

2026 年 09 月 14 日



华鑫证券
CHINA FORTUNE SECURITIES

DeepSeek 发布 V4.1 Flash, GPT-Images-2.5 发布

—计算机行业周报

推荐(维持)

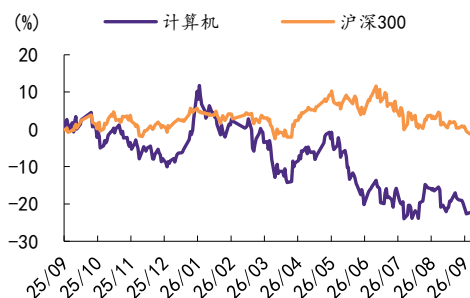
分析师: 任春阳 S1050521110006

rency@cfsc.com.cn

行业相对表现

表现	1M	3M	12M
计算机(申万)	-7.3	-3.8	-22.5
沪深 300	-4.0	-6.2	-0.9

市场表现



资料来源: Wind, 华鑫证券研究

相关研究

- 1、《计算机行业点评报告: 禾赛科技(HSAI): 激光雷达出货量延续高增, 机器人业务开启第二增长曲线》2026-09-07
- 2、《计算机行业点评报告: 文远知行(WRD): 海外业务驱动收入高增, 规模扩张下亏损持续收窄》2026-09-07
- 3、《计算机行业周报: GPT-6Astra 推出, ClaudeFable5.1 发布》2026-09-07

投资要点

■ **算力: 千问办公上线“多人工作台”, DeepSeek 发布 V4.1 Flash 推动显存瘦身与降价**

2026 年 9 月 10 日, DeepSeek 正式发布全新 MoE 架构模型 V4.1 Flash。该模型在 Agentic Benchmark 测试中超越 V4 Pro 与 GLM 5.3 等旗舰模型; 得益于 CSA2 机制、FP4 存储与 SWA 有界重放, HBM 与 SSD 需求分别降至上一代的 1/4 与 1/8, 较初代 KV Cache 累计缩小 437 倍; 官方全面下调 API 资费, 并同步推进 V4 Pro 向新模型的路由过渡与有序下线。

■ **AI 应用: ChatGPT 周访问量环比+4.38%, OpenAI 发布 GPT-Images-2.5**

2026 年 9 月 9 日, OpenAI 正式推出了其图像生成模型的最新版本——GPT-Images-2.5。该模型在生成延迟上最高可降低百分之五十, 同时大幅提升了图像编辑的精准度, 并引入了全新的 Sketch 手绘输入功能。在面向开发者的 API 端, OpenAI 首次将图像模型拆分为“快”与“慢”两个版本, 以满足不同场景下的差异化需求。目前, 该模型已面向 ChatGPT、ChatGPT Work 以及 Codex 的全线用户开放, 其中也包含免费版。

■ **AI 融资动向: Mistral 完成 30 亿欧元 D 轮融资, 三星领投创欧洲纪录**

法国 AI 独角兽 Mistral AI 宣布完成 30 亿欧元 (约合 234.39 亿元人民币) D 轮融资, 投后估值超 210 亿欧元, 由三星电子领投并签署战略合作协议, 创下欧洲科技企业史上最大单笔股权融资纪录。本轮融资进一步夯实其在开源全栈大模型与欧洲“主权 AI”领域的领跑地位。

投资建议

近日, 甲骨文发布其 2027 财年第一季度财报。当季总收入 193 亿美元, 同比增长 30%; 其中云收入 116 亿美元, 同比增长 62%, 云基础设施 (IaaS) 收入 74 亿美元、同比增长 121%, 云应用 (SaaS) 收入 42 亿美元、同比增长 10%; GAAP 营业收入 67 亿美元, 同比增长 57%, Non-GAAP EPS 1.92 美元, 同比增长 30%; 当季经营现金流 230 亿美元、同比增长 184%, 自由现金流 50 亿美元, 主因公司持续投入云基础设施建设。AI 订单与交付同步放量: 当季新签 AI 云合同超 300 亿美元,

剩余履约义务（RPO）同比增加 2,090 亿美元至 6,640 亿美元，公司表示 AI 云训练与推理需求持续快于供给；当季新增交付 850MW 数据中心容量，自上季末以来已向 AI 云客户交付超 30 万块 GPU，交付算力接近 4QFY26 的三倍。指引方面，2QFY27 总收入预期同比增长 30 - 34%、云收入预期同比增长 65 - 71%；FY2027 全年总收入预期至少 900 亿美元。

本次财报的核心信号在于 AI 云“订单—交付—收入”的传导链条全面提速：订单端，RPO 已达 6,640 亿美元且单季再增超 300 亿美元 AI 合同；交付端，单季 850MW 新增容量、GPU 交付量环比近三倍，IaaS 收入 121% 的增速说明产能爬坡正直接转化为收入，而非纸面订单；投入端，自由现金流阶段性转负、单季 200 亿美元股权增发落地，公司处于以资产负债表换取 AI 基建份额的投入高峰期。我们认为，甲骨文的财报与此前英伟达、Snowflake 的指引形成互证：英伟达在供给受限口径下仍给出环比约 12% 的收入指引，Snowflake 产品收入连续三个季度加速，甲骨文则完成了订单向收入的加速转化——AI 算力需求正从芯片、云平台到企业数据层全链条兑现；在订单已不再是瓶颈的背景下，头部云厂商的产能交付能力正成为 AI 云竞争格局的核心变量。

