致力于为全球科技工作者提供探索微观世界的先进工具。公司荣获多项荣誉，包括“2021 年安徽省科学技术奖一等奖”、“朱良漪分析仪器创新奖”、“安徽省新型研发机构”、“安徽省量子精密测量创新中心”以及“安徽省专精特新冠军企业”等。公司面向量子科技、材料科学、化学化工、生物医学、工业领域、科学教育、能源勘探等领域，致力于帮助客户更高效地推动技术的发展，探索并创造人类的未来。

2025 年 5 月，公司在安徽省合肥市量子科仪谷举行的 2025 量子精密测量赋能新质生产力会议上发布钻石单自旋传感器、量子磁力仪、微波场强仪三款自主研制产品。钻石单自旋传感器采用金刚石氮-空位色心技术，空间分辨率达纳米级，可应用于芯片无损检测等领域；量子磁力仪灵敏度远超传统磁力计，实现了极弱磁场高精度测量，可用于检测心脏跳动产生的磁场变化，为心脑血管疾病的早期诊断提供全新手段；微波场强仪基于里德堡原子量子传感技术，具备高动态范围、高精度的微波电场测量能力，适用于微波计量测试、电磁环境监测等领域。

图 42：许克标博士发布前沿量子传感器



资料来源：中国科技网，科技日报，国元证券研究所

公司量子钻石系列设备与扫描电镜在 2025 年助力客户科研团队取得多项突破性成果：在材料科学领域，曲阜师范大学团队借助公司的量子钻石单自旋谱仪，实现 ^{13}C 天然丰度金刚石中 NV-色心 720μs 超长相干时间，创下该领域最高水平，该研究通过高温高压技术，实现了高相干单 NV-色心的可控制备，为量子信息技术的发展提供了重要基础；在生物医学领域，南京理工大学团队利用量子钻石单自旋谱仪对 FND 进行探测，测到了纳米金刚石 NV 色心激发态的光探测磁共振谱，对激发态 NV 性质进行了研究，并测试了 FND 在外部磁场 B=0,1,2,3,4,10 mT 作用下的连续波谱；在海洋地质领域，海洋二所团队通过量子钻石显微镜(QDM)探讨了普遍存在的氧化风化作用是否会影响 SMS 矿床的磁性特征。

3.4.2 本源量子

本源量子计算科技（合肥）股份有限公司是国内量子计算龙头企业，2017 年成立于合肥市高新区，团队技术起源于中科院量子信息重点实验室。公司聚焦量子计算产业生态建设，打造自主可控工程化量子计算机，围绕量子芯片、量子计算测控一体机、量子操作系统、量子软件、量子计算云平台和量子计算科普教育核心业务，全栈研制开发量子计算，积极推动量子计算产业落地，聚焦生物科技、化学材料、金融分析、轮船制造、大数据等多行业领域，探索量子计算产业应用，争抢量子计算核心专利。

2025 年 5 月，公司推出支持 500+量子比特的中国第四代自主量子计算测控系统“本源天机 4.0”，标志着我国量子计算产业已具备可复制、可迭代的工程化生产能力，为百比特级量子计算机量产奠定产业化基础。量子计算测控系统是量子计算机的“神经中枢”，承担着量子芯片精密信号生成、采集与控制的核心职能。“本源天机 4.0”是继 3.0 版本应用于我国第三代自主超导量子计算机“本源悟空”后的再次升级，在扩展性、集成度、性能稳定性及自动化水平方面实现跨越式提升。公司自主研发的系列底层软硬件架构，进一步增强对量子芯片的高效控制与精准读取，可大幅缩短量子计算机的研发与交付时间。“本源天机 4.0”搭载四大核心软件——量子计算测控系统服务端管理软件 Naga&Venus、超导量子比特底层操控服务软件 Monster、全界面量子芯片调控分析应用软件 Visage、量子计算机操作系统连接软件 Storm。其中，全界面量子芯片调控分析应用软件 Visage，将颠覆超导量子芯片调试的传统模式。

2025 年 8 月，公司联合中国科学技术大学、合肥综合性国家科学中心人工智能研究院成功实现基于量子边编码技术的药物分子性质预测应用，并在我国第三代自主超导量子计算机“本源悟空”上完成真机验证，该项技术为分子性质预测与药物研发开辟了全新的技术路径。在药物研发过程中，精准预测分子性质是高效筛选候选药物的关键环节。此次研究团队创新构建了量子嵌入图神经网络架构，引入首创的量子边编码技术与量子节点嵌入模式，在量子层面实现对原子与化学键的同步处理，大幅提升了分子行为的预测精度，显著提升药物发现效率。

截至 2025 年 10 月底，公司团队研制的中国第三代自主超导量子计算机“本源悟空”，已向世界提供量子算力服务超一年——163 个国家和地区的用户访问超 3600 万次，成功完成了 71 万个全球量子计算任务。

图 43：“本源悟空”量子计算机模型



资料来源：今日头条，人民日报，国元证券研究所

4. 风险提示

1) **人工智能大模型技术在软件及服务领域的落地节奏低于预期的风险：**人工智能大模型技术的发展日新月异，受多种因素的影响，在软件及服务领域的落地节奏可能会低于市场的预期，进而可能会影响相关上市公司业绩的兑现节奏；

