中国银河证券 | CGS

AI Agent 智能体经济释放“新供给侧改革”乘法效应

—— 计算机行业 2025 年中期策略报告

计算机行业首席分析师：吴砚靖

计算机行业分析师：邹文倩

计算机行业研究助理：胡天昊

AI Agent 智能体经济释放“新供给侧改革”乘法效应

—— 计算机行业 2025 年中期策略报告

2025 年 6 月 21 日

核心观点

- **2025 年上半年回顾：**2025 年以来，计算机行业跑赢上证指数及沪深 300，受 Deepseek 催化，行业 Q1 整体走强，后受年报季及关税风波等因素影响，二季度持续回调。近期受稳定币概念催化，金融 IT 及跨境支付板块走强。上半年涨幅排名前三子板块分别为 SAAS、云计算和智能汽车。
- **2025 年下半年展望：当下国内 AI 投资呈现长短错配，下半年恰是布局良机。**宏观角度，在中国经济动能转换过程中，不同于传统供给侧改革依赖“减法逻辑”（去产能、去库存），当下的“新供给侧改革”乘法效应本质是从“要素替代”到“系统重构”，而 AI Agent 经济正通过“消费创造（场景创新×需求激活）、投资提质（能算协同×空间智联）、出口升级（质量溢价×跨境服务 AI 化）三维路径，促使全要素生产率提升与生态系统重构，中长期而言，**“新供给侧改革”乘法效应将带来十倍以上投资机会。**
- **预期催化剂 DeepSeek R2 版本或于下半年亮相，国内 AI 全产业链受益度有望提升。**DeepSeek 在 5 月发布的版本 DeepSeek-R1-0528 思维深度与推理能力显著提升，市场预期 DeepSeek R2 版本或将在多模态融合、实时决策能力及垂直场景泛化性上实现跃升，AI 应用及算力端有望充分受益。
- **AI Agent 智能体经济已经全新开启，技术于产品迭代呈现不可逆趋势，对应投资机会包括：**1、全球推理算力供需剪刀差不断扩大 2、字节在 AI 应用生态领域已构建起相对优势，建议关注字节生态合作伙伴。3、建议关注在 AI Agent 方面布局领先的垂直领域卡位 SAAS 企业。
- **信创产业进入深水区，国产生态繁荣将提升客户粘性进而提升盈利空间。**
- **投资建议：**下半年建议重点关注以下细分赛道及个股，1、国产算力产业链：工业富联、中科曙光、曙光数创、海光信息、龙芯中科、地平线机器人-W 等；2、第三方 IDC 服务商：润泽科技、宝信软件等；3、信创厂商：中国软件、软通动力、达梦数据等；4、AI+应用：科大讯飞、金山办公、同花顺、嘉和美康、国能日新、彩讯股份、恒生电子、万兴科技等；5、云计算服务商：金蝶国际、金山云、优刻得、深信服；6、一体机及端侧 AI：神州数码、拓维信息、虹软科技、海康威视、中科创达、华勤技术、萤石网络等；7、数据要素产业链中供给、流通、应用公司：拓尔思、深桑达 A、上海钢联等；8、稳定币及 RWA：众安在线（6060.HK）、朗新集团。
- **风险提示：**宏观经济不及预期风险，政策推进不及预期风险，技术研发不及预期风险，行业竞争加剧风险，贸易摩擦风险。

重点公司盈利预测与估值（截至 6.21）

股票代码	股票名称	EPS			PE			投资评级
		2025E	2026E	2027E	2025E	2026E	2027E	
002230.SZ	科大讯飞	0.41	0.59	0.80	114.17	79.69	58.50	推荐
688041.SZ	海光信息	1.35	1.93	2.61	99.46	69.45	51.35	推荐
300033.SZ	同花顺	4.56	5.56	6.44	53.10	43.54	37.60	推荐
0268.HK	金蝶国际	0.04	0.09	0.16	379.84	147.51	86.53	推荐
301162.SZ	国能日新	1.05	1.35	1.72	47.95	37.20	29.23	推荐

资料来源：Wind，中国银河证券研究院

计算机行业

推荐 维持

分析师

吴砚靖

☎：010-66568589

✉：wuyanjing @chinastock.com.cn

分析师登记编码： S0130519070001

邹文倩

☎：010-86359293

✉：zouwenqian@chinastock.com.cn

分析师登记编码：S0130519060003

研究助理：胡天昊

☎：010-20252650

✉：hutianhao_yj @chinastock.com.cn

相对沪深 300 表现图 2025/6/21

资料来源：Wind，中国银河证券研究院

相关研究

1.【中国银河宏观】宏观专题报告_瞭望 2035：人智能重构大国经济新范式

2.【银河计算机】2025 年年度策略_AI Agent 繁荣代开启，科技内需迎新篇章

3.【中国银河证券】Deepseek 冲击波：AI 重塑未来计算机行业价值新起点

目录
Catalog

一、行情回顾

（一）计算机行业年初至今估值情况与市场表现

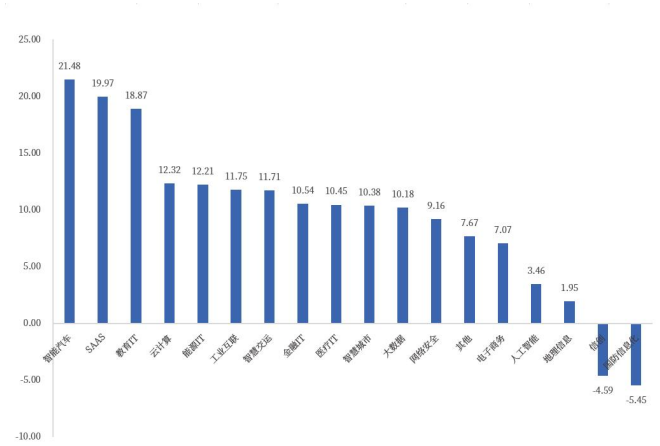
2025 年上半年，计算机行业呈现“先扬后抑”的震荡走势，整体跑赢大盘但分化显著。一季度在 DeepSeek 等大模型流量爆发及 AI Agent 技术突破的催化下强势领涨，行业估值显著抬升；二季度受年报季业绩下滑、美国对华 AI 硬件加征关税等影响持续回调。5 月香港《稳定币条例》正式生效，推动金融 IT 与跨境支付板块逆势走强，成为上半年重要结构性机会。

细分领域表现差异明显，年初至今，计算机子行业中涨幅排名前三分别为 SaaS、云计算和智能汽车：SaaS 板块受益于企业级（to B）智能体落地预期驱动，涨幅居首；云计算板块受国产大模型催化带来本地化部署需求激增表现次之；智能汽车在车载 AI 系统渗透率提升下成为细分板块增长第三位。技术演进上，AI Agent 完成从“工具调用”到部分“自主决策”的范式升级，ClaudeOpus4 等模型已具备 7 小时长任务规划能力，推动生产力智能体在编程、教育等多场景高速增长。

图 1：年初至今计算机指数跑赢沪深 300

资料来源：WIND，中国银河证券研究院

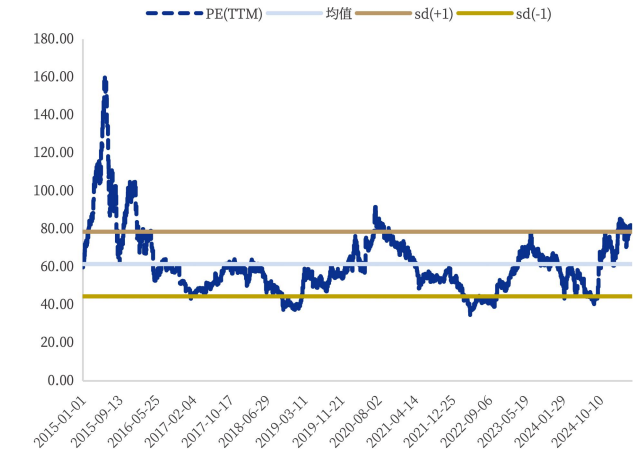
图 2：年初至今计算机子行业涨跌幅（%）



资料来源：WIND，中国银河证券研究院

计算机行业指数过去 10 年 PE（TTM）均值为 61.44 倍；PS（TTM）均值为 3.92 倍。当前计算机行业指数估值处于历史十年均值偏高位置，指数 PE（TTM）值为 79.30 倍，PS（TTM）值为 3.19 倍。

图 3：计算机行业指数近 10 年 PE(TTM)情况



资料来源：WIND，中国银河证券研究院

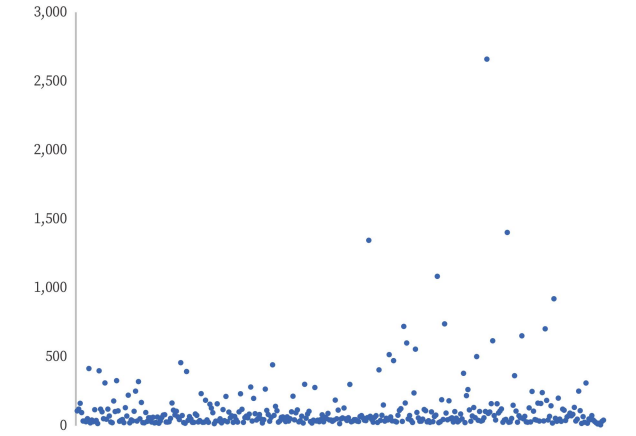
图 4：计算机行业指数近 10 年 PS(TTM)情况

资料来源：WIND，中国银河证券研究院

（二）市值分层表现及涨幅前十个股复盘

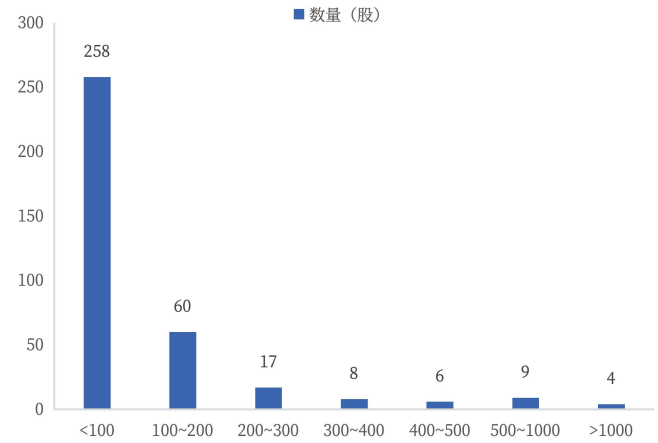
截至 5 月中旬，计算机板块个股共 362 只。据统计，总体市值分布较集中于小于 100 亿区间内，占总数近七成总计 258 只；100 至 200 亿市值区间总计 60 只；200 至 300 亿市值区间总计 17 只；300 至 400 亿市值区间总计 8 只；400 至 500 亿市值区间总计 6 只；超 500 亿市值共 13 只；超 1000 亿市值共 4 只，分别为金山办公、科大讯飞、海康威视和同花顺，总市值分别为 1343.84 亿元、1082.33 亿元、2659.16 亿元和 1400.99 亿元。

图 5：计算机行业个股市值分布情况（亿元）



资料来源：WIND，中国银河证券研究院

图 6：计算机行业个股市值区间分布情况（亿元）



资料来源：WIND，中国银河证券研究院

截至 6 月 21 日，A 股计算机行业年初至今涨幅前十个股如下表所示。

表 1：年初至今涨幅前十个股复盘（截至 6 月 21 日）

证券代码	证券简称	年初至今涨幅降次（%）
301396.SZ	宏景科技	186.79
300468.SZ	四方精创	143.22
300766.SZ	每日互动	123.07
300368.SZ	汇金股份	119.82
688316.SH	青云科技-U	85.67

300546.SZ	雄帝科技	82.81
688171.SH	纬德信息	81.41
688288.SH	鸿泉物联	70.25
688291.SH	金橙子	68.65
002261.SZ	拓维信息	62.59

资料来源：Wind，中国银河证券研究院

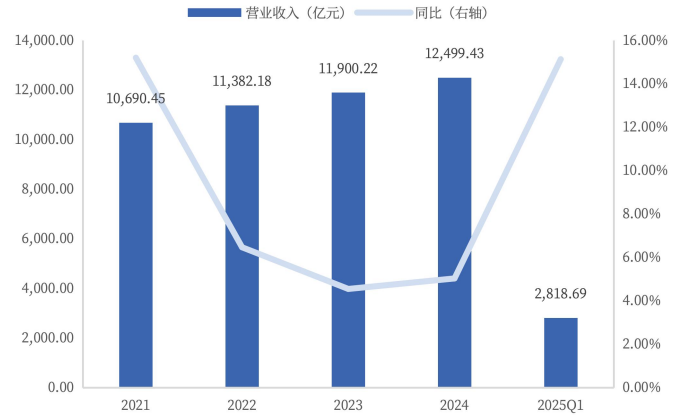
（三）一季报财务指标营收端改善，控费效应下净利润同比增长

1、行业一季度营收同比改善，归母净利润同比大幅增长

2025 年一季度，系经济复苏带动政府及企业信息技术方面支出意愿，付费能力进一步改善，提振计算机行业营收及净利润。2025 年一季度，计算机行业营收同比改善，一季度实现营收 2818.69 亿元，同比增长 15.14%；归母净利润实现 23.29 亿元，同比大幅增长 790.55%，主要系营收增长的同时费用端控制良好，销售、管理及研发费用率较去年同期均有下降。

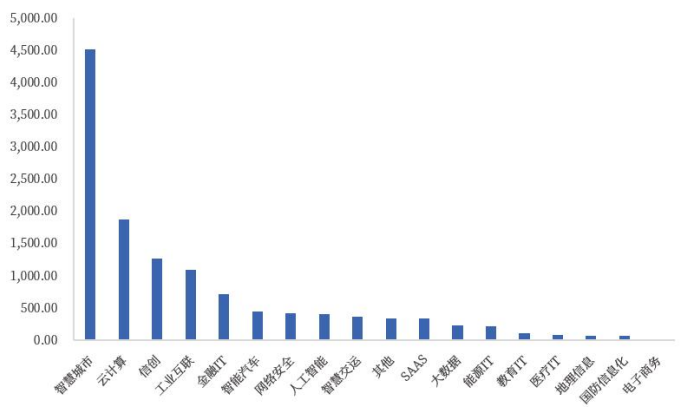
2025 年一季度，计算机板块销售、管理及研发费用率分别降至 7.60%、5.29%、9.33%，分别同比下降 1.11pct、0.85pct、1.64pct。

图 7：计算机行业一季度营收同比改善



资料来源：WIND，中国银河证券研究院

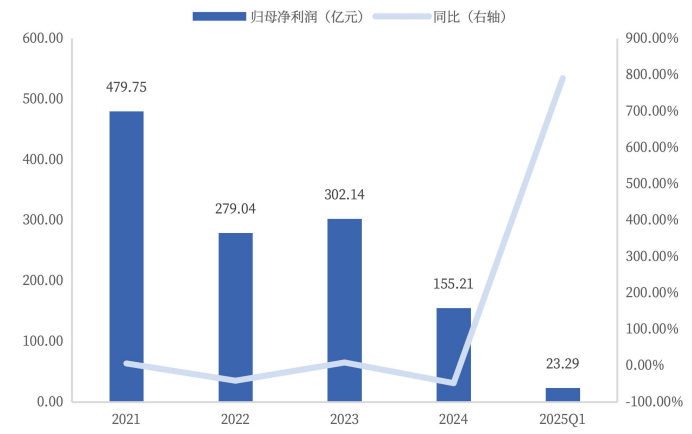
图 8：计算机子行业年初至今营业收入（亿元）



资料来源：WIND，中国银河证券研究院

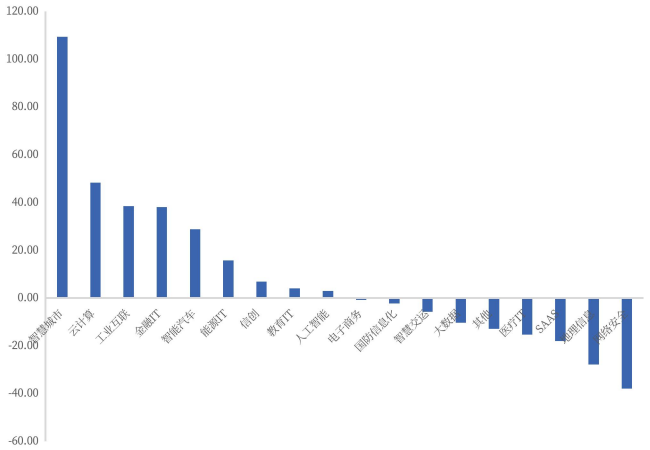
计算机子行业方面，年初至今营收排名前三的分别是智慧城市、云计算和信创；净利润排名前三的分别是智慧城市、云计算和工业互联；其中 9 个子行业板块亏损。

图 9：计算机行业一季度归母净利润同比大幅增长



资料来源：WIND，中国银河证券研究院

图 10：计算机子行业年初至今归母净利润（亿元）

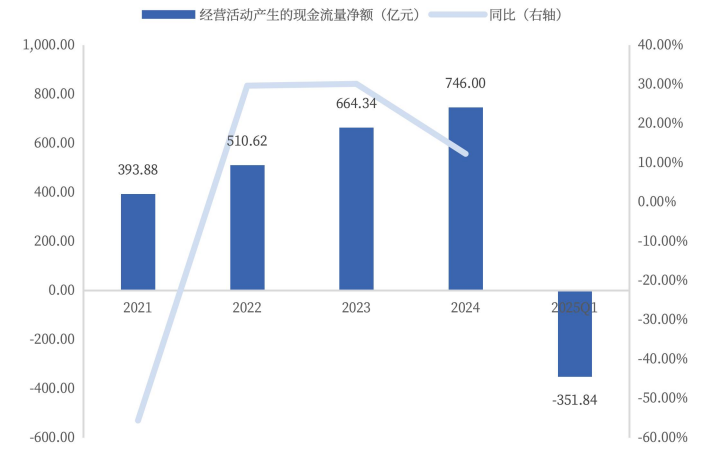


资料来源：WIND，中国银河证券研究院

2、行业一季度经营性现金流同比负额收窄，应收账款周转率提升

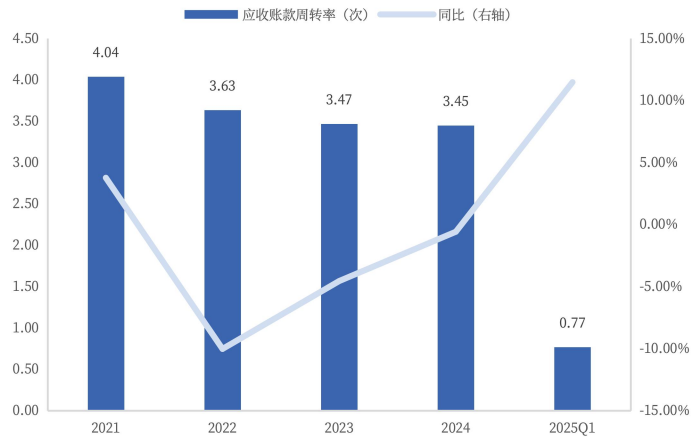
计算机行业一季度经营活动产生的现金流量净额为-351.84 亿元，同比负额收窄，主要系行业需求回暖，应收账款周转率提升至 0.77 次，同比增长 11.48%，可见年初至今行业整体现金回款情况较去年有明显改善。