中长期，建议关注专注于功率半导体及模拟芯片测试系统研发的联动科技（301369.SZ）、专注于半导体等高端制造业的罗博特科（300757.SZ）、新能源业务高增并供货科尔摩根等全球电机巨头的唯科科技（301196.SZ）、AI 智能文字识别与商业大数据领域巨头的合合信息（688615.SH）、深耕工业 AI 与软件并长期服务高端装备等领域头部客户的能科科技（603859.SH）。

风险提示

- 1) AI 底层技术迭代速度不及预期。
- 2) 政策监管及版权风险。
- 3) AI 应用落地效果不及预期。
- 4) 推荐公司业绩不及预期风险。

重点关注公司及盈利预测

公司代码	名称	2026-09-14 股价	EPS			PE			投资评级
			2025	2026E	2027E	2025	2026E	2027E	
300757.SZ	罗博特科	612.30	-0.40	2.13	4.49	-1530.75	287.46	136.37	买入
301196.SZ	唯科科技	81.38	2.05	1.92	2.39	39.28	42.36	34.00	买入
301369.SZ	联动科技	124.72	0.48	1.08	2.62	259.83	115.48	47.60	买入
603859.SH	能科科技	38.10	0.92	1.05	1.32	41.41	36.29	28.86	买入
688615.SH	合合信息	88.19	3.24	4.22	5.25	27.22	20.90	16.80	买入

资料来源：Wind，华鑫证券研究

正文目录

1、 算力动态：千问办公上线“多人工作台”，DEEPSEEK 发布 V4.1 FLASH 推动显存瘦身与降价..... 4

 1.1、 Tokens 跟踪..... 4

 1.2、 云厂动态：千问办公上线行业首个“多人工作台”，探索 B 端无代码协作与 SaaS 生产新范式. 6

 1.3、 产业动态：DeepSeek 发布 V4.1 Flash 模型，架构创新推动 KV Cache 大幅瘦身且全面下调 API 资
 费 7

2、 AI 应用动态：CHATGPT 周访问量环比+4.38%，OPENAI 发布 GPT-IMAGES-2.5..... 9

 2.1、 周流量跟踪：ChatGPT 周访问量环比+4.38%..... 9

 2.2、 产业动态：OpenAI 发布 GPT-Images-2.5，生成提速一半且编辑精准度大幅提升..... 9

3、 AI 融资动向：MISTRAL 完成 30 亿欧元 D 轮融资，三星领投创欧洲科技股权融资纪录 12

4、 行情复盘 14

5、 投资建议 16

6、 风险提示 17

图表目录

图表 1：TOKENS 规模 LEADERBOARD（TOKEN 消耗量为统计口径）..... 6

图表 2：文本类模型市场份额占据示意图（REQUEST 调用量为统计口径）..... 6

图表 3：DEEPSEEK-V4.1-FLASH 与主流前沿模型在 AGENTIC BENCHMARK 上的性能对比..... 7

图表 4：DEEPSEEK V4.1 FLASH 定价表..... 8

图表 5：2026.9.5-2026.9.11 AI 相关网站流量..... 9

图表 6：GPT-IMAGES-2.5 发布..... 10

图表 7：GPT-IMAGES-2.5 生成的“山水宜昌”旅游指南 DEMO 10

图表 8：MANUS 团队对 GPT-2.5 的评测..... 11

图表 9：上周 AI 初创公司融资动态 12

图表 10：上周（2026.9.7-2026.9.11 日）指数日涨跌幅..... 14

图表 11：上周（2026.9.7-2026.9.11 日）AI 算力指数内部涨跌幅度排名 14

图表 12：上周（2026.9.7-2026.9.11 日）AI 应用指数内部涨跌幅度排名 15

图表 13：FICONTEC2025 年年中至今公告订单..... 16

图表 14：重点关注公司及盈利预测 17

1、算力动态：千问办公上线“多人工作台”，DeepSeek 发布 V4.1 Flash 推动显存瘦身与降价

1.1、Tokens 跟踪

根据 OpenRouter 公开数据，2026 年 9 月 7 日至 9 月 13 日，周度 Token 消耗量持续走高，消耗量达 127T，环比上周增加 10.43%。在 Tokens 规模 Leaderboard 前五名中，OpenAI 的 GPT-5.6 Luna 以 18.2T tokens 跃居榜首；Tencent 的 Hy4 preview 以 16.8T tokens 位列第二；Z-ai 的 GLM 5.3 Flash 以 11.9T tokens 位列第三；DeepSeek 的 DeepSeek V4 Flash 0731 以 11.6T tokens 位居第四；Xiaomi 的 MiMo-V2.5 以 7.77T tokens 位居第五。文本类模型的市场份额方面，openai 以 1.31B requests 占据 23.6% 的份额，升至第一；deepseek 以 1.25B requests 占据 22.6% 的份额，位居第二；google、tencent、z-ai 分别以 1.03B、400M、386M requests，对应占据 18.6%、7.2%、7.0% 的市场份额。

图像生成类模型调用量显著回升，调用量为 6.27M requests，环比上周增加约 9.04%。在 Image Leaderboard 中，OpenAI 的 GPT Image 2 以 1.32M requests 跃居第一；Nano Banana 2 (Gemini 3.1 Flash Image) 以 1.07M requests 位居第二；Nano Banana 2 Lite 以 832K requests 位居第三；Nano Banana 以 492K requests 位列第四；Nano Banana 2 (Gemini 3.1 Flash Image Preview) 以 428K requests 位居第五。从市场份额来看，google 以 3.4M requests 占据 54.2% 的份额，位居第一；openai 以 1.88M requests 占据 30.0%，位居第二；black-forest-labs、x-ai、bytedance-seed 则分别以 236K、203K、184K requests，对应占据 3.8%、3.2%、2.9% 的市场份额。