2) **量子信息技术开发的风险：**量子信息技术仍处于快速演进阶段，技术路径尚未收敛。在量子通信领域，新型协议和器件持续迭代，技术更新节奏快，产品化与市场需求匹配存在不确定性；在量子计算领域，多条技术路线并行推进，但尚无成熟方案实现实用化，整体仍处于工程验证阶段。如果相关公司未能准确把握技术发展方向，或关键技术研发和产品转化进展不及预期，相关技术和产品可能被替代，从而对行业商业化进程和投资回报产生不利影响；

3) **量子信息技术市场开拓与商业化进展不及预期的风险：**当前量子通信应用主要集中于政务、金融、电力、通信等重点领域，市场拓展在一定程度上依赖相关基础设施建设和政策推进节奏，商业化应用的规模化落地仍存在不确定性。同时，量子计算与量子精密测量整体尚处于早期发展阶段，产品成熟度和应用场景培育仍需时间。若商业化进展不及预期，可能对相关公司的收入增长和投资回报形成制约；

4) **行业竞争加剧的风险：**随着市场逐步成熟，市场规模不断扩大，计算机行业各细分领域将吸引越来越多的企业进入，市场竞争可能会变的更加激烈。新竞争者的进入以及现有国内、外现有企业的竞争加剧，可能导致相关产品和服务价格的下滑、产品更新换代加快、市场份额难以保持的风险；

5) **宏观经济环境变动的风险：**计算机行业的发展趋势和宏观经济环境密切相关，国家整体经济的持续稳定增长和高质量转型发展导向为行业的发展提供了有利的宏观环境，如果全球政治经济形势发生重大变化，可能会导致下游产业链的需求放缓，可能会对行业的发展环境和市场需求造成不利影响。

6) **国际贸易与技术管制带来的风险：**计算机行业上游的部分核心技术、关键材料及高端科研仪器对进口存在一定依赖，在国际经贸环境复杂化、技术管制趋严的背景下，相关限制措施可能对产业链协同和新业务拓展形成约束。尽管国产替代持续推进，但在部分高端环节仍存在不确定性，可能对行业技术迭代进度和长期发展带来不利影响。

投资评级说明

（1）公司评级定义		（2）行业评级定义	
买入	股价涨幅优于基准指数 15%以上	推荐	行业指数表现优于基准指数 10%以上
增持	股价涨幅相对基准指数介于 5%与 15%之间	中性	行业指数表现相对基准指数介于-10%~10%之间
持有	股价涨幅相对基准指数介于-5%与 5%之间	回避	行业指数表现劣于基准指数 10%以上
卖出	股价涨幅劣于基准指数 5%以上		

备注：评级标准为报告发布日后的 6 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 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 指数或纳斯达克指数，新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的），北交所基准指数为北证 50 指数。

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或相当的专业胜任能力，以勤勉的职业态度，独立、客观地出具本报告。本人承诺报告所采用的数据均来自合规渠道，分析逻辑基于作者的职业操守和专业能力，本报告清晰准确地反映了本人的研究观点并通过合理判断得出结论，结论不受任何第三方的授意、影响，特此声明。

证券投资咨询业务的说明

根据中国证监会颁发的《经营证券业务许可证》（Z23834000），国元证券股份有限公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询业务是指取得监管部门颁发的相关资格的机构及其咨询人员为证券投资者或客户提供证券投资的相关信息、分析、预测或建议，并直接或间接收取服务费用的活动。证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。

法律声明

本报告由国元证券股份有限公司（以下简称“本公司”）在中华人民共和国境内（台湾、香港、澳门地区除外）发布，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。若国元证券以外的金融机构或任何第三方机构发送本报告，则由该金融机构或第三方机构独自为此发送行为负责。本报告不构成国元证券向发送本报告的金融机构或第三方机构之客户提供的投资建议，国元证券及其员工亦不为上述金融机构或第三方机构之客户因使用本报告或报告载述的内容引起的直接或连带损失承担任何责任。本报告是基于本公司认为可靠的已公开信息，但本公司不保证该等信息的准确性或完整性。本报告所载的信息、资料、分析工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买证券或其他投资标的的投资建议或要约邀请。本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。本公司建议客户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。在法律许可的情况下，本公司及其所属关联机构可能会持有本报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取投资银行业务服务或其他服务，上述交易与服务可能与本报告中的意见与建议存在不一致的决策。

免责条款

本报告是为特定客户和其他专业人士提供的参考资料。文中所有内容均代表个人观点。本公司力求报告内容的准确可靠，但并不对报告内容及所引用资料的准确性和完整性作出任何承诺和保证。本公司不会承担因使用本报告而产生的法律责任。本报告版权归国元证券所有，未经授权不得复印、转发或向特定读者群以外的人士传阅，如需引用或转载本报告，务必与本公司研究所联系并获得许可。

网址：www.gyzq.com.cn

国元证券研究所

合肥	上海	北京
地址：安徽省合肥市梅山路 18 号安徽国际金融中心 A 座国元证券	地址：上海市浦东新区民生路 1199 号证大五道口广场 16 楼国元证券	地址：北京市朝阳区安定路 5 号院 3 号楼中建财富国际中心 5 层
邮编：230000	邮编：200135	邮编：100029