图 11：计算机行业一季度经营活动净现金流为负，同比负额收窄



资料来源：WIND，中国银河证券研究院

图 12：计算机行业一季度应收账款周转率为 0.77 次，同比增长 11.48%

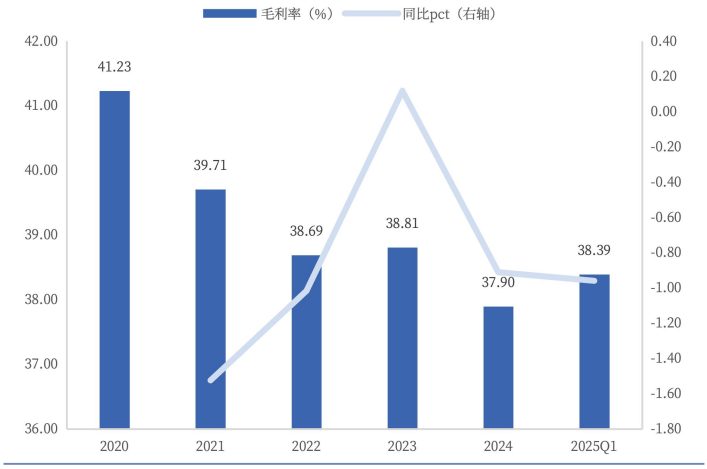


资料来源：WIND，中国银河证券研究院

3、行业一季度毛利率同比小幅下滑，净利率同比增长

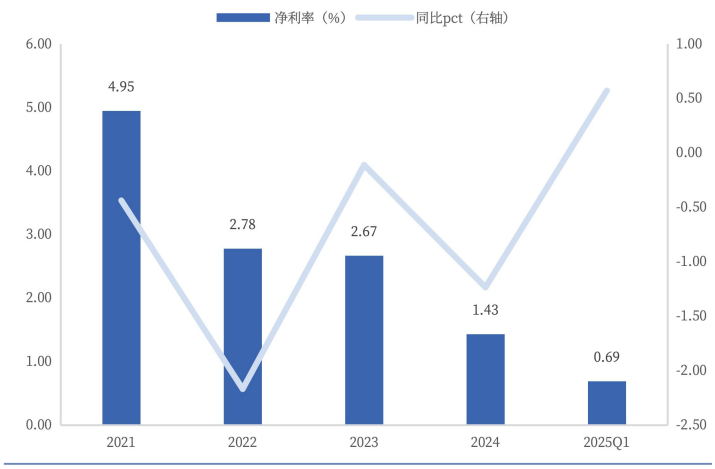
2025 年一季度，计算机行业毛利率降至 38.39%，同比下降 0.96pct；净利率提升至 0.69%，同比上升 0.57pct。

图 13：计算机行业一季度行业平均毛利率同比下降 0.96pct



资料来源：WIND，中国银河证券研究院

图 14：计算机行业一季度行业平均净利率同比上升 0.57pct

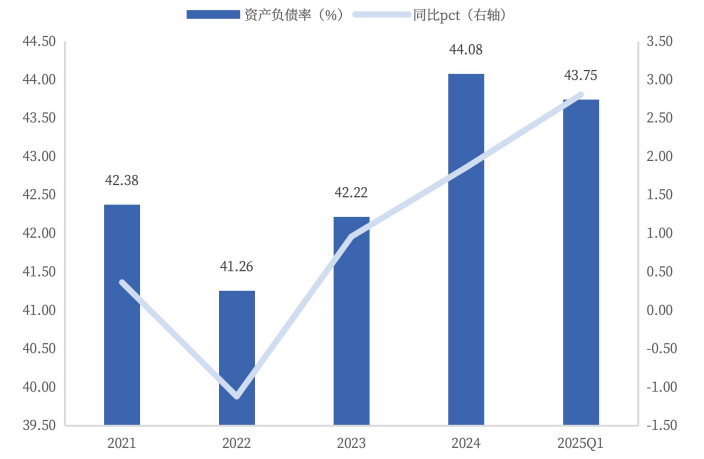


资料来源：WIND，中国银河证券研究院

4、行业一季度资产负债率上升，ROE 同比下降

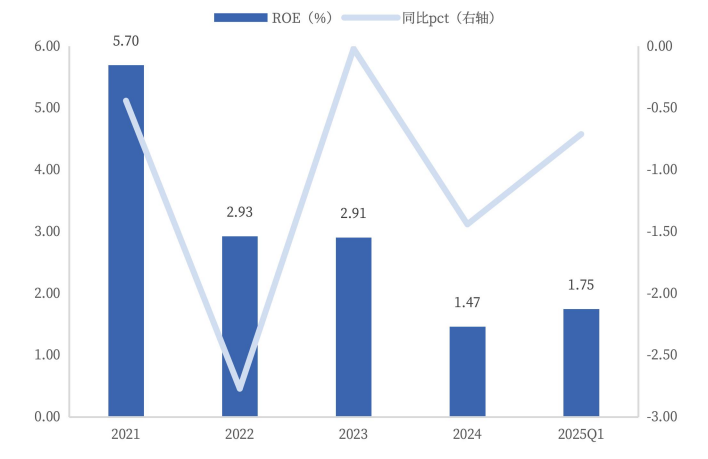
2025 年一季度，计算机行业资产负债率为 43.75%，同比上升 2.81pct；摊薄 ROE 为 1.75%，同比下降 0.71pct。行业发展仍面临一定压力，偿债压力相较去年同期增大，一定程度上影响短期资产盈利水平，但考虑现金流及应收账款周转率情况改善，偿债能力方面相对乐观。

图 15：计算机行业一季度资产负债率为 43.75%，同比上升 2.81pct



资料来源：WIND，中国银河证券研究院

图 16：计算机行业一季度摊薄 ROE 为 1.75%，同比下降 0.71pct



资料来源：WIND，中国银河证券研究院

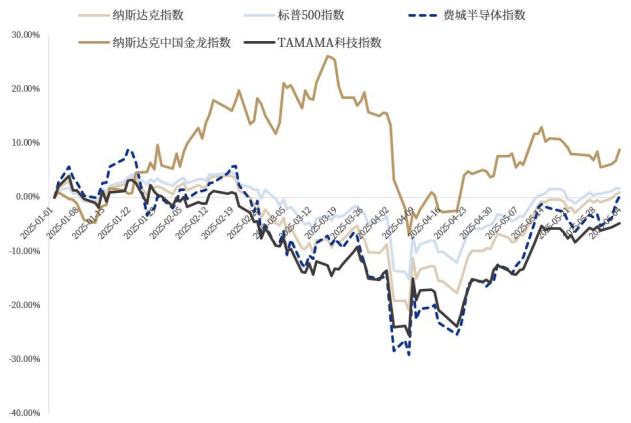
（四）全球科技股行情回顾

1、一季度美股科技震荡下行，五月后出现回升趋势，中概股大涨回调后回升

2025 年一季度，美股科技震荡下行，美股科技跑输大盘，TAMAMA 指数跑输纳斯达克指数。一季度后美股科技整体持续震荡。

2025 年一季度，中概股涨幅可观，经过 3 月后短暂回调后，4 月中旬后恢复上涨趋势。截至 2025 年 6 月上旬，恒生科技指数涨幅为 16.81%，纳斯达克金龙指数涨幅为 7.26%。

图 17：一季度美股科技震荡下行，五月后出现回升趋势



资料来源：WIND，中国银河证券研究院

图 18：美股科技表现不佳，港股、中概股科技及 A 股计算机均上涨

指数代码	指数简称	涨跌幅%				市盈率PE（TTM）
		2025年初至今	2024	2023年	2022年	
SPX.GI	标普500指数	1.52	23.31	24.23	-19.44	26.74
IXIC.GI	纳斯达克指数	0.78	28.64	43.42	-33.10	40.31
SOX.GI	费城半导体指数	1.07	19.27	64.90	-35.83	46.88
8884057.WI	TAMAMA科技指数	-5.03	46.87	67.81	-38.15	32.63
HXC.GI	纳斯达克中国金龙指数	-7.26	4.43	-3.39	-24.63	19.48
HSTECH.HI	恒生科技指数	-16.81	13.70	-8.83	-27.19	20.58
CI005027.WI	计算机	-5.05	9.91	8.90	-25.14	139.88

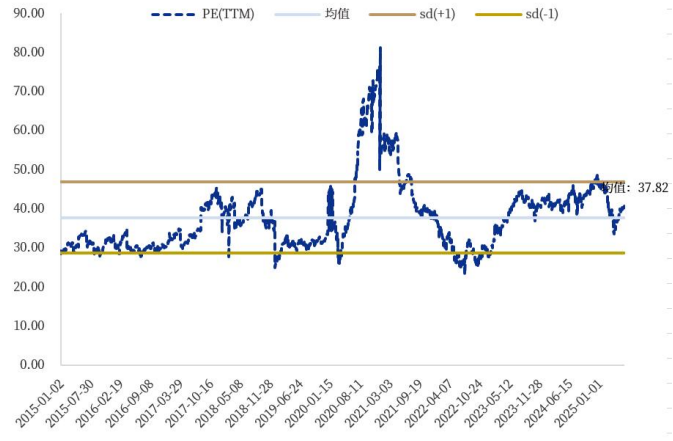
资料来源：WIND，中国银河证券研究院

美股科技纳斯达克指数过去 10 年 PE(TTM)均值为 61.58;标普 500 指数 10 年 PE(TTM)均值为 37.82。**相较过去 10 年历史水平，当前纳斯达克指数估值处均值偏高位置。**截至美东时间 6 月 18 日，纳斯达克指数 PE（TTM）值为 79.63，标普 500 指数 PE（TTM）值为 40.43。

图 19：纳斯达克指数近 10 年 PE(TTM)情况

资料来源：WIND，中国银河证券研究院

图 20：标普 500 指数近 10 年 PE(TTM)情况



资料来源：WIND，中国银河证券研究院

二、AI Agent 智能体经济全新开启

（一）AI Agent 技术范式革命：从工具到自主决策，从个体到协作

2024Q4，AI agent 模型技术实现重大突破，其标志性突破在于能够通过自然语言实现与硬件的交互，从而独立调用硬件操作界面并执行用户指令。代表性模型比如 Claude 3.5 Sonnet、智谱 AutoGLM。Anthropic 于 2024 年 10 月 23 日发布了 Claude 3.5 Sonnet 模型，得益于 Anthropic 推出的 API，Claude 可以感知并与计算机界面进行交互，开发者可以通过集成这一 API，将用户的指令翻译成计算机可以执行的指令，使得 AI 可以模拟人类与计算机的交互方式，包括移动光标、点击屏幕以及通过虚拟键盘输入信息。智谱同样于 2024 年 10 月推出了 AutoGLM，聚焦于设备操

控能力，支持通过工具调用完成具体任务（如操作手机、电脑等），能理解屏幕信息、规划任务、自我判断调整操作，如调节亮度模式、规划路线等。

如果说 2024 年的 AI Agent 像是一个操控工具，那么经历了 2025 年至今的技术演进，AI Agent 已经从“被动工具”迈向“自主决策体”，并且智能体从个体走向协作，AI Agent 开发平台出现，AI Agent 应用逐渐形成生态。2025 年 AI Agent 的技术演进主要可以从以下四个方面来看：

1、环境感知：从文本到多模态融合

Anthropic 于 2025 年 5 月发布 Claude Opus 4 和 Claude Sonnet 4，再次将代码、高级推理和 AI 智能体，推向全新标准。Claude Opus 4 具备强大的图像理解与跨模态推理能力，能处理复杂图文信息、支持多图对比和图像驱动的工具调用，为 AI Agent 提供更接近人类的环境感知和自主决策能力，推动智能体从单一任务执行迈向多模态协作。Claude Opus 4 和 Claude Sonnet 4 强化了多模态输入到多步骤任务的连续链式思维。

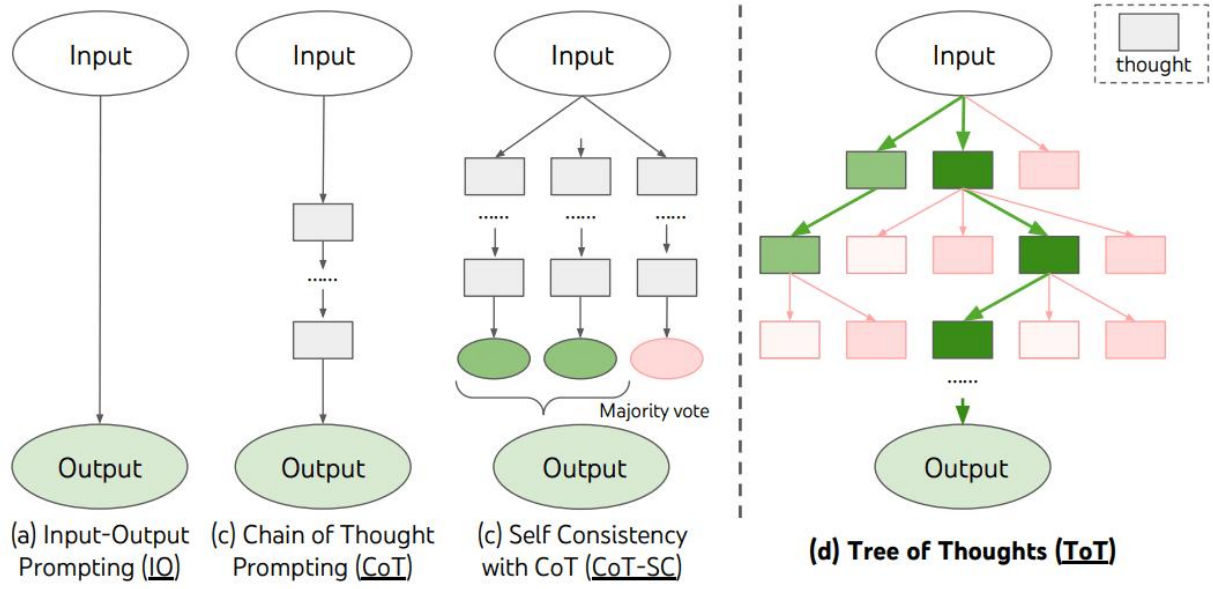
Manus 是中国初创公司 Monica 于 2025 年 3 月发布的通用型 AI 智能体。Manus 的核心能力是基于多模态感知做“真实世界任务”，可以批量读图、提取表格、拍图识别文件结构从而自动生成幻灯片、代码、项目说明等，实现了图文交互，是“认知+行动”的强化。

字节跳动 2025 年通过 Agent TARS、BAGEL 多模态模型、Seed 系列等模型体现出强环境理解+工具控制+多模态融合的综合实力。Agent TARS 从文本语言模型到视觉语言融合感知，在 GUI 中执行真实操作；BAGEL 多模态模型具有统一感知与理解生成能力，向“通用多模态大脑”过渡，增强跨模态理解与关联；Seed 系列能够长上下文记忆，结合网页、地图、图像的输入推理地理、结构、关系等，创新环境理解。

2、规划能力：从线性推理到自主决策

2024 年：基于思维链（CoT）和思维树（ToT）的提示词工程，依赖人工设计流程。比如 CoT 使用线性分步骤推理，应用于数学题、逻辑判断、代码生成等场景，依然依赖于人类提前设定结构模版，无法自动发现逻辑，不具备任务全局感知能力；ToT 运用多方案路径搜索，运用于产品设计、规划任务、文本生成等应用场景，模拟“思考+比较+选择”过程，但是模型本身不具备“主动分支”能力，并未具备真正的自主规划、感知、执行能力。

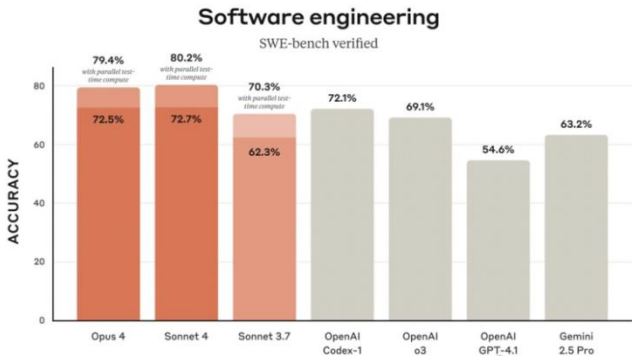
图 21：基于思维树（ToT）的提示词工程



资料来源：《Prompt Engineering Guide》，中国银河证券研究院

2025 年，AI Agent 迈入自主决策体，代表性的技术进展：OpenAI 推出的 Operator 具有自主执行任务拆解、重试、选择路径、调用工具的能力；Anthropic 推出的 Claude Opus 4 在编码和复杂问题的解决方面表现出色，能独立运行长达七小时，具有长期任务规划能力，Extended Thinking 是一种增强的推理能力，使模型回答前更好的分解问题、规划解决方案并寻找不同的解决方法；Manus 采用多智能体架构，能够自动完成复杂任务，例如研究、执行、交付结构化结果；字节跳动 Agent TARS 能够通过自然语言与计算机图形用户界面交互，实现文件管理、浏览器导航等自动化，UI-TARS-1.5 是视觉语言模型，能有效执行多种任务增强模型推理和适应能力。

图 22：Claude Opus 4 在 SWE-bench 测试中领先



资料来源：新智元，中国银河证券研究院

图 23：Claude Opus 4 测试碾压 OpenAI 最强推理模型 o3

	Claude Opus 4	Claude Sonnet 4	Claude Sonnet 3.7	OpenAI o3	OpenAI GPT-4.1	Gemini 2.5 Pro Preview (05-06)
Agentic coding SWE-bench Verified[1,3]	72.5% / 79.4%	72.7% / 80.2%	62.3% / 70.3%	69.1%	54.6%	63.2%
Agentic terminal coding Terminal-bench[1,4]	43.2% / 50.0%	35.5% / 41.3%	35.2%	30.2%	30.3%	25.3%
Graduate-level reasoning GPQA Diamond[1]	79.6% / 83.3%	75.4% / 83.8%	78.2%	83.3%	66.3%	83.0%
Agentic tool use TAU-bench	Retail 81.4% Airline 59.6%	Retail 80.5% Airline 60.0%	Retail 81.2% Airline 58.4%	Retail 70.4% Airline 52.0%	Retail 68.0% Airline 49.4%	– –
Multilingual Q&A MMMLU[3]	88.8%	86.5%	85.9%	88.8%	83.7%	–
Visual reasoning MMMU (validation)	76.5%	74.4%	75.0%	82.9%	74.8%	79.6%
High school math competition AIME 2025[4,5]	75.5% / 90.0%	70.5% / 85.0%	54.8%	88.9%	–	83.0%

Methodology
1. Opus 4 and Sonnet 4 achieve 72.5% and 72.7% pass@1 with bash/editor tools (averaged over 10 trials, single-attempt patches, no test-time compute, using nucleus sampling with a top_p of 0.95).
2. Opus 4 and Sonnet 4 score 39.2% and 35.5% pass@1 with the same agent as non-Claude models, the above reported 43.2% and 35.5% with Claude Code as agent framework.
3. Claude scores on MMMLU are the average over 14 non-English languages.
4. Opus 4 and Sonnet 4 were run on AIME using nucleus sampling with a top_p of 0.95.
5. On SWE-Bench, Terminal-Bench, GPQA and AIME, we additionally report results that benefit from parallel test-time compute by sampling multiple sequences and selecting the single best via an internal scoring model.