嵌入模型调用量微幅回落，调用量为 289M requests，环比上周下降 2.03%。在 Embedding Leaderboard 前五名中，OpenAI 的 Text Embedding 3 Small 以 119M requests 位居榜首；Qwen 的 Qwen3 Embedding 8B 以 55.5M requests 位列第二；BAAI 的 bge-m3 以 20.5M requests 位列第三；Google 的 Gemini Embedding 2 以 19.8M requests 位列第四；OpenAI 的 Text Embedding 3 Large 以 19.5M requests 位列第五。从市场份额来看，openai 以 142M requests 占据 49.0% 的份额，位居第一；qwen 以 61.3M requests 占据 21.2%，位列第二；google、baai、perplexity 则分别以 41.5M、21.1M、10.9M requests，对应占据 14.3%、7.3%、3.8% 的市场份额。

重排序模型调用量大幅上升，调用量为 11.8M requests，环比上周增加 42.17%。在 Rerank Leaderboard 中，rerank-2.5-lite 以 4.08M requests 跃居榜首；rerank-2.5 以 2.8M requests 位居第二；Rerank v3.5 以 2.33M requests 位列第三；Rerank 4 Fast 以 1.54M requests 位列第四；Qwen3 Reranker 8B 以 656K requests 位居第五。从市场份额来看，voyageai 以 6.88M requests 占据 58.2% 的份额，升至第一；cohere 以 4.2M requests 占据 35.5% 的份额，位列第二；qwen、nvidia 分别以 656K、94K requests，对应占据 5.5%、0.8% 的市场份额。

视频生成类模型调用量继续上升，调用量为 195K requests，环比上周增加 4.84%。在 Video Leaderboard 前五名中，Google 的 Veo 3.1 Lite 以 35K requests 升至第一；ByteDance 的 Seedance 2.0 Mini 以 23K requests 位列第二；ByteDance 的 Seedance 2.5 与 Seedance

2.0 Fast 均以 18K requests 并列第三；Google 的 Veo 3.1 Fast 以 11K requests 位列第五。从市场份额来看，bytedance 以 78K requests 占据 39.7% 的份额，位列第一；google 以 52K requests 占据 26.8% 的份额，位居第二；alibaba、x-ai、minimax 则分别以 16K、16K、14K requests，对应占据 8.3%、8.1%、6.9% 的市场份额。

语音生成类模型调用量小幅回调，调用量为 2.32M requests，环比上周下降 5.69%。在 Speech Leaderboard 前五名中，Google 的 Gemini 3.1 Flash TTS Preview 以 908K requests 位居第一；Hexgrad 的 Kokoro 82M 以 637K requests 位列第二；xAI 的 Grok Voice TTS 1.0 以 261K requests 位列第三；Fish-audio 的 S2.1 Pro 与 S2.1 Pro Free (free) 分别以 115K、104K requests 位列第四和第五。从市场份额来看，google 以 1.25M requests 占据 50.8% 的份额，位居第一；hexgrad 以 512K requests 占据 20.8%，位列第二；x-ai、fish-audio、deepgram 则分别占据 7.9%、7.1%、7.0% 的市场份额。

语音转写类模型调用量保持平稳，调用量为 16.6M requests，环比上周持平。在 Transcription Leaderboard 前五名中，OpenAI 的 Whisper Large V3、Whisper Large V3 Turbo 分别以 7.79M、3.8M requests 位居第一和第二；Microsoft 的 MAI-Transcribe 2 以 1.11M requests 升至第三；OpenAI 的 GPT-4o Mini Transcribe 以 899K requests 位列第四；NVIDIA 的 Parakeet TDT 0.6B v3 以 837K requests 位列第五。从市场份额来看，openai 以 13.1M requests 占据 79.2% 的份额，位居第一；microsoft 以 1.14M requests 占据 6.9% 的份额，位列第二；nvidia、qwen、mistralai 则分别以 894K、600K、415K requests，对应占据 5.4%、3.6%、2.5% 的市场份额。

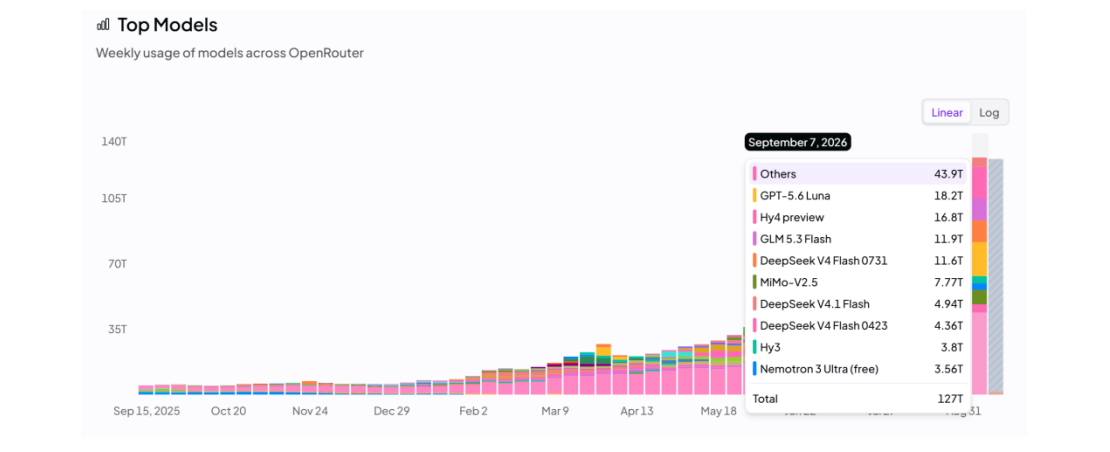
* 统计口径说明：以上数据均来源于 OpenRouter。基于官方统计口径调整，当前 Rankings 采用 Token 与 Request 两套统计维度：其中，仅 Tokens 规模 Leaderboard 仍以模型 Token 消耗量为统计口径，其余类目榜单及市场份额板块则统一采用 Request 调用量为统计口径，两者统计维度不同，不宜直接对应或进行横向比较。

9 月 7 日合肥中科类脑智能技术有限公司宣布完成数亿元 B+ 轮战略融资，由央企龙头旗下中车资本领投，资金将主要用于算电协同 Token 工厂异构算力纳管调度、博弈决策智能技术攻关及输出标准化 Token 工厂落地模式，以降低单位 Token 综合成本并加速规模化落地。

9 月 8 日北京首家 Token 工厂“北京壹号词元工厂”在亦庄正式落地并同步发布开源国内首个词元工厂性能基准。该项目由软通动力建设，对标电网稳定运行标准，一期每日 Token 产能达 1.4 万亿，远期规划日产 10 万亿，未来将联动张家口、乌兰察布等绿电基地构建京津冀一体化算力集群，以集约化与标准化模式推动高端算力普惠。

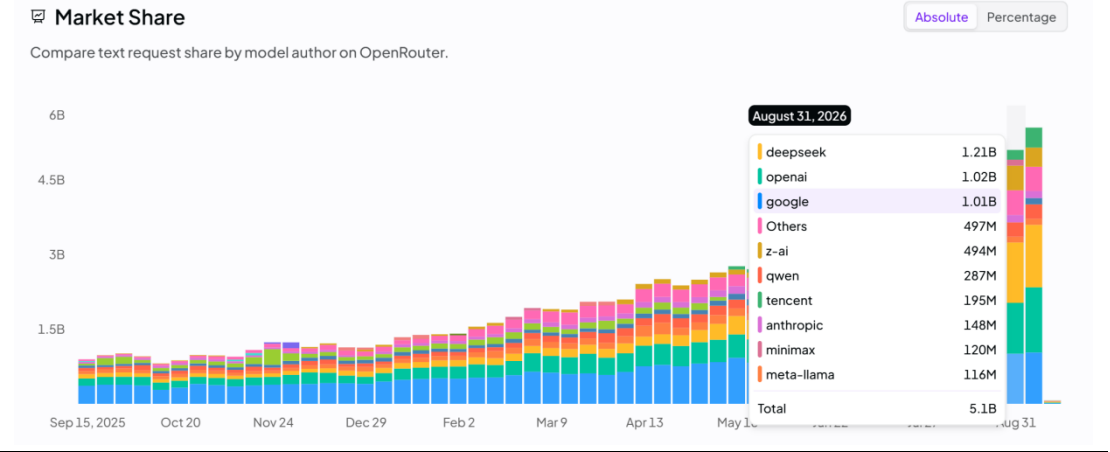
9 月 12 日 PEC 2026 AI 创新者大会暨第三届提示工程峰会在北京召开，大会以“赢得 Token，赢得世界”为主题，围绕前沿模型、Token 工厂与产业落地展开探讨。业内代表指出未来几年 Token 消耗的核心增量将来自生产力智能体，多智能体协同协作有望引爆企业级词元需求并驱动数据基础设施迭代。

图表 1：Tokens 规模 Leaderboard（Token 消耗量为统计口径）



资料来源：OpenRouter，华鑫证券研究

图表 2：文本类模型市场份额占据示意图（Request 调用量为统计口径）



资料来源：OpenRouter，华鑫证券研究

1.2、云厂动态：千问办公上线行业首个“多人工作台”，探索 B 端无代码协作与 SaaS 生产新范式

近日，千问办公正式上线行业首个“多人工作台”功能。该功能面向活动组织、家校协同与企业协作等多元业务场景，用户仅需在千问办公首页下方功能入口或技能中心搜索安装该技能，通过自然语言简单描述需求，千问办公即可自动拆解并执行任务，在十几分钟内生成并发布支持最多百人同时在线协作的交互式Web 网页。与市面上主要面向个人用户用于生成在线简历、作品集或单页打卡器等本地化、单向信息展示的常规 AI 工作台不同，“多人工作台”重点攻克企业办公与系统开发在“呆板冗长的在线协作文档”与“高成本、长周期定制软件开发”之间的中间地带难题，将个人级 Agent 工具升级为承载复杂业务流程的组织级协同系统。

核心能力与系统架构方面，千问办公“多人工作台”具备四项核心能力：一是角色与操作权限配置，系统支持为管理员与不同职级普通成员划分差异化数据权限与操作界面，例如在编辑部选题系统实测中支持自动区分主编、编辑与记者权限，在同城相亲角场景下