资料来源：新智元，中国银河证券研究院

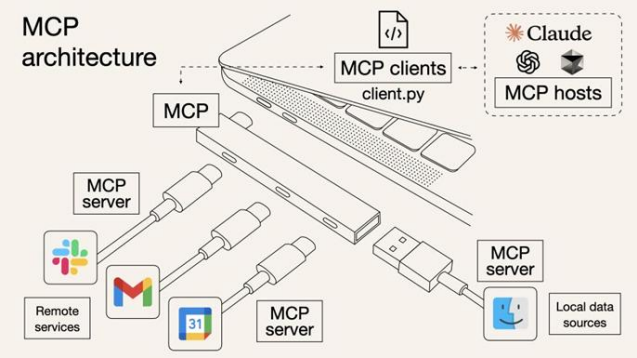
3、工具使用：从 API 调用到交互协作

2024 年：API 调用阶段。早期模型比如 Claude 3.5 Sonnet 通过生成 API 调用文本与外部系统交互，但受限于接口覆盖范围，模型本身不具备“操作工具”的能力，而是通过输出结构话文本，由外部系统解析和执行任务，所有操作必须提示明确触发指令，无法自主“调用工具”，同时，一次智能处理一个工具调用，缺乏任务拆解和工具序列执行的能力。这个阶段的交互是单向的，并不具备持续控制、反馈处理或多工具协同能力。

2025 年：视觉交互与协议标准化（MCP 与 A2A）。工具调用体系在 2025 年迎来关键转折——从结构化接口调用向通用视觉交互与协议化调度推进。首先，视觉交互能力的提升让 AI 能理解内容并进行自动化操作，比如 Anthropic 的 Browser Use，开源网页自动化接口，无需调用预定义 API，允许 Claude 智能体控制浏览器，完成数据提取、填写表单等操作。在 MCP 出现之前，AI 工具调用面临接口碎片化的痛点，每个 LLM 使用不同的指令格式，每个工具 API 也有独特的数据结构，开发者需要为每个组合编写定制化连接代码。MCP（Model Context Protocol），简称模型上下文协议，是 Anthropic 公司于 2024 年 11 月推出的开放标准协议，让各种不同的大型语言模型能够无缝地与各种外部数据源和工具（如业务软件、数据库、代码库等）进行交互操作。开发者只需按 MCP 标准开发一次接口，即可被多个模型调用。OpenAI、Google、阿里、腾讯、百度、字节等巨头已相继宣布接入 MCP 协议。

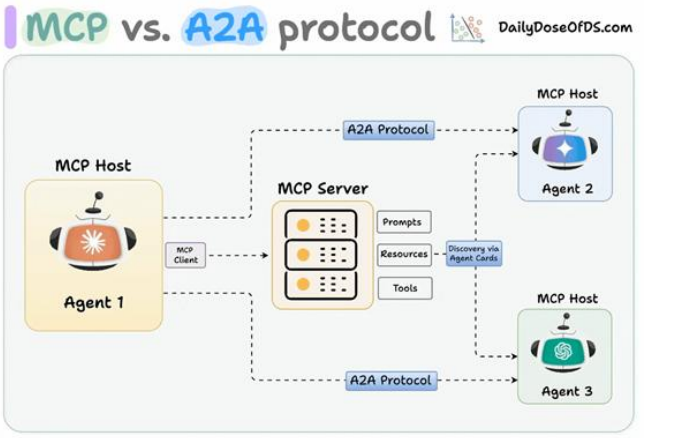
A2A（Agent2Agent）协议是谷歌于 2025 年 4 月推出，作为 AI 智能体间的“通用语言”，允许不同厂商、不同框架的智能体协作，标志着 AI 智能体从“单兵作战”迈向“群体智能”。

图 24：MCP 技术架构三个核心部分



资料来源：腾讯研究院，中国银河证券研究院

图 25：MCP 与 A2A 的协作机制



资料来源：谷歌，中国银河证券研究院

OpenAI 的 Agent SDK 使得开发以标准方式构建、测试、发布 AI Agent 工具，支持多模型调用，强调工具的模块化。AI Agent 从“依赖提示指令”迈向视觉交互和协议标准化的自主协作体，标志着智能体正式具备类人工作流的完整能力。

4、记忆能力：从短期缓存到长期记忆增强

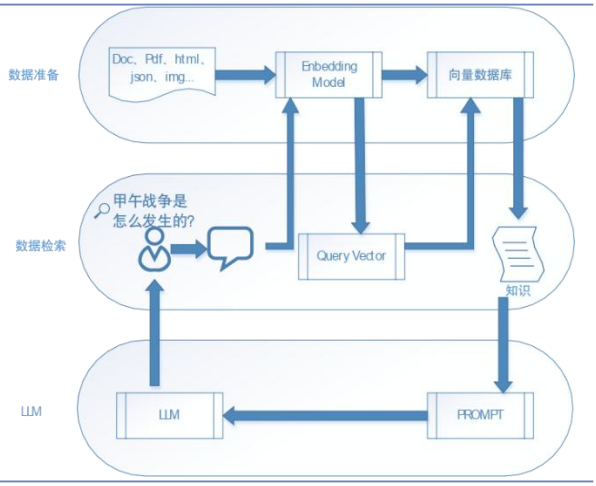
2024 年：短期记忆优化。通过扩展上下文窗口（如 GPT-4 支持 128K Tokens）提升对话连贯性，使得模型可以保持多轮对话的一致性，特别在客户服务、技术支持等场景的体验得到了提升，可以在不分拆摘要的情况下一次性处理长文档、财报等。但对话结束后就“失忆”，无法做到跨会话记忆，并且为被动式记忆，记忆内容需人工重新附加到 Prompt，模型无法记住用户的个性化偏好，处理超长上下文成本较高。

2025 年：长期记忆增强。RAG 技术结合向量数据库，实现外部知识检索；MemGPT 的分层记忆管理支持任务执行中的信息存储与调用。2025 年，AI 智能体实现了从“短期缓存”向“长期记忆”的跃迁，AI 不再仅依赖于上下文窗口，具备跨会话、跨任务的持续记忆能力。**RAG（检索增强生成）**就是通过检索获取相关的知识并将其融入 Prompt，让大模型能够参考相应的知识从而给出合理回答；因此，可以将 RAG 的核心理解为“检索+生成”，向量数据库用来存放向量化之后的知识库，支持信息的持久化存储与调用，并提供向量检索能力，为 RAG 系统实现对知识的初步检索。

MemGPT（Memory-GPT）由伯克利大学的研究团队开发，被誉为最专业的 LLM 记忆管理框架。该技术灵感来源于传统操作系统中的分层内存系统，通过快速内存和慢速内存之间的数据移动提供较大内存资源的可能。MemGPT 也是一个智能管理不同内存层次的系统，以便在 LLM 有限的上下文窗口内有效地提供扩展上下文，并利用中断来管理其自身和用户之间的控制流程。MemGPT 可以分析远超底层 LLM 上下文窗口的大型文档，并且可以创建会话代理，通过与用户的长期交互来记忆、反映和动态发展。

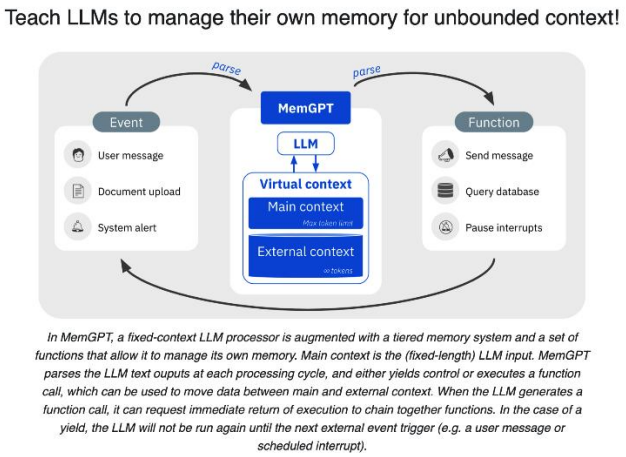
OpenAI 为 GPT-4o 引入 Memory API，使模型能够记住个性化偏好、历史对话和任务进度，在与 Books GPT 的交互中，模型能够记住用户偏好的书籍类型，提供个性化推荐。Anthropic 的 Claude Opus 4 与 Sonnet 4 两款模型引入了持久记忆功能，能够在长时间任务中保持上下文。Manus 的短期记忆方面，在活动对话中缓存上下文的信息，在长期记忆方面，使用向量数据库存储领域知识和历史数据，支持信息的持久化存储与调用，同时捕捉用户偏好，实现个性化交互。

图 26：RAG 系统结合向量数据库的运行流程



资料来源：魔搭社区，中国银河证券研究院

图 27：MemGPT 如何扩展 LLM 的上下文范围



资料来源：智源社区，中国银河证券研究院

（二）全球 AI 大模型动态更新：功能与趋势

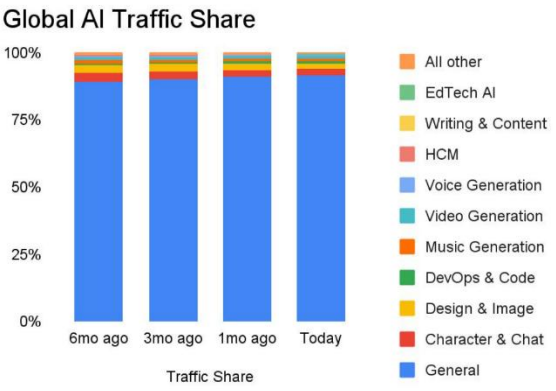
从全球 AI 大模型 2024 年 12 月至 2025 年 5 月的流量趋势来看，不同功能的大模型流量分化较为明显。AI 工具类模型的总流量增速经历了先抑后扬的走势，2025 年 3 月以来的同比增速保持在 20%以上；其中通用类、编程类大模型流量增速最快，2025 年 3 月以来的同比增速分别保持在 25%、75%以上。近一个月流量同比下降比较明显的领域依次为法律、客服、写作、图片生成。

图 28：按大模型功能划分的流量趋势（2024.12-2025.5）

12周变化	12/6	12/20	1/3	1/17	1/31	2/14	2/28	3/14	3/28	4/11	4/25	5/9
General	25%	14%	-8%	4%	14%	18%	23%	25%	55%	46%	34%	22%
性格与对话	1%	7%	18%	8%	-1%	4%	-1%	-5%	-7%	-5%	1%	-2%
设计与图像生成	-6%	-4%	-10%	-4%	-1%	2%	8%	2%	12%	28%	2%	-6%
DevOps & Code 完成	36%	51%	56%	74%	83%	74%	72%	98%	125%	106%	103%	75%
音乐生成	3%	-5%	-3%	-7%	-6%	-3%	-5%	-1%	-4%	7%	12%	10%
视频生成	-21%	-17%	-9%	3%	5%	15%	16%	8%	3%	9%	5%	-5%
语音生成	0%	-4%	-5%	9%	14%	11%	17%	15%	15%	7%	8%	8%
HCM	-13%	-11%	-21%	11%	16%	16%	31%	29%	52%	2%	-2%	-4%
写作与内容	-4%	-7%	-23%	-13%	-7%	-11%	-9%	-6%	14%	-3%	-12%	-11%
教育科技人工智能	-15%	-22%	-18%	-9%	-11%	0%	5%	8%	6%	-7%	1%	-9%
Customer Support & 经验	-9%	-2%	-12%	-9%	-6%	-7%	-7%	-11%	5%	-4%	-11%	-12%
法律	-12%	-14%	-24%	-2%	-4%	-3%	12%	2%	26%	-32%	-70%	-73%
数据分析	24%	3%	-11%	2%	32%	25%	42%	172%	177%	108%	1%	-1%
所有其他	-1%	2%	-17%	14%	21%	24%	36%	33%	62%	12%	-2%	-7%
AI工具总览	22%	12%	-7%	4%	13%	17%	22%	24%	50%	43%	32%	20%

资料来源：similarweb，中国银河证券研究院

图 29：按大模型功能划分的流量份额（2024.12-2025.5）



资料来源：similarweb，中国银河证券研究院

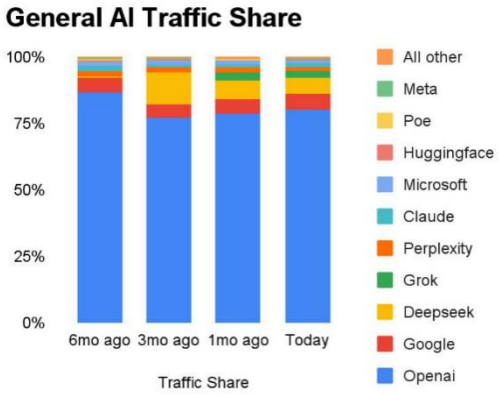
通用类大模型里，2025 年 3 月以来的流量同比增速表现较好的依次为 Grok、Google、Meta；OpenAI、Claude 以及 Microsoft 流量增速保持稳定，分别稳定在 30%、25%、23%左右。Deepseek 流量在 2025 年初经历了爆发式增长，但近一个月同比下降；Huggingface 和 Poe 近一个月流量也同比下降。

图 30：通用类大模型的流量趋势（2024.12-2025.5）

12周变化	12/6	12/20	1/3	1/17	1/31	2/14	2/28	3/14	3/28	4/11	4/25	5/9
Openai	26%	15%	-9%	4%	5%	5%	9%	11%	38%	33%	35%	27%
Google	-2%	-5%	-18%	-11%	-1%	3%	15%	14%	50%	54%	53%	51%
Deepseek	211%	223%	1013%	3029%	17694%	17701%	8658%	6804%	1688%	532%	-6%	-41%
Grok	172%	82%	na	na	na	na	524539%	1096167%	169235%	25911%	11307%	5243%
困惑度	46%	28%	3%	8%	-5%	4%	25%	30%	52%	18%	11%	6%
Claude	22%	1%	-16%	-7%	-18%	-18%	2%	30%	61%	29%	25%	25%
Microsoft	125%	83%	23%	-15%	-14%	-15%	-6%	1%	23%	43%	29%	23%
Huggingface	24%	15%	16%	17%	53%	58%	42%	29%	26%	15%	-14%	-29%
Poe	4%	-12%	-30%	-25%	-29%	-28%	-21%	-16%	-3%	-10%	-6%	-17%
Meta	-6%	21%	-15%	-17%	-6%	5%	5%	-22%	-21%	7%	59%	68%

资料来源：similarweb，中国银河证券研究院

图 31：通用类大模型的流量份额（2024.12-2025.5）



资料来源：similarweb，中国银河证券研究院

（三）AI Agent 商业模式变革：AI Agent 正从“提供工具”向“交付价值”转变

2024Q4，我们认为 AI Agent 模型有望推动 APP 生态逐渐向模型生态转变，AI Agent 应用有望取代 APP 的地位；2025 年至今，我们已经看到多个拥有头部模型的大厂也推出了 AI 智能体开发平台，正在逐步构建起各自的 AI Agent 生态。

伴随着 AI Agent 从“被动工具”迈向“自主决策体”，并且从个体走向协作，AI Agent 的商业模式也将发生变革，AI Agent 应用的竞争点正从“提供工具”向“交付价值”转变。能真正提升下游企业利润的 AI Agent 应用将会胜出，从这个角度来说，对应垂直行业 know how 型卡位公司的投资机会相对提升，能融入智能体能力的 SAAS 企业有望迎来价值重估的机会。

1、AI 智能体开发平台：大厂逐步构建各自的 AI Agent 生态

AI 智能体开发平台具有汇聚流量的优势，如果在上面的 Agent 应用丰富起来，构建起 AI Agent 生态，则有望成为 AI 时代的“安卓”圈，因此平台的开放性和模型技术优势也是关键驱动力。字节跳动发布“Coze”，以零代码方式构建多模态、可记忆的智能体，应用于内容创作、教育辅导等场景，提升营销效率，也可辅助教学；阿里推出“百炼平台”，支持从模型调用到插件集成的全流程智能体开发，应用于电商服务助手、日程管理助手等场景；腾讯的“元器”，结合混元大模型与微信生态，实现一站式创建和分发，应用于客服助手、内容创作等场景，可以提升服务效率，辅助创作。

根据 IDC 报告显示，2024 年中国公有云上大模型调用量达 114.2 万亿 tokens（不包含出海群体使用的海外 MaaS 平台的调用量），按照大模型调用量的市场份额来看，字节火山引擎占据了 46.4%的市场份额，位列第一，其次为百度智能云（19.3%）和阿里云（19.3%）。此外，腾讯云、中国移动、天翼云等其他厂商整体占据 15%的市场份额。2024 年模型调用量仍然以文本类的能力为主，2024Q4 语音类模型调用量也开始增长。预计 2025 年图像、视频类大模型的调用量也将开始起量，成为驱动未来 2 年大模型 tokens 增长的重要力量。