可限制普通用户隐私信息仅由管理员后台查看；二是云端数据库托管，生成的工作台不再局限于本地运行网页，所有业务数据统一存储于云端 Supabase (PostgreSQL) 数据库中，系 QW Pages 在发布工作台时开通的专属托管数据库，实现与其他用户数据完全物理隔离且永久在线；三是管理后台集中管控，统一留存成员提交的数据、任务状态及工作进展，便于统筹统计与复盘；四是一键在线发布与跨端协作，网页发布后生成独立访问链接，团队成员在移动端与 PC 端均可通过账号密码直接登录，最高支持百人实时在线协同。

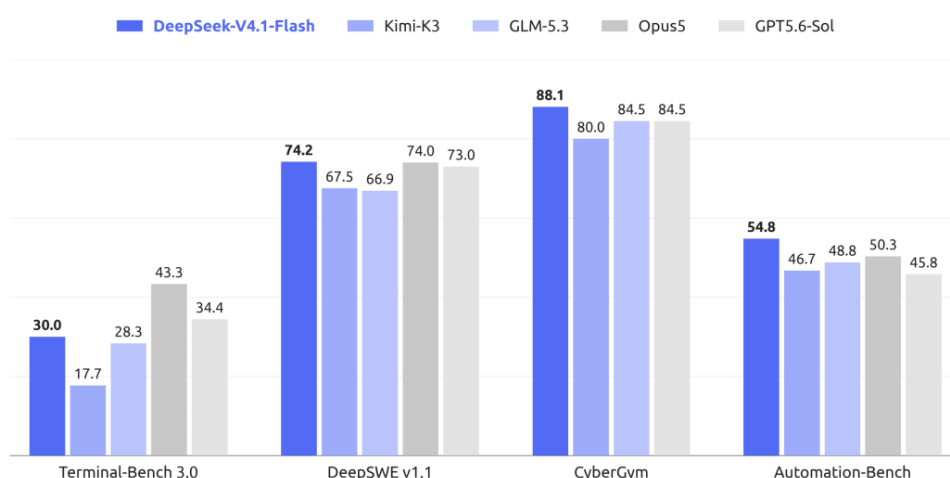
场景落地与生态协同方面，该功能适用于家校协作、达人投放、连锁店铺及百人市集等多元实际业务场景，且支持用户在业务流程变化后继续通过自然语言对网页与后台逻辑进行更新重构。此外，千问办公底层全面打通钉钉生态，支持群聊总结、文档表格创建与日程邮件收发等协同操作，并开源了上下文基础设施 Mycontext，支持企业将业务流程沉淀为组织级 Skill，持续加速 AI 生成软件在 B 端办公场景的真实生产力转化。

1.3、产业动态：DeepSeek 发布 V4.1 Flash 模型，架构创新推动 KV Cache 大幅瘦身且全面下调 API 资费

2026 年 9 月 10 日，DeepSeek 正式发布全新模型架 DeepSeek 发布 V4.1 Flash 模型，架构创新推动 KV Cache 降逾八成且全面下调 API 资费构系列中尺寸最小的成员——DeepSeek V4.1 Flash，该模型为 552B 参数的 MoE 架构，具备原生多模态视觉理解能力，已在 Hugging Face 同步开源权重与技术报告，网页端及整合了快速、专家和识图模式的 DeepSeek App 亦同步上线新模型。

架构设计方面，V4.1 Flash 采用了全新的 Causal-Encoder-Decoder 非对称结构，输入激活参数仅 8B，输出激活参数为 16B，官方给出的架构设计初衷为能力上限更高、推理速度更快、吞吐更大且可扩展到更大参数规模的模型。在新的预训练方式与更大规模强化学习后训练加持下，V4.1 Flash 在各项基准测试中超越了包括自研前代旗舰 DeepSeek V4 Pro 在内的多款前沿模型；在 Agentic Benchmark 等基准测试中，其性能表现全面超越 DeepSeek V4 Pro、GLM 5.3 以及 Kimi-K3。

图表 3：DeepSeek-V4.1-Flash 与主流前沿模型在 Agentic Benchmark 上的性能对比



资料来源：DeepSeek，华鑫证券研究

生态适配方面，DeepSeek 同步开源了轻量级 xPU 内核 JIT 编译库 DeepJIT，统一接口并支持 NVIDIA CUDA GPU 与华为昇腾 NPU 两大后端，提供跨节点共享缓存能力；同时 DeepSeek Harness v0.1.5 版本上线，新模型针对该框架进行了专项训练与优化，支持在保留已有 KV Cache 的情况下更新系统提示词；腾讯（WorkBuddy、CodeBuddy、腾讯云 TokenHub、ima、Marvis）以及 OpenCode 作为官方合作伙伴已全量接入该模型，其中 OpenCode Go 套餐提供 4 倍使用额度，WorkBuddy 提供限时两周独家优惠。

上下文存储与缓存优化方面，DeepSeek V4.1 Flash 大幅减少了 KV Cache 占用，与上一代模型 DeepSeek V4 Flash 相比，对 HBM 的需求减少至 1/4，对 SSD 的需求减少至 1/8，相较于初代模型 DeepSeek V1，其 KV Cache 累计缩小了 437 倍。该项压缩由架构、缓存精度和部署策略三方协同实现：在架构上，模型采用 CSA2（压缩稀疏注意力 2）机制，将全局 KV 缓存和 Top-K 索引跨层复用，每层仅保留局部注意力缓存与查询向量；在缓存精度上，模型从训练阶段即采用 FP4 格式存储全局 KV 缓存；在部署策略上，新增 SWA 有界重放机制，通过重放最近 n 个 token 近似重建滑动窗口注意力状态，避免 SWA 缓存写入 SSD。上述优化使模型将上下文从 4K 拉长至 1M 时，单 token 解码 FLOPs 仅增加约四分之一。

商业化定价与产品线调整方面，新价格已于 2026 年 9 月 10 日 12:00 正式生效，继续保留闲时为高峰时段半价的峰谷定价机制：闲时输入缓存命中为 0.02 元/百万 Tokens、未命中为 1.0 元/百万 Tokens、输出为 4.0 元/百万 Tokens；高峰时段输入缓存命中为 0.04 元/百万 Tokens、未命中为 2.0 元/百万 Tokens、输出为 8.0 元/百万 Tokens。API 服务方面，用户调用 deepseek-flash 即可接入新模型，旧版本 V4 Flash 与 V4 Flash Vision Exp 已下线并暂时路由至 V4.1 Flash；鉴于 V4.1 Flash 在性能、费用、速度及总用时等指标上全面超越 V4 Pro，官方计划有序下线 V4 Pro 模型，自北京时间 2026 年 9 月 14 日 12:00 起至 V4.1 Pro 上线前，所有 deepseek-v4-pro 请求将全部路由至 V4.1 Flash 并按其单价计费。

图表 4：DeepSeek V4.1 Flash 定价表

DeepSeek-V4.1-Flash API 价格表			
(单位：¥ / 百万Tokens)			
	输入（缓存命中）	输入（缓存未命中）	输出
空闲时段	0.02 元	1.0 元	4.0 元
高峰时段	0.04 元	2.0 元	8.0 元
高峰时段：工作日 9:00-12:00, 14:00-18:00（北京时间）其余均为空闲时段（含周末休息日） 北京时间 2026 年 9 月 10 日 12:00 开始生效			
deepseek			

资料来源：DeepSeek，华鑫证券研究

2、 AI 应用动态： ChatGPT 周访问量环比 +4.38%， OpenAI 发布 GPT-Images-2.5

2.1、 周流量跟踪： ChatGPT 周访问量环比+4.38%

本期（2026.9.5-2026.9.11） AI 相关网站流量数据：访问量前三位分别为 ChatGPT（1381.0M）、 Bing（838.0M）和 Gemini（615.2M），访问量环比增速第一为 ChatGPT（4.38%）；平均停留时长前三位分别为 Character.AI（00:15:41）、 Discord（00:10:55）和 NotionAI（00:08:48）；平均停留时长环比增速第一为 ChatGPT（1.00%）。

图表 5： 2026.9.5-2026.9.11 AI 相关网站流量

应用	应用类型	归属公司	周平均访问量 (M)	访问量环比	平均停留时长	时长环比
ChatGPT	聊天机器人	OpenAI	1381.0	4.38%	6:45	1.00%
Bing	搜索	微软	838.0	-5.49%	7:06	-1.39%
Gemini	聊天机器人	谷歌	615.2	2.01%	6:45	-0.98%
Canva	在线设计	Canva	215.7	3.26%	5:52	-0.56%
Github	代码托管	微软	148.9	0.74%	6:27	0.00%
Discord	游戏社区	微软	144	-1.37%	10:55	-0.46%
Character.AI	聊天机器人	Character.AI	35.42	1.17%	15:41	-1.57%
DeepL	翻译工具	DeepL	22.61	0.00%	2:27	0.68%
Perplexity	AI 搜索	Perplexity	21.56	1.13%	4:21	-1.51%
QuillBot	释义工具	QuillBot	8.041	0.61%	2:48	-1.18%
Kimi	聊天机器人	Moonshot AI	6.86	-11.14%	5:57	-10.30%
NotionAI	文本/笔记	Notion	3.07	2.99%	8:48	76%
文心一言	聊天机器人	百度	0.00	-51.66%	0:24	09%

资料来源： similarweb, 华鑫证券研究

2.2、 产业动态： OpenAI 发布 GPT-Images-2.5，生成提速一半且编辑精准度大幅提升

2026年9月9日， OpenAI 正式推出了其图像生成模型的最新版本——GPT-Images-2.5。该模型在生成延迟上最高可降低百分之五十，同时大幅提升了图像编辑的精准度，并引入了全新的 Sketch 手绘输入功能。在面向开发者的 API 端， OpenAI 首次将图像模型拆分为“快”与“慢”两个版本，以满足不同场景下的差异化需求。目前，该模型已面向 ChatGPT、 ChatGPT Work 以及 Codex 的全线用户开放，其中也包含免费版。

图表 6: GPT-Images-2.5 发布



资料来源：新智元，华鑫证券研究

从规模数据来看，OpenAI 公布了 ChatGPT Images 与 GPT-Image API 的合计生成量，已达到每周三十亿张图片。按此换算，日均约有四点三亿张图片经过这套系统处理。在演示中，GPT-Images-2.5 展现出了令人惊艳的能力，尤其是此前困扰 AI 生图的中文乱码问题已消失不见。通过生成的“山水宜昌”旅游指南 Demo 可以看出，模型能够准确渲染中文标题、景点介绍、美食推荐等复杂排版内容，文字清晰且逻辑连贯，场景一致性强。