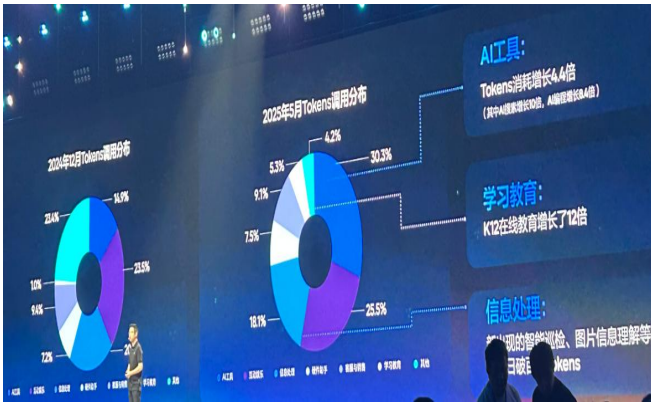
图 32：中国 AI Agent 行业图谱

资料来源：非凡产研，中国银河证券研究院

2、生产力智能体：通用型和 AI 工具类增长最快

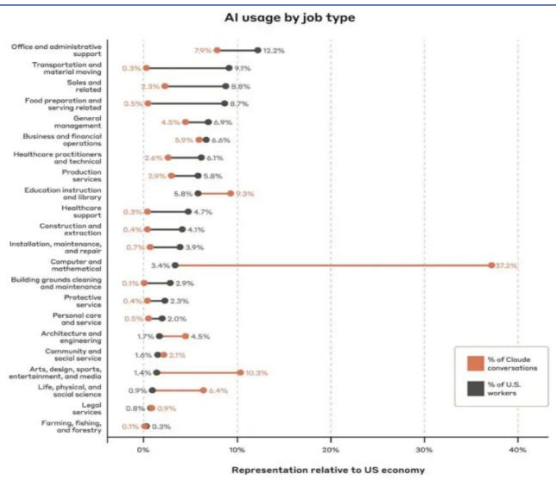
生产力智能体是指主要以提升效率为核心的智能系统，除了通用型智能体外，编程、教育、创作等领域增长最快。根据火山引擎数据，其 AI 工具类场景 tokens 消耗 5 个月增长 4.4 倍，其中 AI 搜索增长 10 倍，AI 编程增长 8.4 倍，K12 在线教育增长 12 倍。智能巡检、视频检索等新场景突破日均百亿 tokens。同时，根据 Anthropic 发布的《AI 经济指数》报告，Claude 模型的使用量里计算机和数学占比 37.2%（编程、开发）、艺术和创作 10.3%、教育/档案 9.3%。

图 33：火山引擎 AI 工具类场景 tokens 消耗 5 个月增长 4.4 倍



资料来源：火山引擎，中国银河证券研究院

图 34：Claude 模型的使用量按工作类型划分：编程开发、艺术创作靠前



资料来源：Anthropic，中国银河证券研究院

通用型：Manus 智能体和 Genspark 智慧搜索核心用于跨领域信息整合以及自动化执行。Manus 的商业模型为基于任务效果“后付费”或“按结果计费”，用户只为成功输出付费，近 20 人团队支撑千万级收入；Genspark 聚焦于知识工作者、学生、程序员等使用场景，支持生活类、

学术类、代码等查询，更注重“本地化内容”以及“中文搜索习惯”，结合多模态能力，支持图片和链接的查询。

创作型：Liblib AI 图片生成服务主要为设计师、插画师、自媒体创作者等提供低门槛、高效率的服务，同时构建内容共创与分享的垂直社区生态，产品涵盖多样化的图像生成方式、丰富的模型资源、活跃的社区互动以及多模态扩展等。2025 年 2 月完成数亿人民币的 A+轮融资，投资方包括洪策资本、顺为资本和巨人网络。

编程型：MGX 智能体旨在通过模拟人类软件开发团队的协作流程，实现从需求分析到部署的全流程自动化开发。产品核心功能与架构为多智能体协作、自然语言编程、全流程自动化开发以及标准化操作流程，适合于无深入编程技能的个人开发者、辅助编程教学的教育与培训、数据分析以及开发创意工具等应用场景。

图 35：中国 AI 产品年收入榜单

排序	产品名称	分类	Web收入（百万美金）	App收入（百万美金）	产品收入（百万美金）	所属公司
1	美图秀秀	图片编辑	0	105	105	美图公司
2	KLING AI	视频生成	92	9	101	快手
3	PictureThis	物体识别	0	100	100	睿琪软件
4	夸克	浏览器	0	83	83	阿里巴巴
5	manus	智能体	41	7	48	蝴蝶效应
6	HeyGen	视频生成	41	0	41	HeyGen
7	Airbrush	视频编辑	4	35	40	美图公司
8	BeautyPlus	图片编辑	0	28	28	美图公司
9	OpusClip	视频编辑	27	0	27	OpusClip
10	PLAUD	文章摘要	0	24	24	PLAUD
11	Genspark	智慧搜索	22	0	22	MainFunc
12	美颜相机	图片编辑	0	22	22	美图公司
13	Clipto.AI	社媒工具	22	0	22	Clipto.AI
14	Openart	图片生成	21	0	21	Openart
15	Fotor	图片编辑	13	6	20	恒图科技
16	Wink	视频编辑	0	19	19	美图公司
17	PolyBuzz	情感陪伴	0	19	19	作业帮
18	Monica	写作软件	15	2	18	蝴蝶效应
19	YouCam Makeup	图片编辑	0	16	16	玩美移动
20	Solvely	教育	0	15	15	加勒比熊猫
21	Filmora	视频编辑	0	15	15	万兴科技
22	Notta	会议助手	10	4	14	思维巡航
23	AI Mirror	形象生成	0	14	14	Polyverse

资料来源：非凡产研，中国银河证券研究院

图 36：豆包大模型日均 tokens 使用量超过 16.4 万亿

资料来源：火山引擎，中国银河证券研究院

3、企业级智能体

企业级智能体指专为企业环境设计和部署的 AI 执行系统，能够在业务流程中模拟人类员工的部分决策和执行行为。当前国内外企业级智能体正加速落地，成为企业提效增能的重要工具。**不同于生产力智能体，企业级智能体的参与者更大比例是原有深耕垂直领域的 SAAS 服务商，它们对业务流程有更深刻的理解，比如用友网络、致远互联、汉得信息、鼎捷数智、凌志软件、焦点科技等。**

图 37：企业架构转型（从 PC 时代到 AI 时代）

资料来源：火山引擎，中国银河证券研究院

图 38：致远互联 CoMi 企业 AI 智能体平台

资料来源：致远互联，中国银河证券研究院

（四）AI Agent 推理算力供需剪刀差测算

我们根据以下假设来测算，未来 3 年全球 AI Agent 应用每日消耗的算力总量。

（1）假设目前全球 AI Agent 日活人数与 AI Web 总日活人数相当，即 2025 年 AI Agent 全球（不含中国）渗透率约为 7%。我们预期 2026-2028 年渗透率分别为 11%、14%、16%。

根据非凡产研统计的全球 AI Web 产品月活数据，海外整体 AI Web 产品活跃度较国内更高，2025 年 5 月前 20 名海外产品合计月活量约为 8.47 亿，国内约为 1.05 亿；海外合计月活量呈现逐月增长的趋势（月增速 4%左右），而国内合计月活量下降比较明显（5 月环比 4 月仍下降约 5%）。

考虑到 AI Web 日常使用率较高，我们假设 DAU（日活）/MAU（月活）=38%，则 AI Agent 全球（不含中国）日活人数约为 3.25 亿。若按照 2026-2028 年渗透率分别为 11%、14%、16%计算，则 AI Agent 全球（不含中国）日活人数分别为 5.24 亿、6.84 亿、8.00 亿。

图 39：2025 年 5 月全球 AI Web 产品月活数据

排名	产品	市场	分类	网址	活跃用户（万人）	环比变化
1	ChatGPT	海外	聊天机器人	chatgpt.com	41,142	-0.54%
2	Gemini	海外	聊天机器人	gemini.google.com	10,820	27.21%
3	deepseek	国内	聊天机器人	chat.deepseek.com	5,607	-6.50%
4	Quizlet	海外	教育	quizlet.com	2,526	-4.82%
5	Grammarly	海外	写作软件	grammarly.com	2,416	-4.47%
6	Grok	海外	个人助理	grok.com	2,399	-18.90%
7	Microsoft Copilot	海外	个人助理	copilot.microsoft.com	2,241	
8	Quillbot	海外	改写润色	quillbot.com	2,108	-1.04%
9	Quizizz	海外	教育	quizizz.com	1,884	7.96%
10	Perplexity	海外	智慧搜索	perplexity.ai	1,848	4.59%
11	NotebookLM	海外	知识管理	notebooklm.google.com	1,668	39.74%
12	Google AI Studio	海外	聊天机器人	aistudio.google.com	1,662	14.47%
13	Claude	海外	聊天机器人	claude.ai	1,479	13.16%
14	Suno	海外	音乐生成	suno.com	1,342	5.81%
15	Gamma	海外	PPT生成	gamma.app	1,019	7.51%
16	Fotor	出海	图片编辑	fotor.com	959	-20.57%
17	ZeroGPT	海外	内容检测	zerogpt.com	931	2.52%
18	Character.AI	海外	情感陪伴	character.ai	921	-1.57%
19	豆包	国内	聊天机器人	doubao.com	894	10.80%
20	Hugging Face	海外	模型训练	huggingface.co	854	-5.01%

资料来源：非凡产研，中国银河证券研究院

图 40：2025 年 5 月中国 AI Web 产品月活数据

排名	产品	分类	网址	活跃用户（万人）	环比变化	所属公司
1	deepseek	聊天机器人	chat.deepseek.com	5,607	-6.50%	深度求索
2	豆包	聊天机器人	doubao.com	894	10.80%	字节跳动
3	百度AI搜索	智慧搜索	chat.baidu.com	752	-8.53%	百度
4	知乎直答	智慧搜索	zhida.zhihu.com	442	4.78%	知乎
5	Kimi	聊天机器人	kimi.moonshot.cn	422	-12.31%	月之暗面
6	腾讯元宝	个人助理	yuanbao.tencent.com	284	-16.96%	腾讯
7	百度橙篇	写作软件	cp.baidu.com	265	42.65%	百度
8	deepseek开放平台	开发工具	platform.deepseek.com	206	-28.16%	深度求索
9	秘塔AI搜索	智慧搜索	metaso.cn	178	-4.33%	秘塔网络
10	文心一言	聊天机器人	yiyan.baidu.com	174	-16.11%	百度
11	纳米AI搜索	智慧搜索	n.cn	169	-11.76%	360
12	火山方舟	模型训练	volcengine.com	166	-1.69%	字节跳动
13	即梦AI	图片生成	jimeng.jianying.com	146	-11.86%	字节跳动
14	沉浸式翻译	翻译	immersivetranslate.com	143	-17.84%	书同文网络
15	有道翻译	翻译	fanyi.youdao.com	137	-2.69%	网易
16	FlowUs	改写润色	flowus.cn	126	-4.21%	云上绿洲
17	扣子	智能体	coze.cn	116	-12.47%	字节跳动
18	熊猫办公	写作软件	tukuppt.com	115	4.39%	迷南文化
19	C知道	智慧搜索	so.csdn.net	111	-7.14%	创新乐知
20	问小白	智慧搜索	wenxiaobai.com	96	5.75%	元石科技

资料来源：非凡产研，中国银河证券研究院

图 41：2025 年 5 月全球 AI APP 产品月活数据

排名	产品	市场	分类	应用名称	活跃用户（万人）	环比变化
1	ChatGPT	海外	聊天机器人	ChatGPT	60,284	13.65%
2	Gemini	海外	聊天机器人	Google Gemini	7,730	270.58%
3	Nova A.I.	海外	聊天机器人	AI Chatbot - Nova	6,713	-0.13%
4	deepseek	国内	聊天机器人	DeepSeek - AI 智能助手	5,619	-74.94%
5	Edge	海外	浏览器	Microsoft Edge: AI Browser	5,438	0.78%
6	百度AI搜索	国内	智慧搜索	百度-DeepSeek满血接入	5,142	-72.46%
7	Photomath	海外	教育	Photomath	4,425	-10.03%
8	豆包	国内	聊天机器人	豆包-抖音旗下AI智能助手	3,988	-73.94%
9	Grok	海外	个人助理	Grok	3,445	35.85%
10	Character.AI	海外	情感陪伴	Character AI: Chat, Talk, Text	3,160	1.91%
11	夸克	国内	浏览器	夸克-你的AI搜索	3,099	-20.69%
12	Talkie	出海	情感陪伴	Talkie AI: Chat With Character	3,011	2.90%
13	ChatOn	海外	聊天机器人	ChatOn - AI Chat Bot Assistant	2,843	-2.12%
14	VivaCut	出海	视频编辑	VivaCut - AI 专业级视频剪辑编辑软件	2,624	0.49%
15	Chat & Ask AI	海外	个人助理	Chat & Ask AI by Codeway	2,623	0.18%
16	Genius	海外	营销工具	Genius: AI Art Photo Editor	2,597	-28.15%
17	B612	海外	图片编辑	B612 AI Photo&Video Editor	2,567	-0.90%
18	UpFoto	海外	图片增强	UpFoto - AI Photo Enhancer	2,509	-9.96%
19	Chatbot AI	海外	写作软件	Chatbot AI-AI Writing for me	2,504	
20	Remini	海外	图片编辑	Remini - AI Photo Enhancer	2,450	-13.49%

资料来源：非凡产研，中国银河证券研究院

图 42：2025 年 5 月中国 AI APP 产品月活数据

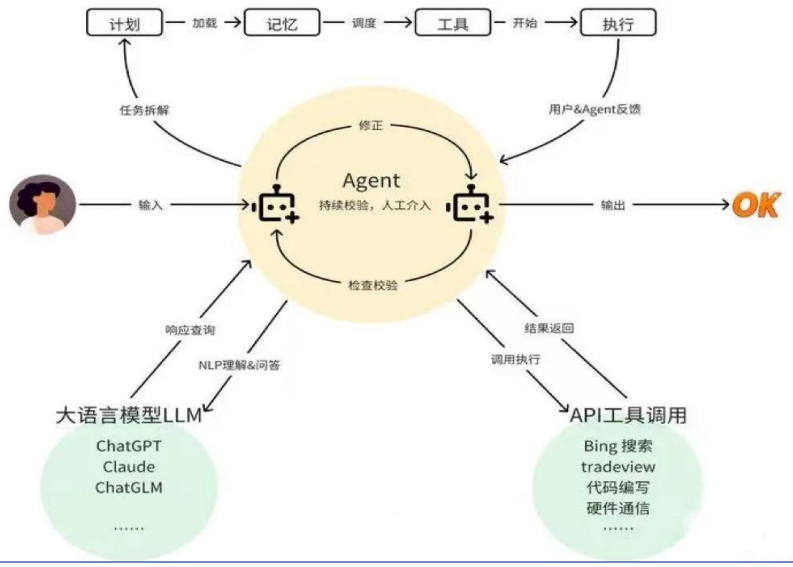
排名	产品	分类	应用名称	活跃用户（万人）	环比变化	所属公司
1	deepseek	聊天机器人	DeepSeek - AI 智能助手	22,426	2.69%	深度求索
2	百度AI搜索	智慧搜索	百度-DeepSeek满血接入	18,669	4.99%	百度
3	豆包	聊天机器人	豆包-抖音旗下AI智能助手	15,301	4.26%	字节跳动
4	作业帮	教育	作业帮-中小学家长作业检查和辅导工具	10,582	-0.15%	作业帮
5	美图秀秀	图片编辑	美图秀秀	9,078	1.71%	美图公司
6	美颜相机	图片编辑	美颜相机	4,103	1.04%	美图公司
7	夸克	浏览器	夸克-你的AI搜索	3,907	0.14%	阿里巴巴
8	腾讯元宝	个人助理	腾讯元宝-发现AI新体验	3,875	14.29%	腾讯
9	Kimi	聊天机器人	Kimi智能助手	3,584	5.28%	月之暗面
10	大学搜题酱	教育	大学搜题酱—教材网课答案全收录	3,302	3.04%	作业帮
11	小猿口算	教育	小猿口算 - 中小学检查作业和学习工具	2,777	27.80%	猿辅导
12	小猿搜题	教育	小猿搜题 - 中小学家长辅导学习和作业检查工具	2,644	4.17%	猿辅导
13	快对	教育	快对AI-智能学习助手	2,481	1.16%	
14	纳米AI搜索	智慧搜索	纳米AI搜索 - DeepSeek R1满血版 已上线	2,042	-27.95%	360
15	文小言	智慧搜索	文小言-原文心一言APP	1,613	3.31%	百度
16	Wink	视频编辑	Wink - 像修图一样修视频	1,549	3.53%	美图公司
17	天天跳绳	动作捕捉	天天跳绳-欢乐AI运动场	1,520	0.47%	微芒科技
18	快影	视频编辑	快影	1,189	7.30%	快手
19	即梦AI	图片生成	即梦AI - 即刻造梦	1,130	7.33%	字节跳动
20	讯飞听见	会议助手	讯飞听见	1,096	2.75%	科大讯飞

资料来源：非凡产研，中国银河证券研究院

（2）假设每个日活用户 2025 年每日仅使用 1 次 AI Agent 应用，2026-2028 年分别增加到 2、3、4 次（场景增加）；且随着应用深度的增加，单次使用 AI Agent 产生的请求次数也会增加，假设 2025-2028 年分别为 50、80、100、120 次。

上文提到，AI Agent 是一种可感知环境变化、独立自主做出决策，并能够主动执行相应行动的 AI 系统，那么每一次自主规划或 API 调用都对应着一次对模型的请求，每循环一次至少对应着 10 次请求，若一个任务拆解为 5 次循环，则至少需要请求 50 次。未来 MCP 的成熟应用有望使循环次数的增长放缓，因为通俗来讲，对于某些问题 Agent 可以直接通过 MCP 接口获得结果而非重新计算。

图 43：AI Agent 工作流程



资料来源：中国银河证券研究院

（3）假设每一次请求需要的 token 数，2025-2028 年分别为 2000、2500、3000、3500。目前主流 AI 智能体完成一个简单的任务（比如制作一张数据图表）大约消耗 10 万 token，比较复杂的任务（比如制作一个 Web 应用）能达到消耗 90 万 token 以上。

我们假设 2025 年平均使用一次 AI Agent 应用消耗 10 万 token，对应 50 次请求，则单次请求消耗 token 数为 2000。未来随着多模态数据更广泛地应用，单次请求消耗 token 数可能大幅增长，比如一张 512*512 像素的图片对应约 334 tokens，Kimi 的 Vision 模型实行按量计费方式，单张图片按 1024 tokens 计算。

（4）假设目前平均单 token 消耗算力约为 8 TFLOPs，随着模型参数量以及多模态数据量的增加，单 token 消耗算力有望逐年增长，我们假设 2025-2028 年平均单 token 消耗算力分别为 8、10、12、14 TFLOPs。

我们基于通义千问 2 多模态代表模型 Qwen2-VL-2B-Instruct 的官方测试结果来推算单 token 消耗算力，推理速度测试基于 NVIDIA A100 80GB，测试了生成 2048 个 token 时，输入长度分别为 1、6144、14336 等 token 时的速度和内存，A100 在 BF16 下的算力为 312 TFLOPs，36 tokens/s 的速度对应单 token 消耗算力约为 8 TFLOPs。