图表 7: GPT-Images-2.5 生成的“山水宜昌”旅游指南 Demo



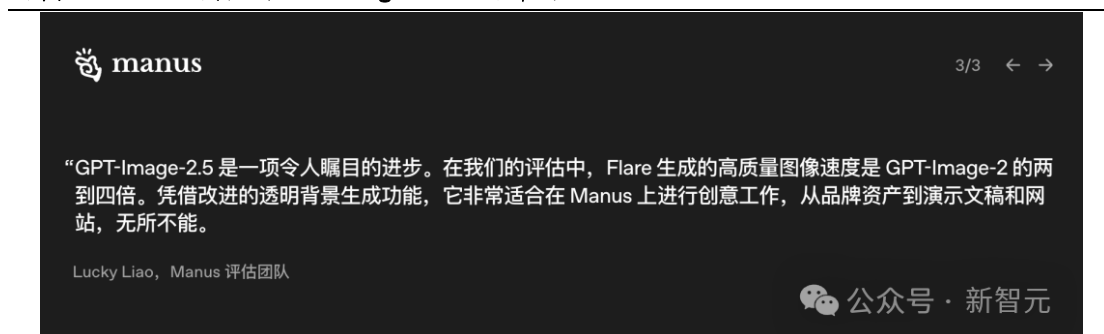
资料来源：新智元，华鑫证券研究

在核心能力提升方面，速度提升最为直观。该模型的生成延迟比 Images 2.0 最高降低了百分之五十，同时照片和纹理的渲染质量得到改善，参考照片中人物主体的还原度也更高。精准编辑则针对了生图工具长期存在的老问题：用户要求修改局部时，模型往往会连带改动未指定的部分。而 Images 2.5 能够仅修改指定区域，保留画面其余元素的构图、光照和主体特征，在复杂背景和多主体场景下的稳定性明显提高。用户甚至可以在图片上直接放置评论标注，圈定需要修改的位置，操作逻辑接近设计工具 Figma。此外，多轮编辑的一致性也是重要改进，在长对话中反复修改同一张图片时，前几轮的编辑结果能够保持住，画质不会随轮次下降。

产品功能层面，Images 2.5 带来了几项新特性。首先是 Sketch 功能，用户可在 ChatGPT 对话框输入 @Sketch 打开手绘面板，画一张草图并配上文字描述，ChatGPT 便能以草图为构图参考生成成品图，其门槛不在于画功，而在于画出空间关系和大致比例。其次是模板功能，覆盖海报、商品图、传单等高频格式，用户选择模板并填入信息即可按预设结构出图，省去了从零写 Prompt 的试错过程。第三是 Prompt 分享，用户可将生成好图的完整 Prompt 发出，他人带入自己的照片和细节，便能在同一框架下生成个人版本。例如“80 年代复古肖像” Prompt 已在社交平台流传，用户上传自拍即可生成同风格照片。

在 API 端，OpenAI 同步推出了 GPT-Image-2.5 Flare 和 GPT-Image-2.5 Sunburst 两款模型。Flare 主打速度和批量生成，被定位为大多数应用的默认选型，适用于社交内容、产品体验图、快速原型和高频出图。据 Manus 评测团队的数据，Flare 的出图速度达到 GPT-Image-2 的两到四倍，透明背景生成的改进对程序化生成品牌素材、演示文稿和网页配图直接有用。Sunburst 则定位高端创意工作流，用更长的生成时间换取更精细的编辑控制，典型场景包括品牌营销的成品级素材和精修产品图。

图表 8：Manus 团队对 GPT-Images-2.5 的评测



资料来源：新智元，华鑫证券研究

3、AI 融资动向：Mistral 完成 30 亿欧元 D 轮融资，三星领投创欧洲科技股权融资纪录

当地时间 2026 年 9 月 8 日，法国人工智能公司 Mistral AI 宣布完成总额 30 亿欧元（约合 234.39 亿元人民币）的 D 轮融资，投后估值超过 210 亿欧元（约合 1640.75 亿元人民币），创下欧洲科技企业迄今规模最大的单笔股权融资纪录。本轮融资由韩国三星电子领投，EQT 管理的 Scaleup Europe Fund（欧盟委员会承诺 10 亿欧元锚定投资的增长期基金）与现有投资方 PSG Equity 担任联合领投；新增投资方包括 Advent、贝莱德管理的基金及账户、卢森堡大公国；现有股东 a16z、ASML、法国巴黎银行 CIB、Bpifrance、DST Global、Eurazeo、General Catalyst、Index Ventures、Lightspeed、英伟达及 Salesforce Ventures 等机构继续跟投。从 C 轮由半导体光刻设备巨头 ASML 领投，到本轮由全球先进制造龙头三星电子领投，产业资本的连续注入反映出全球半导体与先进工业技术领域对 Mistral 全栈开放技术路径的认可。

产业协同与业务拓展方面，三星电子同日在巴黎举行的韩法国家峰会上与 Mistral AI 签署战略合作协议。根据合作框架，三星计划在半导体业务中全面部署 Mistral 的 AI 服务与解决方案，包括旗舰模型 Mistral Large，重点推进面向半导体芯片设计与制造环节的本地部署定制化 AI 模型，确保高度敏感的技术与运营数据完全留存于三星内部基础设施；针对日趋复杂的先进制造工艺，三星计划在缺陷检测与设备优化等关键环节落地专用 AI 模型，覆盖其先进存储与逻辑芯片产线，以加速研发周期、提升制造精度并稳定产线良率。资金用途与产品定位方面，Mistral 计划将本轮资金重点投向扩大前沿模型研究、提升训练算力、扩充底层基础设施并加速全球商业化布局。目前 Mistral 业务已覆盖 20 个国家，为空客、ASML、汇丰银行等 125 家以上全球大型企业提供关键任务 AI 转型服务；公司将自身定位为 AI 生成提供商，依托开放权重模型、算力基础设施与企业级 AI 产品的全栈布局，支持政企客户在数据治理权与底层控制权的前提下，实现关键生产业务的本地化与自主可控部署。