在实际应用中，单 token 消耗算力不是确定值，它与模型版本和用户并发量高度相关，同一个模型在公有云和私有化部署下的值也区别很大。我们可以参考 OpenAI 的毛利率来计算，OpenAI 认为其成本主要是推理计算消耗的算力，根据财务文件，OpenAI 预计其 2025 年毛利率能达到 49%。那么我们根据 OpenAI 目前 token 收费就能计算出单 token 消耗算力值。参考 H200 租赁价格约 3 美元/小时，H200 在 BF16 下的算力为 1979 TFLOPs，那么 3 美元相当于能买到 720 万 TFLOPs。截至 2025 年 6 月的最新价格，价格处于中位的 GPT-4o 输出价格为 15 美元/百万 token，毛利率 49%对应算力成本约 7.5 美元/百万 token，计算出单 token 消耗算力达到 18 TFLOPs。

对于多模态模型而言，多模态数据使得 token 数急剧增加，单 token 消耗算力也会相应增长。比如 OpenAI 在 4 月发布的图像生成模型 GPT-image-1，在价格方面，GPT-image-1 按 token 定价，图像定价是文本的 8 倍：

文本输入 token（提示文本）：每 100 万 token 5 美元；

图像输入 token（输入图像）：每 100 万 token 10 美元；

图像输出 token（生成的图像）：每 100 万 token 40 美元。

图 44：Qwen2-VL-2B-Instruct 的官方测试结果

- 我们使用批量大小为1，并尽可能少地使用GPU进行评估。
- 我们测试了生成2048个token时，输入长度分别为1、6144、14336、30720、63488和129024个token时的速度和内存。
- 2B (transformers)

模型	输入长度	量化	GPU数量	速度(tokens/s)	GPU内存(GB)
Qwen2-VL-2B-Instruct	1	BF16	1	35.29	4.68
		GPTQ-Int8	1	28.59	3.55
		GPTQ-Int4	1	39.76	2.91
		AWQ	1	29.89	2.88
	6144	BF16	1	36.58	10.01
		GPTQ-Int8	1	29.53	8.87
		GPTQ-Int4	1	39.27	8.21
		AWQ	1	33.42	8.18
	14336	BF16	1	36.31	17.20
		GPTQ-Int8	1	31.03	16.07
		GPTQ-Int4	1	39.89	15.40
		AWQ	1	32.28	15.40

资料来源：阿里通义千问官网，中国银河证券研究院

图 45：A100、H100 等算力卡的参数

型号		H100 80GB SXM5	H800 80GB SXM5	H100 80GB PCIe	H800 80GB PCIe	A100 80GB SXM4	A800 80GB SXM4	A100 80GB PCIe	A800 80GB PCIe
应用场景		AI/HPC 科学计算	AI	AI/HPC 科学计算	AI	AI/HPC 科学计算	AI/HPC 科学计算	AI/HPC 科学计算	AI/HPC 科学计算
GPU架构		Hopper	Hopper	Hopper	Hopper	Ampere	Ampere	Ampere	Ampere
GPU核心版本		GH100	GH100	GH100	GH100	GA100	GA100	GA100	GA100
单精度浮点核心(CUDA Core)		16896	16896	14592	14592	6912	6912	6912	6912
显存容量		80GB HBM3	80GB HBM3	80GB HBM2e	80GB HBM2e	80GB HBM2e	80GB HBM2e	80GB HBM2e	80GB HBM2e
显存带宽		3.35TB/s	3.35TB/s	2TB/s	2TB/s	2039 GB/s	2039 GB/s	1935 GB/s	1935 GB/s
NVLink		NVLink 4.0 NVSwitch 900GB/s	NVLink 4.0 NVSwitch 400GB/s	NVLink bridge 600 GB/s	NVLink bridge 400 GB/s	NVLink 3.0 NVSwitch 600 GB/s	NVLink 3.0 NVSwitch 400 GB/s	NVLink bridge 600 GB/s	NVLink bridge 400 GB/s
张量运算核心(Tensor Core)		528 (4代)	528 (4代)	456 (4代)	456 (4代)	432	432	432	432
光线追踪核心 (RT Core)		-	-	-	-	-	-	-	-
性能指标 (PEAK)	FP64浮点 (TFLOPS)	34	1	26	0.8	9.7	9.7	9.7	9.7
	FP32浮点 (TFLOPS)	67	60	51	51	19.5	19.5	19.5	19.5
	FP64 Tensor Core (TFLOPS)	67	60	51	51	19.5	19.5	19.5	19.5
	TF32 Tensor Float (TFLOPS)	989	989	756	756	156	156	156	156
	BF16 Tensor Core (TFLOPS)	1979	1979	1513	1513	312	312	312	312
	FP16 Tensor Core (TFLOPS)	1979	1979	1513	1513	312	312	312	312
	INT8 Tensor Core (TOPS)	3958	3958	3026	3025	624	624	624	624
	INT4 Tensor Core (TOPS)	-	-	-	-	1248		1248	1248
最大功耗		700W	700W	350W	350W	400W	400W	300W	300W

资料来源：昊源诺信，中国银河证券研究院

（4）基于以上假设，我们计算出未来 3 年全球（不含中国）AI Agent 应用每日消耗的推理算力总量，2026-2028 年的增速分别达到 8 倍、3.5 倍、2.5 倍。在 40%的算力利用率下，对应 2025

年 H200 的需求量为 380.54 万块，2026 年 B200 的需求量为 1347.87 万块。AI 芯片性能的进化无法赶超推理算力需求的急剧增长，全球推理算力供需剪刀差不断扩大。

表 2：2025-2028 年全球 AI Agent 应用推理算力需求总量测算

	2025	2026	2027	2028
全球互联网用户（单位亿，不含中国）	46.48	47.64	48.83	50.03
Agent 渗透率	7.00%	11.00%	14.00%	16.00%
Agent 日活人数（单位亿）	3.25	5.24	6.84	8.00
增速		61%	30%	17%
每个日活用户使用次数	1	2	3	4
每个使用的请求数	50	80	100	120
每日请求总数（单位亿）	162.67	838.48	2050.65	3842.15
模型参数	200B	400B	800B	1600B
每次请求需要的 token 数	2000	2500	3000	3500
总 token 数（单位亿）	325332	2096204	6151950	13447526
单 token 消耗算力（TFLOPs）	8	10	12	14
每日 Agent 应用算力消耗（亿 TFLOPs）	**2602656**	**20962040**	**73823400**	**188265370**
增速		**8x**	**3.5x**	**2.5x**
单块 H200 在 BF16 下的 24h 算力（PFLOPs）	170985.60	170985.60	170985.60	170985.60
算力利用率	40%	40%	40%	40%
H200 需求（万块）	380.54	3064.88	10793.80	27526.49
或者：				
单块 B200 在 BF16 下的 24h 算力（PFLOPs）	388800	388800	388800	388800
算力利用率	40%	40%	40%	40%
B200 需求（万块）	167.35	1347.87	4746.88	12105.54

资料来源：中国银河证券研究院

（五）产业链投资机会

1、看好海外算力需求持续增长，建议关注国内 NV 链相关企业。根据英伟达产品计划，其 Blackwell Ultra（GB300）芯片，Blackwell Ultra NVL72 平台将于 2025 下半年推出，在 NVL72 状态下(72 颗芯片互联)AI 性能是 GB200 的 1.5 倍。其下一代 AI 芯片 Rubin，Vera Rubin NVL144 将于 2026 下半年推出，性能是 GB300 NVL72 的 3.3 倍；而更强的 Rubin Ultra NVL576 将于 2027 年下半年推出，性能是 GB300 NVL72 的 14 倍。我们可以看到，AI 芯片性能的进化有其自身的限制。在海外 AI 应用月活量呈现逐月增长的趋势下，结合我们的测算，可以预期 AI 芯片性能的进化无法赶超推理算力需求的急剧增长，全球推理算力供需剪刀差将会不断扩大。

2、字节在 AI 应用生态领域已构建起相对优势，建议关注字节生态合作伙伴。截至 2025 年 5 月底，豆包大模型日均 tokens 使用量超过 16.4 万亿，较去年 5 月刚发布时增长 137 倍。同时，字节火山引擎占据了国内公有云上大模型调用量的 46.4%的市场份额，位列第一。

3、建议关注在 AI Agent 方面布局领先的垂直领域卡位 SAAS 企业。从应用层面来说，伴随着 AI Agent 从“被动工具”迈向“自主决策体”，并且从个体走向协作，AI Agent 的商业模式也将发生变革，AI Agent 应用的竞争点正从“提供工具”向“交付价值”转变。能真正提升下游企业利

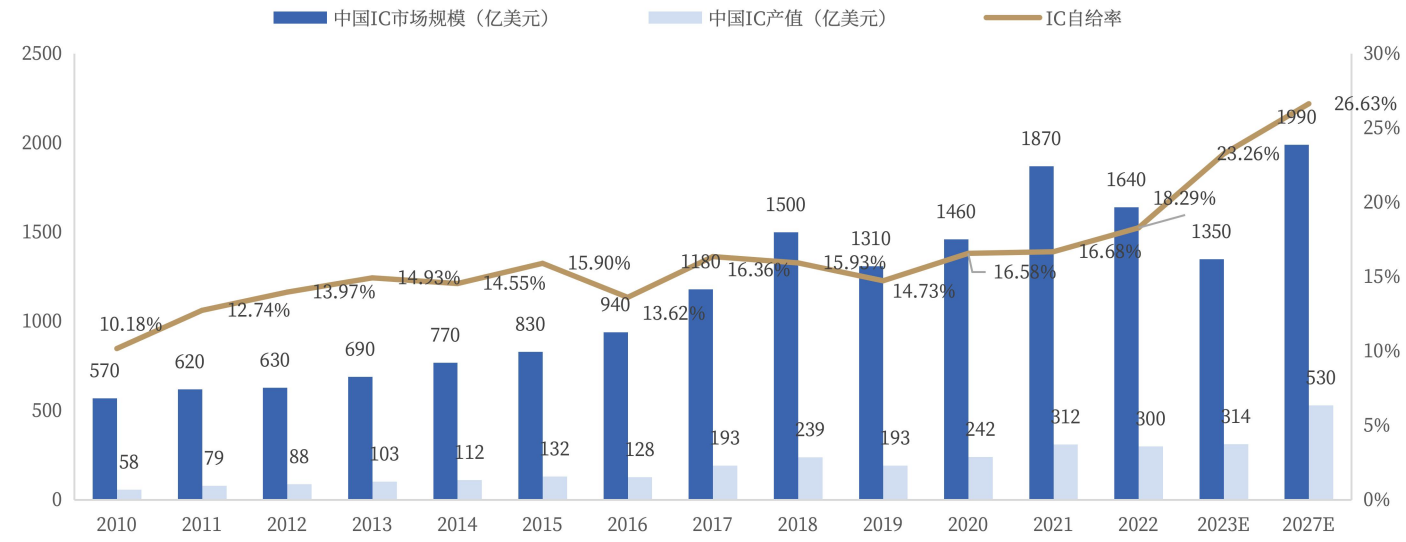
润的 AI Agent 应用将会胜出，从这个角度来说，对应垂直行业 know how 型卡位公司的投资机会相对提升，能融入智能体能力的 SAAS 企业有望迎来价值重估的机会。

三、信创进入深水区，盈利能力有望持续提升

（一）算力国产化自给率不断提升

中国半导体自给率仍然处在较低水平，国产替代潜力巨大。根据 TechInsights 数据显示，2022 年中国 IC 市场规模为 1640 亿美元，中国 2022 年 IC 制造业产值（中资制造业产值与外资制造业产值之和）为 300 亿美元，整体来看 2022 年中国芯片自给率约为 18.3%，其中中国大陆本土企业制造的芯片产值为 152 亿美元，外资制造的芯片产值产值为 148 亿美元，如果只计算中国大陆本土企业制造的芯片，自给率大概只有 9.15%。我国半导体自给率相对 2010 年已经有很大提升，但目前仍处于较低水平，预计 2027 年中国半导体自给率为 26.63%。

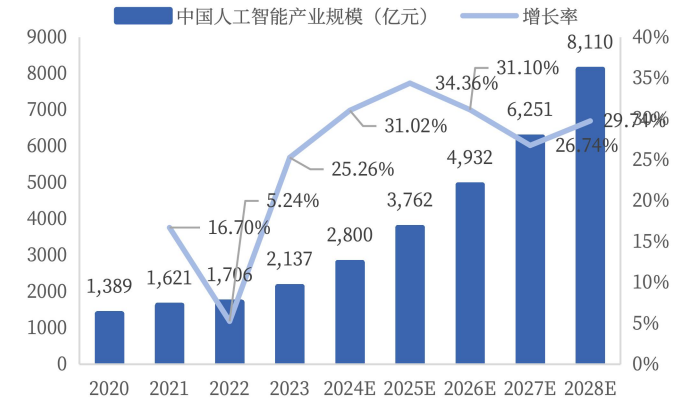
图 46：中国半导体自给率及预测



资料来源：TechInsights、艾瑞咨询，中国银河证券研究院

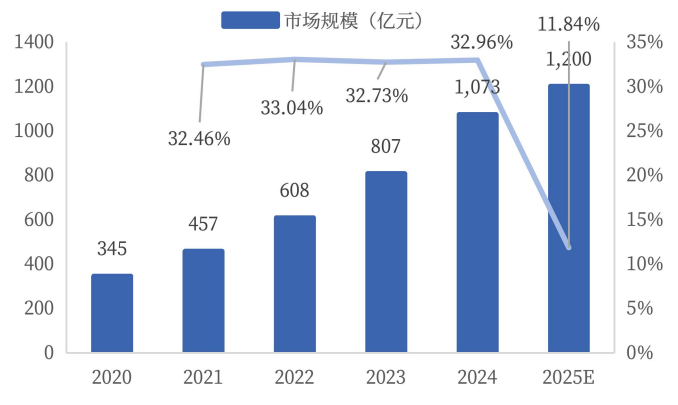
国产 GPU 市场规模猛增，产业规模快速扩容，当下产能仍然是制约国产芯片的瓶颈之一。根据艾瑞咨询和中商产业研究院数据，2025 年中国人工智能市场规模预计为 3762 亿元，预计 28 年市场规模达到 8110 亿元，预计 25 年中国 GPU 市场规模为预计为 1200 亿元，2021-2025 年期间，中国 GPU 市场规模年复合增长率为超 20%，伴随 DeepSeek 推动推理算力需求强劲增长叠加美国对中国科技封锁，国产 GPU 需求高速增长，但目前在产能扩张方面仍存在瓶颈。

图 47：中国人工智能市场规模



资料来源：艾瑞咨询，中国银河证券研究院

图 48：中国 GPU 市场规模及增速



资料来源：中商产业研究院，中国银河证券研究院

（二）信创突围：技术突破与政策驱动市场放量

信创产业是中国为突破关键核心技术“卡脖子”问题、实现科技自立自强而重点发展的战略性产业。以国产化为核心，涵盖基础硬件（如芯片、服务器、电脑、网络交换机等硬件）、基础软件（操作系统、数据库、中间件等）、云基础设施、应用软件等领域，旨在构建自主可控的 IT 全产业链生态。近年来，在政策推动和市场需求的双重助力下，信创产业加速从党政机关向金融、电信、能源等关键行业拓展，逐步形成了“2+8+N”的体系，成为推动数字经济高质量发展、保障国家信息安全的重要引擎。未来随着技术迭代和生态完善，信创产业将进一步释放创新活力，赋能千行百业数字化转型。

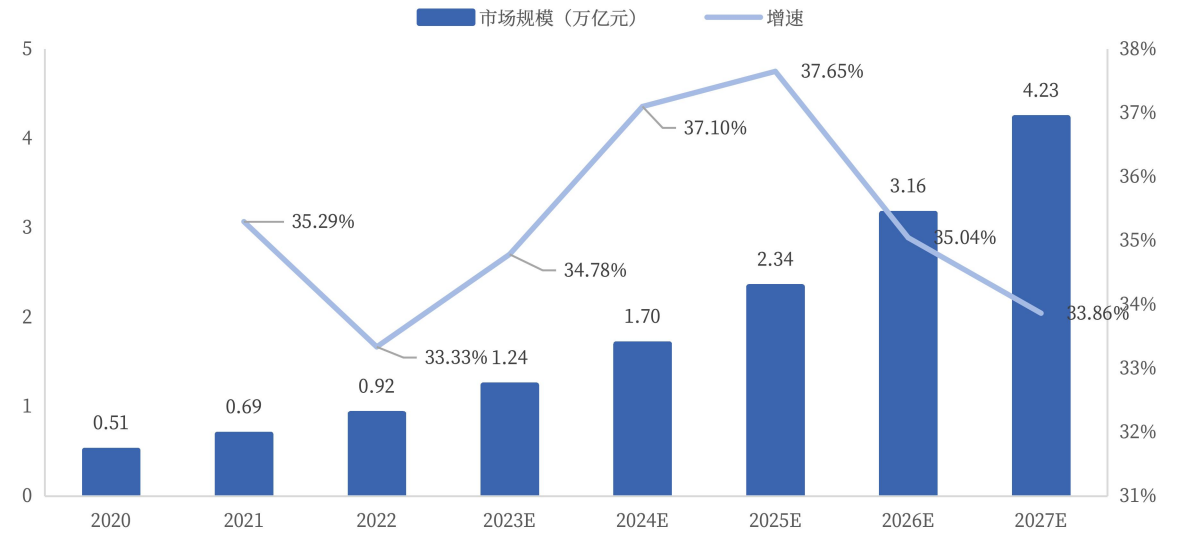
图 49：信创产业链全景图

资料来源：艾瑞咨询，中国银河证券研究院

中国信创市场规模快速增长，25 年迎来信创关键加速之年。2021 年中国信创行业市场规模 6900 亿元，同比增长 35.29%，由于受到宏观环境影响，国央企信息化预算相对缩减，企业对信息化资本投入更为谨慎，22 年信创市场增速有所放缓，行业整体市场规模 9200 亿元，同比增长 33.3%，预计 25 年信创行业市场规模 2.34 万亿元，同比增长 37.65%，我们认为 25 年伴随政策驱动、信创

招标加速推进，将迎来我国信创产业发展的关键节点，25 年“十四五”收官之年，我们预计下半年党政、金融、能源、电新等重点行业信创替代率进一步提升，安全可靠测评及生态适配认证体系将进一步统一，推动市场规范化，信创产业将进入规模化、深度化应用阶段。

图 50：信创行业市场规模与增速



资料来源：亿欧智库，中国银河证券研究院

我国信创产业技术逐渐突破、生态加速协同。信创技术方面，近年在国产 CPU、操作系统、数据库等核心领域逐渐拉近与国际主流水平的差距。国产 CPU 方面，龙芯、鲲鹏等国产 CPU 性能奋起直追，麒麟、统信等国产操作系统市场份额逐渐提升，为信创产业发展提供坚实基础技术底座。生态方面，我国信创呈现出多元协同发展的格局，在数据库领域达梦、人大金仓等数据库逐渐替换，此外云计算与中间件等逐渐形成了稳定的竞争格局，信创产业内部协同发展形成了从硬件到软件的完成产业链生态。

表 3：国产 CPU 主要厂商及技术能力对比

		海光	兆芯	龙芯	飞腾	鲲鹏	申威
	合作方	AMD/中科曙光	VIA/上海国资委	中科院研究所	天津飞腾/CEC	华为	江南计算所/CETC
综合评价	优势	X86 最新授权，性能较强，应用生态丰富	国资委背景，上海地区覆盖广，X86 应用生态丰富	起步最早，适配厂商多，自主化程度高	ARM 前景广阔，产品线丰富，性能不断提升，架构层级授权自主化程度较高	ARM 前景广阔，产品线极其丰富，性能最强，党政+商用市场接受程度高	在军方市场占有率较高，底层应用、超算为主力方向
	劣势	进内核层级授权，自主化程度较低，无桌面授权	早期的 x86 内核层级授权，市场开拓不足	MIPS 生态应用匮乏，性能一般，不利于商品市场拓展	产品起步较晚，性能相对弱势	受制裁中，未来发展存在一定不确定性	超算为主要方向，商用产品开发不足
技术能力	指令体系来源	X86 授权+自研	X86/ARM 授权	MIPS 授权+自研	SPARC/ARM 授权+自研	ARM 授权+自研	Alpha 授权+自研
	授权层级/创新可信程度	X86 内核层级永久授权，自主化	自主层级授权，自主化程度较低	获得 MIPS 指令集修改权限，自	ARM v8 架构层级永久授权，自主化	ARM v8 架构层级永久授权，自	已基本完全实现创新可信

		程度较低		主化程度大	程度高	主化程度高	
	核心技术	AMD 授权 X86 指令集架构，“禅定”X86 CPU	CPU、GPU 芯片组核心技术	MIPS 授权架构的 CPU 及生态圈，跨指令兼容的二进制翻译技术	自研 ARM v8 架构处理器、片上并行系统(PSoC)体系结构	ARM v8 授权架构、达芬奇架构 NPU	申威 64 自主可控架构
	代工厂	格罗方德、三星半导体	台积电	意法半导体	台积电	台积电	中芯国际
	最小制程	14nm	16nm	28nm	16nm	7nm	28nm

资料来源：头豹研究院，中国银河证券研究院

政策层层加码为信创注入新动能，DeepSeek 为信创加速提供强劲动力。自信创产业开始规模化推广以来，政策层面持续加码，为信创产业发展注入了新动能，79 号文明确提出到 2027 年实现国央企信创 100%替代的目标，其中覆盖基础硬件、基础软件、操作系统、中间件等领域。当前面对全球政治格局的变化，关税摩擦常态化，美国对中国科技制裁愈演愈烈，倒逼中国信创加速替代，此外 DeepSeek 的横空出世加速 AI 引用在信创领域落地，为信创产业发展提供了强劲动力，预计 2025 年新创产业将迎来爆发式增长，开启国产化替代的新篇章。

表 4：近年来国家信创政策梳理

时间	政策名称	颁发部门	内容
2025 年 3 月	《2025 年政府工作报告》	国务院	推进高水平科技自立自强。充分发挥新型举国体制优势，强化关键核心技术攻关和前沿性、颠覆性技术研发，加快组织实施和超前布局重大科技项目。优化国家战略科技力量布局，推进科研院所改革，探索国家实验室新型科研组织模式，增强国际和区域科技创新中心辐射带动能力。
2024 年 9 月	《工业重点行业领域设备更新和技术改造指南》	工信部	提出到 2027 年完成约 200 万套工业软件和 80 万台套工业操作系统更新换代任务；力争到 2027 年，80%的规模以上制造业企业基本 实现网络化改造，边缘网关、边缘控制器等产品部署超过 100 万台，“5G+工业互联网”项目数超过 2 万个。
2024 年 3 月	《关于更新中央国家机关台式计算机、便携式计算机批量集中采购配置标准的通知》	中央国家机关政府采购中心	乡镇以上党政机关，以及乡镇以上党委和政府直属事业单位及部门所属为机关提供支持保障的事业单位在采购台式计算机、便携式计算机时，应当将 CPU、操作系统**符合安全可靠测评要求纳入采购需求。**
2023 年 12 月	《台式计算机政府采购需求标准（2023 年版）》	财政部、工信部	为提高台式计算机政府采购需求管理的科学化、规范化水平，进一步落实政府采购公平竞争原则，优化营商环境，营造良好的产业生态，财政部、工业和信息化部制定了《台式计算机政府采购需求标准（2023 年版）》。
2023 年 2 月	《数字中国建设整体布局规划》	国务院	到 2025 年，基本形成横向打通、纵向贯通、协调有力的一体化推进格局，数字中国建设取得重要进展。数字基础设施高效联通，数据资源规模和质量加快提升，数据要素价值有效释放，数字经济发展质量效益大幅增强，政务数字化智能化水平明显提升，数字文化建设跃上新台阶，数字社会精准化普惠化便捷化取得显著成效，数字生态文明建设取得积极进展，数字技术创新实现重大突破，应用创新全球领先，数字安全保障能力全面提升，数字治理体系更加完善，数字领域国际合作打开新局面。

2022 年 1 月	《“十四五”数字经济发展规划》	国务院	到 2025 年，数字经济迈向全面扩展期，数字经济核心产业增加值占 GDP 比重达到 10%，数字化创新引领发展能力大幅提升，智能化水平明显增强，数字技术与实体经济融合取得显著成效，数字经济治理体系更加完善，我国数字经济竞争力和影响力稳步提升。着力提升基础软硬件、核心电子元器件、关键基础材料和生产装备的供给水平，强化关键产品自给保障能力。
2021 年 12 月	《“十四五”推进国家政务信息化规划》	国家发展改革委	到 2025 年，政务信息化建设总体迈入以数据赋能、协同治理、智慧决策、优质服务为主要特征的融慧治理新阶段，跨部门、跨地区、跨层级的技术融合、数据融合、业务融合成为政务信息化创新的主要路径，逐步形成平台化协同、在线化服务、数据化决策、智能化监管的新型数字政府治理模式，经济调节、市场监管、社会治理、公共服务和生态环境等领域的数字治理能力显著提升，网络安全保障能力进一步增强，有力支撑国家治理体系和治理能力现代化。

资料来源：头豹研究院，中国银河证券研究院

（三）GPU 国产厂商性能及制程对比

高端芯片受禁令影响，国产 AI 芯片奋起直追。虽然目前全球 AI 芯片市场被英伟达垄断，国内 AI 芯片投资热度高企，芯片厂商研发投入持续加码。目前国内 AI 芯片第一梯队有华为、寒武纪、海光信息等，单卡算力正在逐渐缩小与高端芯片差距。国产 AI 算力芯片中，华为昇腾 910b 单卡算力达到 640 TOPS（INT 8）；寒武纪思元 370 单卡算力 256 TOPS（INT 8）。

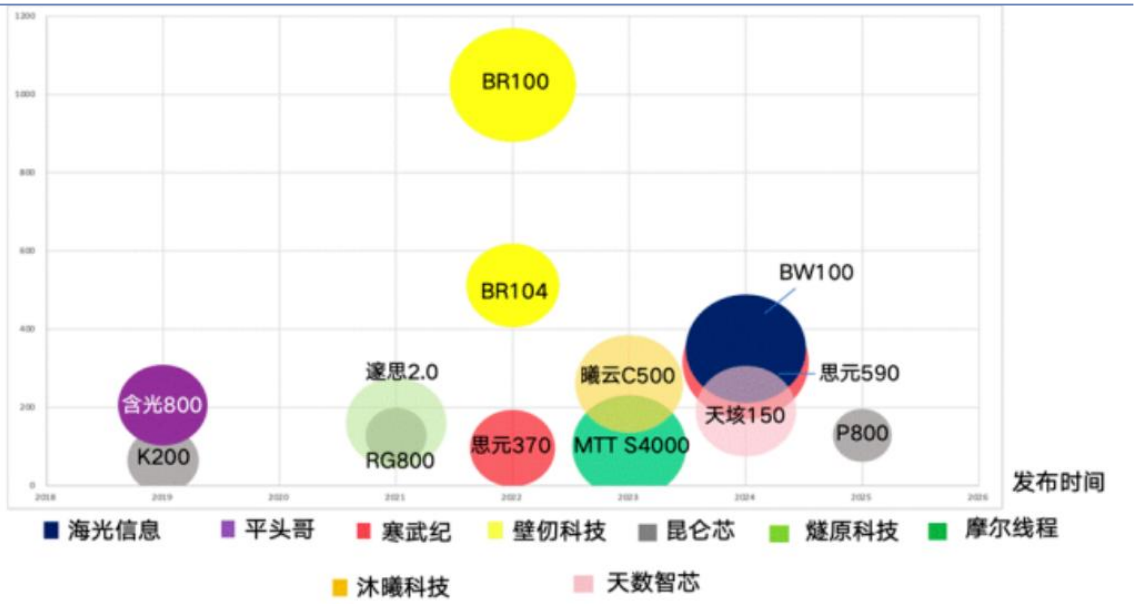
表 5：国产人工智能芯片性能参数对比

	产品	整型 INT8 算力	制程	显存类型	显存容量	显存带宽	互联技术	生态	最大热设计功耗 TDP
华为	昇腾 910b	640 TOPS	7nm	HBM2e	64GB	392GB/s	PCIe 5.0	MindSpore	400W
寒武纪	MLU 370-X8	256TOPS	7nm	LPDDR5	48GB	614.4GB/s	PCIe 4.0	CUDA	250W
壁仞科技	BR100	2048 TOPS	7nm	HBM2e	64GB	2.3TB/s	PCIe 5.0	BIRENSUPA	550w
燧原科技	云燧 T21	320TOPS	7nm	HBM2e	32GB	1.8TB/s	PCIe 4.0	Topsride	400W
摩尔线程	MTT S4000	200TOPS	7nm	GDDR6	48GB	0.8TB/s	MTLink	CUDA	450W
天数智芯	天垓 150	384TOPS	7nm	HBM2e	64GB	1.6TB/s	PCIe 4.0	CUDA	450W
沐曦	MXC 500	480TOPS	7nm	HBM2e	64GB	1.8TB/s	MetaXLink	CUDA	450W
昆仑芯	RB200	256TOPS	7nm	GDDR6	-	0.5TB/s	K-LINK	-	150W
平头哥	含光 800	825TOPS	12nm	-	64GB	-	PCIe 4.0	-	276W

资料来源：半导体综研，中国银河证券研究院

整体来看国产 GPU 单卡性能与英伟达 H100 等中高端产品性能接近。寒武纪的思元 590 和沐曦科技的曦云 MXC500 在算力和功耗上领先，发布时间集中在 2023-2024 年。整体来看，国产芯片的算力水平与英伟达中端产品如 H100 接近，但由于各家厂商技术路线多样性，TDP 热设计功耗方面较为分散，中国 AI 芯片目前逐渐进入爆发期，但与英伟达仍存在代际差。

图 51：国产主流芯片 FP16 算力对比（圆圈大小代表芯片 TDP）



资料来源：半导体综研，中国银河证券研究院

国产 GPU 处在“可用”到“好用”关键转型期。虽然目前国产 AI 芯片在特定领域实现突破，但构建完整技术生态仍需一定周期，AI 大模型算力需求缺口以及美国对中国科技限制加剧，正在倒逼产业加速迭代，也推动全产业链生态的协同进化。

（四）信创基础软件加速渗透

我国信创基础软件生态日趋完善，操作系统与数据库蓬勃发展。

1）数据库方面：形成了传统数据库厂商、云数据库厂商与新兴数据库厂商生态协同发展的格局，其中传统数据库厂商如达梦、人大金仓、南大通用等在信创中占据一定市场份额；互联网等大厂的云数据库如阿里云、腾讯云、华为云等凭借强大的资金支持和技术实力，不断加速市场化竞争；其他新兴数据库厂商如平凯星辰的 TiDB 开源分布式数据库在金融和互联网领域得到广泛认可。

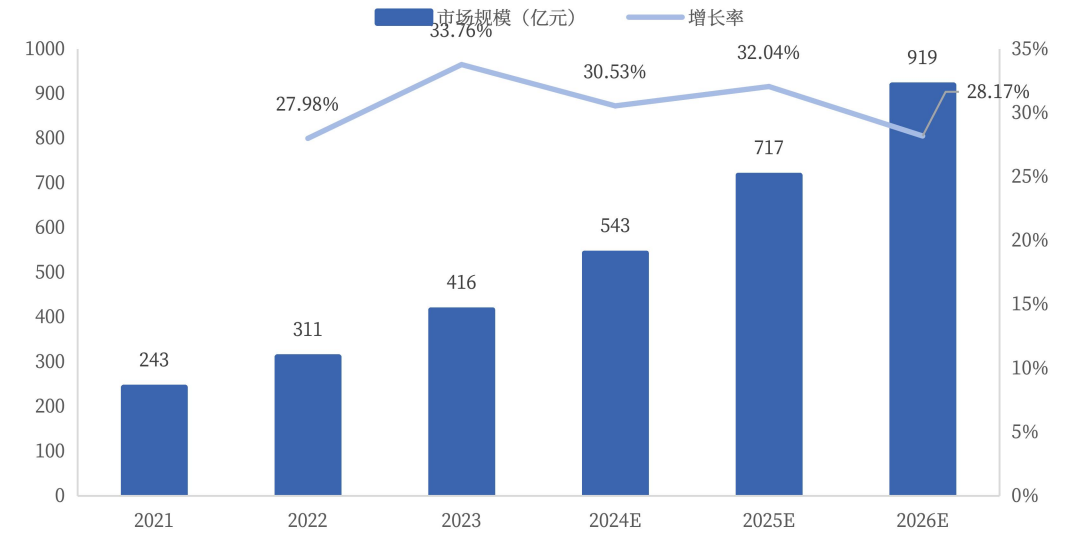
2）操作系统方面：操作系统逐渐形成了麒麟软件与统信软件双寡头竞争格局，占据商业版国产操作系统 90%的市场份额，麒麟软件连续 12 年在中国 Linux 市场占有率第一，又在国防、金融等关键领域具备深厚积累；统信作为桌面操作系统，在功能方面基本上已经追平了 Windows 7，部分功能赶超 Windows 10，但与 Window 11 仍有差距。此外，华为欧拉与鸿蒙分别聚焦服务器云计算与智能终端，根据 CounterPoint Research 数据，在中国市场，2024 年第四季度，鸿蒙系统在中国手机系统中的市场份额达到 19%，iOS 为 17%，鸿蒙连续四个季度超越苹果 iOS，稳居中国市场第二大手机系统。

图 52：中国信创国产基础软件代表厂商

资料来源：第一新声，中国银河证券研究院

预计 2026 年中国信创数据库市场规模达到 919 亿元。2023 年中国数据库行业市场规模为 437 亿元，伴随国内数字化转型加速假设，基础软件领域迎来了高速发展，数据库作为信息系统核心软件以及信创的关键环节，迎来发展黄金期，根据第一新声数据，预计 2026 年市场规模有望达到 919 亿元，同比增长 28.17%。

图 53：中国信创数据库市场规模及增速



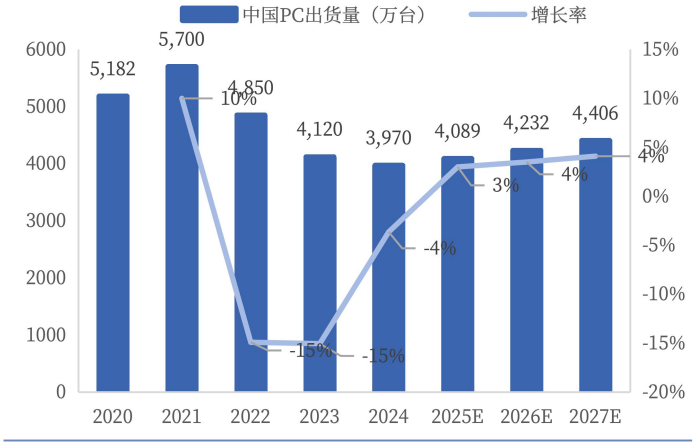
资料来源：第一新声，中国银河证券研究院

国产数据库产业生态逐渐完善。目前，国产数据库于政务、金融、能源等关键领域应用广泛，在互联网、制造业等行业也渐露锋芒。在技术创新层面，国产数据库在分布式架构、云原生、AI 优化等领域持续创新，不断增强事务处理能力、保障数据一致性、提升智能化水平。而且云化趋势显著，云原生数据库作为未来重要发展方向，能更好适配云原生架构，实现轻量化与弹性扩展。国产数据库与主流操作系统、中间件及云平台深度集成，形成完整生态系统。开源社区如 TiDB 社区活跃，吸引全球开发者参与贡献，推动生态繁荣，有助于提升国产数据库的国际竞争力，拓展海外市场。

达梦数据是国内领先的数据库产品开发服务商，是国内数据库基础软件产业发展的关键推动者。公司面向大中型公司、企事业单位、党政机关提供各类数据库软件及集群软件、云计算与大数据产品、数据库一体机等一系列数据库产品及相关技术服务，致力于成为国际顶尖的全栈数据产品及解决方案提供商，在信创数据库领域占据主导地位。

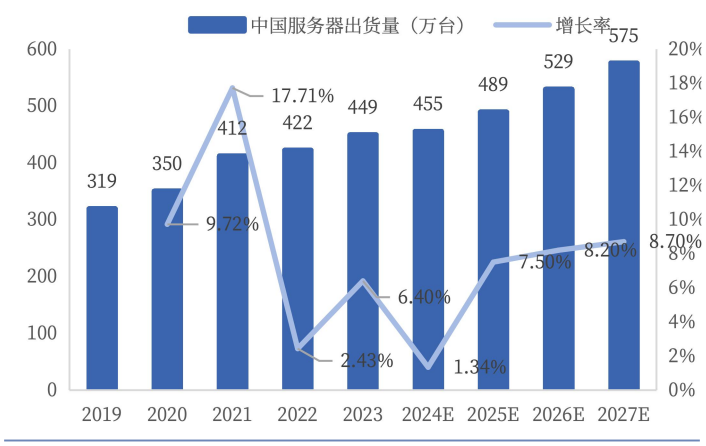
信创操作系统 2027 年市场规模预计达到 610 亿元。若仅考虑 PC 与服务器出货量，假设 2027 年 PC 操作系统单价 600 元，服务器操作系统单价 6000 元，预计 2027 年操作系统市场规模 610 亿元，同比增长 7%。

图 54：中国 PC 出货量及预测



资料来源：Canalys，中国银河证券研究院

图 55：中国服务器出货量及预测



资料来源：IDC、中商产业研究院，中国银河证券研究院

表 6：国产操作系统市场规模及预测

	2020	2021	2022	2023	2024	2025E	2026E	2027E
中国 PC 出货量（万台）	5182	5700	4850	4120	3970	4089	4232	4406
增长率		10.00%	-14.91%	-15.05%	-3.64%	3.00%	3.50%	4.10%
单价（元）	500	500	500	500	550	550	600	600
PC 操作系统市场规模（亿元）	259.09	285.00	242.50	206.00	218.35	224.90	253.93	264.34
中国服务器出货量（万台）	350	412	422	449	455	489	529	575
增长率		17.71%	2.43%	6.40%	1.34%	7.50%	8.20%	8.70%
单价（元）	5000	5000	5500	5500	5500	6000	6000	6000
服务器操作系统市场规模（亿元）	175.00	206.00	232.10	246.95	250.25	293.48	317.54	345.17
总市场规模（亿元）	434.09	491.00	474.60	452.95	468.60	518.38	571.47	609.51

资料来源：Canalys、IDC、中商产业研究院，中国银河证券研究院

当前国产操作系统逐渐从“可用”向“好用”发展。目前来看以桌面操作系统为例，Windows 依旧占据中国市场很大份额，国产操作系统在消费端的市占率还很低，部分行业还存在大量 Windows 系统未替换，国产操作系统还有很大空间。

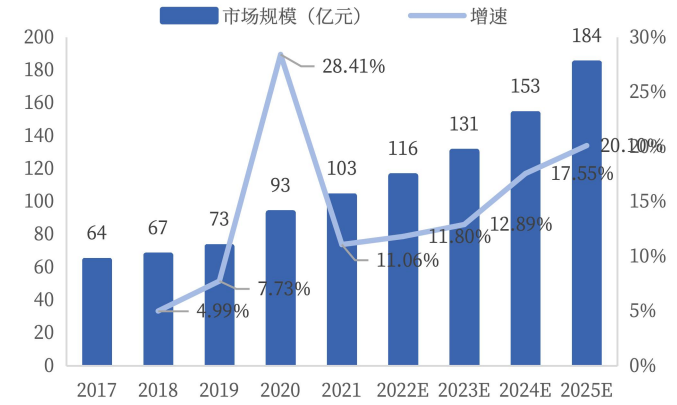
美国升级对华 EDA 管控，国产 EDA 迎新变局。5 月 28 日，美国商务部向 EDA 厂商如新思科技（Synopsys）、楷登电子（Cadence）、西门子 EDA（Siemens EDA）发出通知升级出口禁令。EDA 位于半导体产业链上游，是整个芯片产业的基础，虽然只占芯片成本不到 2%，但是具有杠杆效应，虽然 EDA 市场只有几百亿美元的规模，但是每年撬动约 5000 亿美元的半导体市场。

图 56：EDA 位于集成电路产业链上游支撑关键环节

资料来源：华大九天招股说明书，中国银河证券研究院

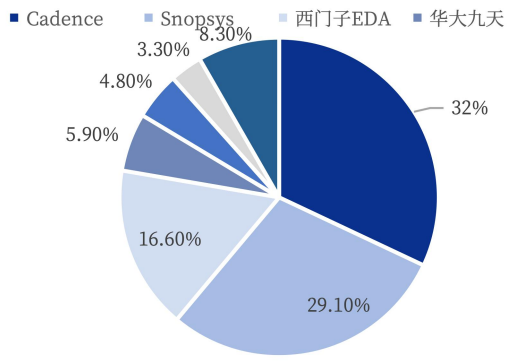
中国 EDA 市场空间超百亿元，国产替代空间广阔。2023 年，新思科技（Synopsys）、楷登电子（Cadence）、西门子 EDA 三家就占了走中国接近 80%的份额，作为国产 EDA 龙头华大九天 2023 年只有 5.9%的市场份额，美国对中国 EDA 断供后，本土 EDA 厂商如华大九天、概伦电子、广立微等迎来新的发展机遇，并购浪潮下叠加国产替代巨大空间，国产 EDA 迎来黄金发展期。

图 57：中国 EDA 市场规模及增速



资料来源：亿渡数据，中国银河证券研究院

图 58：中国 EDA 市场竞争格局



资料来源：华经产业研究院，中国银河证券研究院

1）华大九天：国产 EDA 龙头，公司主要从事 EDA 工具软件的开发、销售及相关服务。EDA 工具是集成电路领域的上游基础工具，应用于集成电路设计、制造、封装、测试等产业链各个环节，是集成电路产业的战略基础支柱之一。公司主要产品包括模拟电路设计全流程 EDA 工具系统、数字电路设计 EDA 工具、平板显示电路设计全流程 EDA 工具系统和晶圆制造 EDA 工具等 EDA 工具软件，并围绕相关领域提供技术开发服务。公司相关产品和服务主要应用于集成电路设计及制造领域。

2）概伦电子：是一家具备国际市场竞争力的 EDA 企业，拥有领先的 EDA 关键核心技术，致力于提高集成电路行业的整体技术水平和市场价值，提供专业高效的 EDA 流程和工具支撑。公司通过 EDA 方法学创新，推动集成电路设计和制造的深度联动，加快工艺开发和芯片设计进程，提高集成电路产品的良率和性能，增强集成电路企业整体市场竞争力。

3）广立微：领先的集成电路 EDA 软件与晶圆级电性测试设备供应商，专注于芯片成品率提升和电性测试快速监控技术。公司提供 EDA 软件、电路 IP、WAT 测试设备以及与芯片成品率提升技术相结合的全流程解决方案，在集成电路从设计到量产的整个产品周期内实现芯片性能、成品率、稳定性的提升，成功打破了集成电路成品率提升领域长期被国外产品垄断的局面。

我们认为，EDA 作为贯穿整个半导体工艺流程的关键工业软件，目前全球 EDA 厂商新思、楷登与西门子占据超过 80%中国市场份额，美国对华半导体产业制裁进一步加码，倒逼国产 EDA 厂商加速突围，提升国产化率，EDA 目前被纳入到“十四五”规划重点卡脖子突破领域，政策上有望持续加码推动国产 EDA 实现破局，建议关注“十五五”规划中 EDA 等领域政策以及产业龙头企业和行业内收并购行为。

四、溢出机会：能源-算力协同革命

（一）算力-绿电绑定模式未来有望加速推广

2023 年以来国家不断出台政策推进算力+绿电协同发展。2023 年，政策明确提出到 2025 年底，算力电力双向协同机制初步形成，国家枢纽节点新建数据中心绿电占比需要超过 80%。2024 年来，国家发改委、工信部、能源局等部门印发《数据中心绿色低碳发展专项行动计划》提出，到 2025 年底，全国数据中心整体上架率不低于 60%，平均电能利用效率降至 1.5 以下，可再生能源利用率年均增长 10%，平均单位算力能效和碳效显著提高。2024 年，国家发改委、数据局、工信部联合印发《国家数据基础设施建设指引》，提出要加强大型风光基地和算力枢纽节点协同联动，把绿色电力转换成绿色算力，并提出要积极消纳风光绿电助力碳达峰碳中和，借助源网荷储模式，加强数据中心能源智慧管理，通过监测分析与负荷预测优化电力系统效率，探索绿电直供。国家发展改革委、国家能源局 2024 年 12 月 20 日印发《电力系统调节能力优化专项行动实施方案（2025—2027 年）》则提出，要推进算力与绿色电力融合，促进绿电消纳。

表 7：国内部分推进算力+绿电协同发展的政策

名称	时间	相关内容
《算力基础设施高质量发展行动计划》	2023.10	全面提升算力设施能源利用效率和算力碳效水平
《深入实施“东数西算”工程加快构建全国一体化算力网的实施意见》	2023.12	2025 年底，算力电力双向协同机制初步形成，国家枢纽节点新建数据中心绿电占比超过 80%。支持国家枢纽节点地区利用“源网荷储”等新型电力系统模式。
《数据中心绿色低碳发展专项行动计划》	2024.7	"东数西算"与算力梯度布局，存改结合淘汰低效设备，新建项目 1.25 能效红线+80%绿电，余冷余热回收+数电联营，液冷/AI 节能/模块化电源协同
《国家数据基础设施建设指引》	2024.12	构建数据底座促流通，优化算力布局强协同，升级网络传输保高效，融合绿电能效助双碳，拓展行业应用筑安全。
《电力系统调节能力优化专项行动实施方案（2025—2027 年）》	2024.12	统筹测算调节需求，推进抽蓄火电调峰改造，建设新型储能电站，强化水电光热气电协同，优化电网资源配置，深挖虚拟电厂负荷潜力，构建源网荷储高效体系。
《关于有序推动绿电直连发展有关事项的通知》	2025.5	新能源不直接接入公共电网，通过直连线路向单一电力用户供给绿电，可实现供给电量清晰物理溯源。并网型绿电直连项目享有平等的市场地位，按照《电力市场注册基本规则》进行注册，原则上应作为整体参与电力市场交易。

资料来源：国务院，国家发改委，国家数据局，工信部，能源局，中国信通院，中国政府网，中国银河证券研究院

2025 年 5 月绿电直连政策出台，将加速算力+绿电绑定模式，企业绿色用能比例有望大幅增加，相关绿电交易企业迎来重大发展机遇。2025 年 05 月 21 日，发改委、能源局印发《关于有序推动绿电直连发展有关事项的通知》，促进新能源就近就地消纳，更好满足企业绿色用能需求。政策指出：

1、绿电直连是指风电、太阳能发电、生物质发电等新能源不直接接入公共电网，通过直连线路向单一电力用户供给绿电，可实现供给电量清晰物理溯源的模式。其中，直连线路是指电源与电力用户直接连接的专用电力线路。按照负荷是否接入公共电网分为并网型和离网型两类。并网型项目作为整体接入公共电网，与公共电网形成清晰的物理界面与责任界面，电源应接入用户和公共电网产权分界点的用户侧。

2、并网型项目应按照“以荷定源”原则科学确定新能源电源类型和装机规模。现货市场连续运行地区可采取整体自发自用为主，余电上网为辅的模式；现货市场未连续运行地区，不允许向公共电网反送。项目整体新能源年自发自用电量占总可用发电量的比例应不低于 60%，占总用电量的比例应不低于 30%，并不断提高自发自用比例，2030 年前不低于 35%。上网电量占总可用发电量的比例上限由各省级能源主管部门结合实际确定，一般不超过 20%。

3、并网型绿电直连项目应通过合理配置储能、挖掘负荷灵活调节潜力等方式，充分提升项目灵活性调节能力，尽可能减小系统调节压力。**项目规划方案应合理确定项目最大的负荷峰谷差率，项目与公共电网交换功率的电力峰谷差率不高于方案规划值。**在新能源消纳困难时段，项目不应向公共电网反送电。项目应按照有关管理要求和技术标准做好无功和电能质量管理。

4、并网型绿电直连项目享有平等的市场地位，按照《电力市场注册基本规则》进行注册，原则上应作为整体参与电力市场交易，根据市场交易结果安排生产，并按照与公共电网的交换功率进行结算。并网型绿电直连项目以项目接入点作为计量、结算参考点，作为整体与公共电网进行电费结算。

建议关注绿电交易相关公司：朗新集团(300682.SZ)、国能日新(301162.SZ)。

朗新集团：2024 年公司在广东、江苏、浙江、山东、四川等多省开展电力市场化交易，AI 辅助电力交易取得显著成果，年度整体交易电量超过 19 亿度，同比增长超过 5 倍。公司优势在于十几年的生活缴费和充电桩运营在负荷端积累了大量客户资源。

国能日新：新能源功率预测系统龙头，业绩近 3 年稳定增长；公司优势在于电源侧的客户资源和多年的功率预测技术积累。电力交易布局全面：电力交易数据服务、电力交易辅助决策支持平台、电力交易托管服务、储能管理系统等。2024 年公司电力交易相关产品已实现在山西、山东、甘肃、广东和蒙西的布局并在上述省份均已陆续应用于部分电力交易客户。

表 8：绿电直连政策对国能日新的深度利好

业务领域	政策关联点	商业价值
功率预测系统	直连项目需实时匹配源荷波动（自发自用率≥60%），依赖高精度发电/负荷预测	新能源功率预测产品需求激增，AI 大模型“旷冥”可提升预测效率 60 倍
储能协同方案	政策强制配置储能调节波动，要求优化充放电策略	光储智慧运营系统渗透率提升，通过负荷预测优化储能调度
电力交易服务	直连项目须整体注册参与电力市场，余电交易依赖电价预测与申报策略	AI 交易平台（自动申报、偏差控制）客户基数扩大，服务费率变现空间提升
零碳园区方案	出口型园区需满足绿电占比要求（2025≥30%）	“源-网-荷-储-碳”一体化平台加速落地，绿电交易撮合服务增收

资料来源：国家发改委，中国银河证券研究院

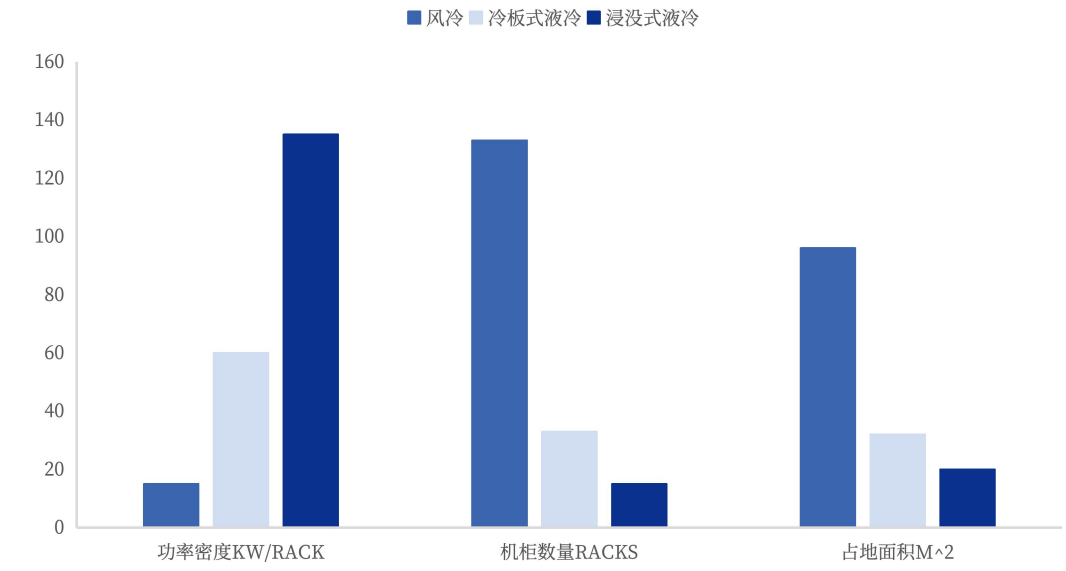
（二）AI 算力驱动液冷渗透率提升，从“可选”向“必选”转变

数据中心 PUE 要求愈发严苛。随着云计算、大数据、人工智能、元宇宙等信息技术的快速发展和传统产业数字化的转型，数据呈现几何级增长，算力和硬件部分能耗也在持续增加，而在“双碳”政策的持续推进下，国家、地方政府、企业层面均在积极推动绿色低碳转型和可持续发展，通讯领域对数据中心节能降耗要求越来越严格。

液冷未来有望逐渐替代风冷，成为 AI 服务器、数据中心标配。AI 训练及推理应用、超算等高算力业务需求持续推升,由此带来的芯片性能需求、服务器功率需求不断提高。场景侧，英伟达 2024 GTC 大会上推出 GB200 NVL72 采用液冷散热方式,并且黄仁勋表示浸没式液冷技术将是未来方向,将带动整片散热市场迎来全面革新。

我们认为，人工智能浪潮下，对算力需求进一步提升，液冷预计将成为最优冷却方案，未来中国液冷服务器市场有望进一步打开竞争格局，产业相关上市公司将受益。目前，中国液冷服务器普及率不足 5%，径普及率并不高。受制于：1）数据中心国家 PUE 标准收紧；2）受制于面积等因素，机柜密度逐渐提升；3）温度过高，芯片故障率升高等客观因素，未来液冷服务器将成为调和快速的算力需求与有限数据中心承载力的共识方案。

图 59：风冷与液冷散热能力对比



资料来源：中兴通讯《液冷技术白皮书》，中国银河证券研究院

表 9：全国主要数据中心 PUE 要求

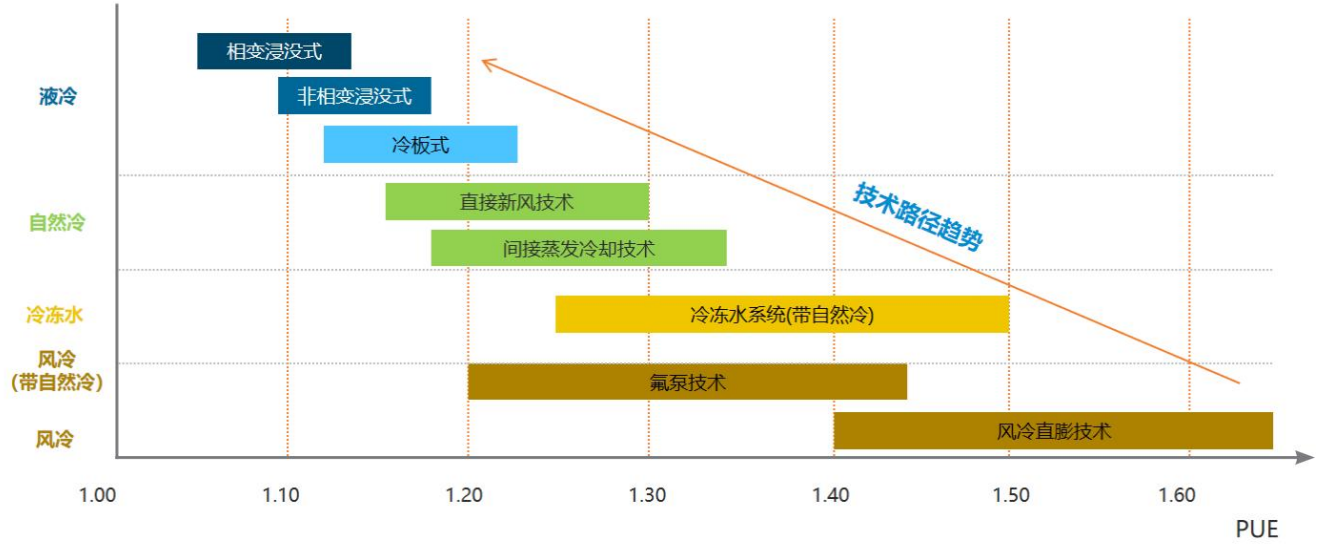
城市	年平均气温℃	数据中心 PUE 要求
北京	12.30	年能源消费量小于 1 万吨标准煤的项目 PUE 值不应高于 1.3；年能源消费量大于等于 1 万吨标准煤，且小于 2 万吨标准煤的项目，PUE 值不应高于 1.25；年能源消费量大于等于 2 万吨标准煤且小于 3 万吨标准煤的项目，PUE 值不应高于 1.2；年能源消费量大于等于 3 万吨标准煤的项目，PUE 值不应高于 1.15；14＜PUE≤18，每度电加价¥0.2；PUE＞18，每度电加价¥0.5
上海	16.60	到 2024 年，新建大型及以上数据中心 PUE 降低到 13 以下，起步区内降低到 1.25 以下。推动数据中心升级改造，改造后的 PUE 不超过 1.4。

广东	22.60	新增或扩建数据中心 PUE 不高于 13，优先支持 PUE 低于 1.25 的数据中心项目，起步区内 PUE 要求低于 1.25
浙江	16.50	到 2025 年，大型及以上数据中心电能利用效率不超过 13，集群内数据中心电能利用效车不得超过 1.25
江苏	15.50	到 2023 年底，全省数据中心机架规模年均增速保持在 20%左右，平均利用率提升到 65%，全省新型数据中心比例不低于 30%，高性能算力占比达 10%，新建大型及以上数据中心电能利用效率（PUE）降低到 1.3 以下，起步区内电能利用效率不得超过 1.25
山东	14.70	自 2020 年起，新建数据中心 PUE 值原则上不高于 1.3，到 2022 年年底，存量改造数据中心 PUE 值不高于 14。到 2025 年，实现大型数据中心运行电能利用效率降到 1.3 以下。优先支持 PUE 值低于 1.25，上架率高于 65%的数据中心新建、扩建项目
青岛	12.70	新建 13，至 2022 年存量改造 14
重庆	18.40	到 2025 年，电能利用效率（PUE）不高于 13。集群起步区内 PUE 不高于 1.25。
四川	15.30	到 2025 年，电能利用效率（PUE）不高于 13。集群起步区内 PUE 不高于 125。各市（州）要充分发挥已建在建数据中心作用，除天府数据中心集群外，区域内平均上架率未达到 60%、平均 PUE 值未达到 1.3 及以下的，原则上不得新建数据中心。
内蒙古	4.30	到 2025 年，全区大型数据中心平均 PUE 值降至 13 以下，寒冷及极寒地区力争降到 1.25 以下，起步区做到 12 以下
宁夏	9.50	到 2025 年，建成国家（中卫）数据中心集群，集群内数据中心的平均 PUE≤1.15，WUE≤0.8，分级分类升级改造国家（中卫）数据中心集群外的城市数据中心，通过改造或关停，到 2025 年，力争实现 PUE 降至 1.2 及以下。
贵州	15.50	引导大型和超大型数据中心设计 PUE 值不高于 1.3；改造既有大型、超大型数据中心，使其数据中心 PUE 值不高于 14。实施数据中心减量替代，根据 PUE 值严控数据中心的能源消费新增量，PUE 低于 1.3 的数据中心可享受新增能源消费量支持。

资料来源：中兴通讯《液冷技术白皮书》，中国银河证券研究院

液冷服务器是大势所趋，数据中心 PUE 可降至 1.25 以下。算力的持续增加，意味着硬件部分的能耗也在持续提升；在保证算力运转的前提下，只有通过降低数据中心辅助能源的消耗，才能达成节能目标下的 PUE 要求。

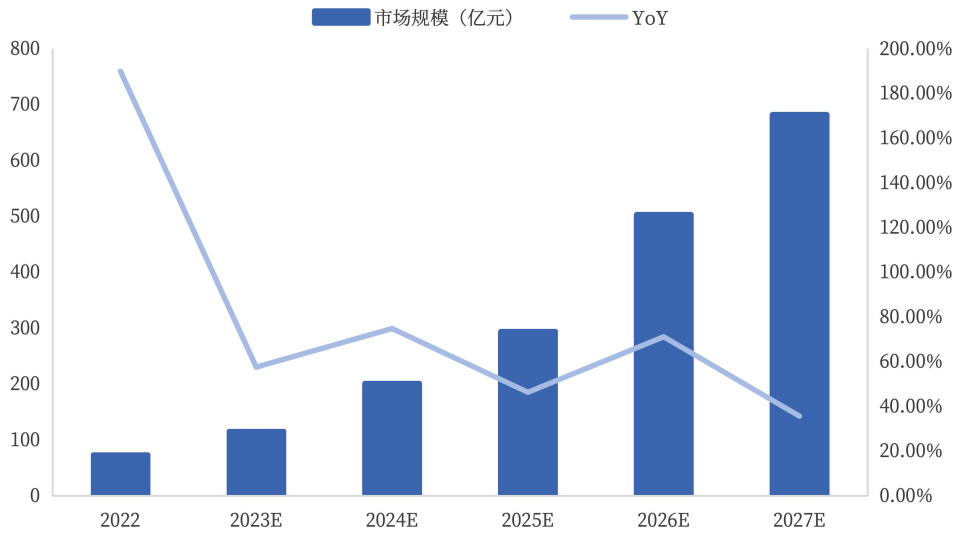
图 60：制冷技术 PUE 对比



资料来源：中兴通讯《液冷技术白皮书》，中国银河证券研究院

2023H1 中国液冷服务器市场同比增长近 3 倍。根据 IDC 发布的《中国半年度液冷服务器市场（2023 上半年）跟踪》报告数据显示，2023 上半年中国液冷服务器市场规模达到 6.6 亿美元，同比增长 283.3%，预计 2023 年全年将达到 15.1 亿美元。IDC 预计，2022-2027 年，中国液冷服务器市场年复合增长率将达到 54.7%，2027 年市场规模将达到 89 亿美元。

图 61：2022 年-2027 年中国液冷服务器市场规模预测

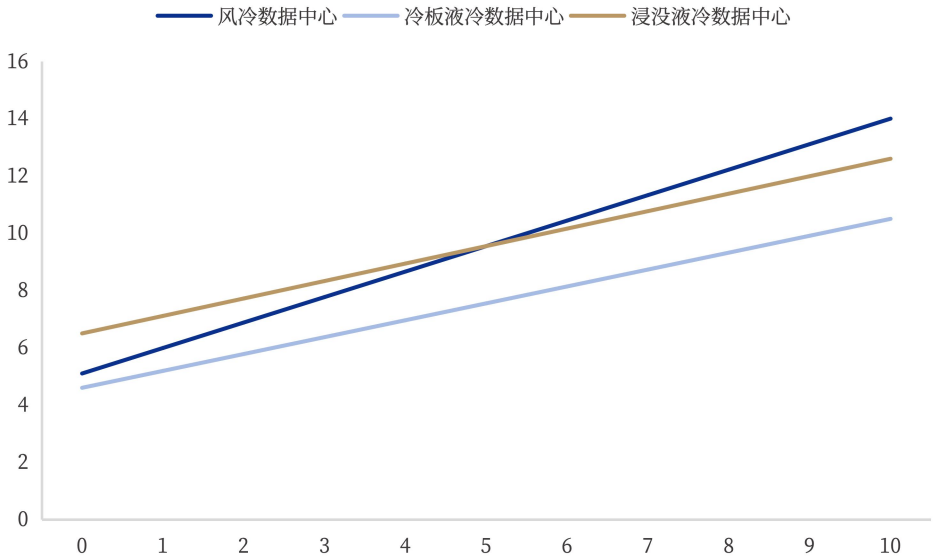


资料来源：IDC、中商产业研究院，中国银河证券研究院

冷板液冷前期投入成本已低于风冷，长期来看液冷（冷板式、浸没式）将有效降低数据中心 TCO。根据施耐德电气，数据中心总成本 TCO 大部分来源于电费等运营成本，大约 50%的数据中心运营支出（不包括 IT 设备）是电力成本。此前，市场普遍认为液冷数据中心基础设施前期投入成本较高，且后期不易维护，但伴随目前液冷技术发展以及对算力需求激增，一方面，冷板液冷数据中心初始建设成本已经低于风冷数据中心。另一方面，浸没式液冷虽前期投入较高，但每年可以节约大

量运营成本（电费）回收投资，数据表明，浸没液冷数据中心运行 4.5 年后 TCO 将出现拐点并且低于风冷数据中心。

图 62：2022 年-2027 年中国液冷服务器市场规模预测



资料来源：曙光数创，中国银河证券研究院

五、投资建议及盈利预测

下半年预期的重要催化剂 DeepSeek R2 版本或于下半年亮相，国内 AI 全产业链受益度有望提升。DeepSeek 在 5 月发布的版本 DeepSeek-R1-0528 思维深度与推理能力显著提升，市场预期 DeepSeek R2 版本或将在多模态融合、实时决策能力及垂直场景泛化性上实现跃升，对应投资机会聚焦四大方向：1）算力基础设施供应商，受益于模型训练和推理需求有望进一步激增；2）部署 DeepSeek API 的 SaaS 服务商；3）采用其技术重构业务流程的上市公司 4）溢出机会，能源-算力协同，能源是 AI 算力的底层约束，若技术突破将重塑产业价值链。

AI Agent 智能体经济已经全新开启，技术于产品迭代呈现不可逆趋势，对应投资机会包括：1、全球推理算力供需剪刀差不断扩大：根据我们测算，未来 3 年海外（不含中国）AI Agent 应用每日消耗的推理算力总量，2026-2028 年的增速分别达到 8 倍、3.5 倍、2.5 倍，若在 40%的算力利用率下，对应 2025 年 H200 的需求量可达 380.54 万，2026 年 B200 的需求量可达 1347.87 万，AI 芯片性能的进化无法赶超推理算力需求的急剧增长，全球推理算力供需剪刀差不断扩大；2、字节在 AI 应用生态领域已构建起相对优势，建议关注字节生态合作伙伴。截至 2025 年 5 月底，豆包大模型日均 tokens 使用量超过 16.4 万亿，较去年 5 月刚发布时增长 137 倍。同时，字节火山引擎占据了国内公有云上大模型调用量的 46.4%的市场份额，位列第一。3、建议关注在 AI Agent 方面布局领先的垂直领域卡位 SAAS 企业。从应用层面来说，伴随着 AI Agent 从“被动工具”迈向“自主决策体”，AI Agent 的商业模式也从“提供工具”向“交付价值”转变，能真正提升下游企业利润的 AI Agent 应用将会胜出，对应垂直行业 know how 型卡位公司的投资机会相对提升，能融入智能体能力的 SAAS 企业有望迎来价值重估的机会。

此外，信创产业进入深水区，国产生态繁荣将提升客户粘性进而提升盈利空间。今年是“十四五”收官之年，信创替代率将逐级落实，我们预计下半年党政、金融、能源、电新等重点行业信

创替代率进一步提升。稳定币长期发展趋势相对明确，短期事件催化下概念股行情预期依然具备反复机会。

投资建议：下半年建议重点关注以下细分赛道及个股，1、国产算力产业链：工业富联、中科曙光、曙光数创、海光信息、龙芯中科、地平线机器人-W 等；2、第三方 IDC 服务商：润泽科技、宝信软件等；3、信创厂商：中国软件、软通动力、达梦数据等；4、AI+应用：科大讯飞、金山办公、同花顺、嘉和美康、国能日新、彩讯股份、恒生电子、万兴科技等；5、云计算服务商：金蝶国际、金山云、优刻得、深信服；6、一体机及端侧 AI：神州数码、拓维信息、虹软科技、海康威视、中科创达、华勤技术、萤石网络等；7、数据要素产业链中供给、流通、应用公司：拓尔思、深桑达 A、上海钢联等；8、稳定币及 RWA：众安在线（6060.HK）、朗新集团。

表 10：重点推荐公司及盈利预测

证券代码	证券简称	股价	EPS				PE				投资建议
		6 月 21 日	2024A	2025E	2026E	2027E	2024A	2025E	2026E	2027E	
002230.SZ	科大讯飞	46.96	0.24	0.41	0.59	0.80	195.67	114.17	79.69	58.50	推荐
9660.HK	地平线机器人-W	6.46	-0.17	-0.16	-0.08	0.03	-38.00	-40.73	-76.81	188.89	-
0268.HK	金蝶国际	13.94	-0.04	0.04	0.09	0.16	-347.63	379.84	147.51	86.53	推荐
301162.SZ	国能日新	50.33	0.93	1.05	1.35	1.72	54.12	47.95	37.20	29.23	推荐
002415.SZ	海康威视	27.45	1.30	1.47	1.69	1.94	21.16	18.67	16.26	14.16	推荐
688208.SH	道通科技	29.89	1.42	1.18	1.48	1.77	21.05	25.41	20.24	16.87	推荐
688246.SH	嘉和美康	26.39	-1.86	0.29	0.65	1.24	-14.19	92.37	40.86	21.30	推荐
688318.SH	财富趋势	97.81	1.66	1.31	1.62	1.92	58.92	74.56	60.37	51.02	-
688568.SH	中科星图	34.54	0.65	0.63	0.87	1.13	53.14	55.22	39.59	30.52	推荐
688692.SH	达梦数据	219	5.44	4.18	5.21	6.39	40.26	52.38	42.03	34.26	-
002920.SZ	德赛西威	98.38	3.62	4.83	6.12	7.65	27.18	20.36	16.06	12.86	-
002970.SZ	锐明技术	46.63	1.66	2.29	3.06	4.04	28.09	20.36	15.26	11.55	-
001339.SZ	智微智能	45.66	0.50	1.03	1.46	1.67	91.32	44.38	31.30	27.30	-
002153.SZ	石基信息	8.33	-0.07	0.03	0.07	0.11	-119.00	243.57	119.00	76.92	-
300017.SZ	网宿科技	10	0.27	0.30	0.34	0.38	37.04	32.95	29.36	26.18	-
300033.SZ	同花顺	242.19	3.39	4.56	5.56	6.44	71.44	53.10	43.54	37.60	推荐
300170.SZ	汉得信息	16.11	0.19	0.25	0.31	0.37	84.79	64.41	52.31	43.38	-
300634.SZ	彩讯股份	25.26	0.51	0.65	0.82	1.00	49.53	38.77	30.86	25.36	推荐
301236.SZ	软通动力	52.6	0.19	0.41	0.57	0.77	276.84	127.08	91.67	68.29	-
301269.SZ	华大九天	117.43	0.20	0.40	0.57	0.76	587.15	296.77	206.56	153.87	-
002236.SZ	大华股份	15.27	0.90	1.02	1.19	1.35	16.97	14.99	12.84	11.32	-
002410.SZ	广联达	13.08	0.15	0.30	0.40	0.48	86.45	43.05	32.84	27.05	-
600536.SH	中国软件	45.65	-0.48	0.09	0.23	0.30	-95.10	502.75	201.10	150.81	-
600570.SH	恒生电子	28.8	0.55	0.65	0.75	0.88	52.36	44.18	38.44	32.74	推荐
603019.SH	中科曙光	67.55	1.31	1.68	2.05	2.49	51.56	40.17	32.96	27.12	推荐
603171.SH	税友股份	39.34	0.28	0.57	0.91	1.25	140.50	68.54	43.27	31.38	推荐
688041.SH	海光信息	134.15	0.83	1.35	1.93	2.61	161.63	99.46	69.45	51.35	推荐
688047.SH	龙芯中科	120.65	-1.56	-0.33	0.17	0.69	-77.34	-361.55	709.29	173.82	-
688095.SH	福昕软件	64.78	0.30	0.47	1.06	1.60	216.73	136.55	61.06	40.51	-

688111.SH	金山办公	270.3	3.56	4.15	5.05	6.13	75.93	65.06	53.57	44.08	推荐
300229.SZ	拓尔思	17.22	-0.12	0.08	0.12	0.17	-147.81	203.31	139.32	103.80	推荐
872808.BJ	曙光数创	58.04	0.31	0.51	0.72	0.97	187.23	114.77	80.72	60.06	推荐

资料来源：Wind，中国银河证券研究院

六、风险提示

（一）宏观经济不及预期风险：计算机行下游客户（政府、金融、企业）IT 预算受宏观经济影响较大，若宏观经济复苏不及预期，可能导致部分企业订单下降机回款周期拉长，进而影响公司现金流。

（二）政策推进不及预期风险：信创、政务信息化等行业受政策影响较大，若国产化替代政策如党政信创等推进不及预期，或财政资金拨付延迟，将直接影响重点行业国产化替代节奏，导致相关企业订单萎缩，营收不及预期。

（三）技术研发不及预期风险：大模型技术迭代放缓，国产 AI 芯片性能或产能不及预期，可能影响 AI 产业发展，导致商业化进程滞后。

（四）行业竞争加剧风险：部分领域竞争格局激烈，企业可能通过打“价格战”方式来换取一部分市场份额，导致行业整体利润承压。

（五）贸易摩擦风险：计算机行业部分企业依赖进口芯片等关键零部件，国际贸易摩擦可能导致供应链中断、成本上升等风险。

图表目录

分析师承诺及简介

本人承诺以勤勉的执业态度，独立、客观地出具本报告，本报告清晰准确地反映本人的研究观点。本人薪酬的任何部分过去不曾与、现在不与、未来也将不会与本报告的具体推荐或观点直接或间接相关。

吴砚靖 计算机/科创板研究负责人，北京大学软件项目管理硕士，10 年证券分析从业经验，历任中银国际证券首席分析师，国内大型知名 PE 机构研究部执行总经理。具备一二级市场经验，长期专注科技公司研究。

邹文倩 计算机/科创板团队分析师，复旦大学金融硕士，复旦大学理学学士；2016 年加入中国银河证券研究院；2016 年新财富入围团队成员。

免责声明

评级标准

评级标准		评级	说明
评级标准为报告发布日后的 6 到 12 个月行业指数（或公司股价）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，北交所市场以北证 50 指数为基准，香港市场以恒生指数为基准。	行业评级	推荐：	相对基准指数涨幅 10%以上
		中性：	相对基准指数涨幅在-5%～10%之间
		回避：	相对基准指数跌幅 5%以上
	公司评级	推荐：	相对基准指数涨幅 20%以上
		谨慎推荐：	相对基准指数涨幅在 5%～20%之间
		中性：	相对基准指数涨幅在-5%～5%之间
		回避：	相对基准指数跌幅 5%以上

联系

中国银河证券股份有限公司 研究院

深圳市福田区金田路 3088 号中洲大厦 20 层

上海浦东新区富城路 99 号震旦大厦 31 层

北京市丰台区西营街 8 号院 1 号楼青海金融大厦

公司网址：www.chinastock.com.cn

机构请致电：

深广地区：程 曦 0755-83471683 chengxi_yj@chinastock.com.cn

苏一耘 0755-83479312 suyiyun_yj@chinastock.com.cn

上海地区：陆韵如 021-60387901 luyunru_yj@chinastock.com.cn

李洋洋 021-20252671 liyangyang_yj@chinastock.com.cn

北京地区：田 薇 010-80927721 tianwei@chinastock.com.cn

褚 颖 010-80927755 chuying_yj@chinastock.com.cn