图表 9：上周 AI 初创公司融资动态

应用	应用类型	领投方	融资轮	融资额	目前累计	目前估值
					融资额	
Mistral	开源大模型	三星电子	D 轮	30 亿欧元	超 80 亿欧元	超 210 亿欧元

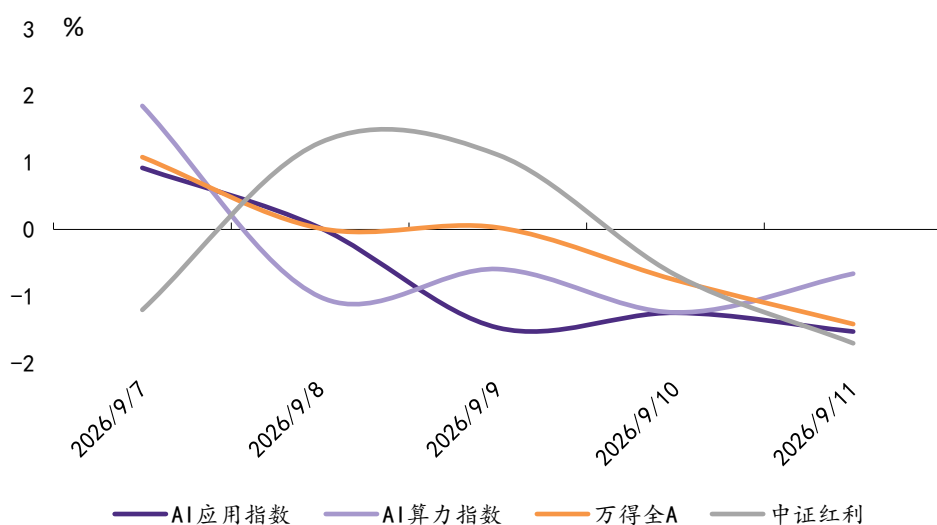
Cognition	AI 代码智能 体	Andreessen Horowitz (a16z)、Accel	E 轮	超 20 亿美 元	超 30 亿 美元	480 亿美 元
Positron	AI 推理芯 片	由 NEA 与 Jim Clark 等共同领投	C 轮	8.75 亿美 元	11.8 亿 美元	50 亿美元

资料来源：Saasverse，全球独角兽企业 500 强，东方财富网，PE 星球，华鑫证券研究

4、行情复盘

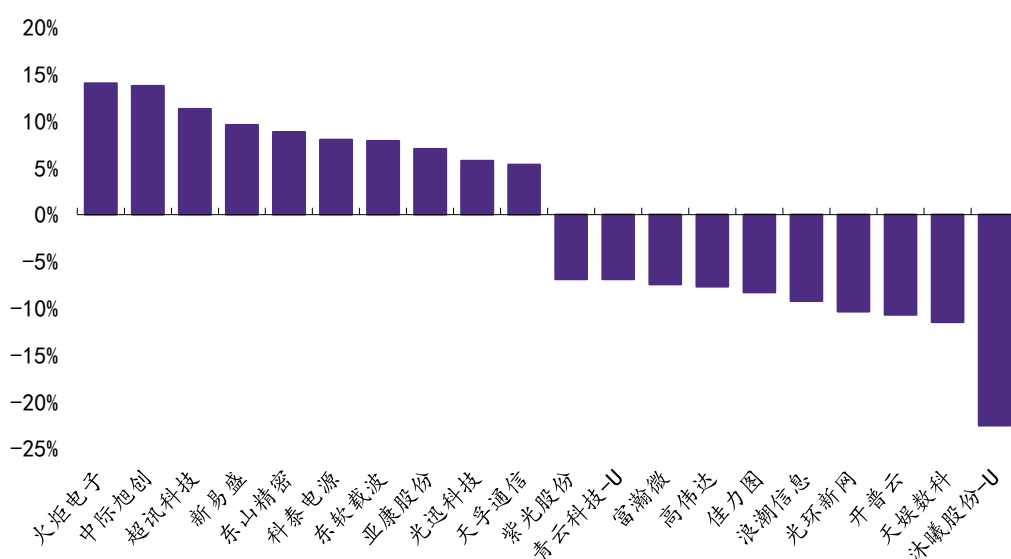
上周（2026.9.7-2026.9.11日），AI应用指数/AI算力指数/万得全A/中证红利日涨幅最大值分别为0.92%/1.85%/1.08%/1.3%，AI应用指数/AI算力指数/万得全A/中证红利日跌幅最大值分别为-1.53%/-1.24%/-1.42%/-1.71%。AI算力指数内部，火炬电子以14.02%录得上周最大涨幅，沐曦股份-U以-22.52%录得上周最大跌幅。AI应用指数内部，深南电路以19.55%录得上周最大涨幅，惠威科技以-13.36%录得上周最大跌幅。

图表 10：上周（2026.9.7-2026.9.11日）指数日涨跌幅



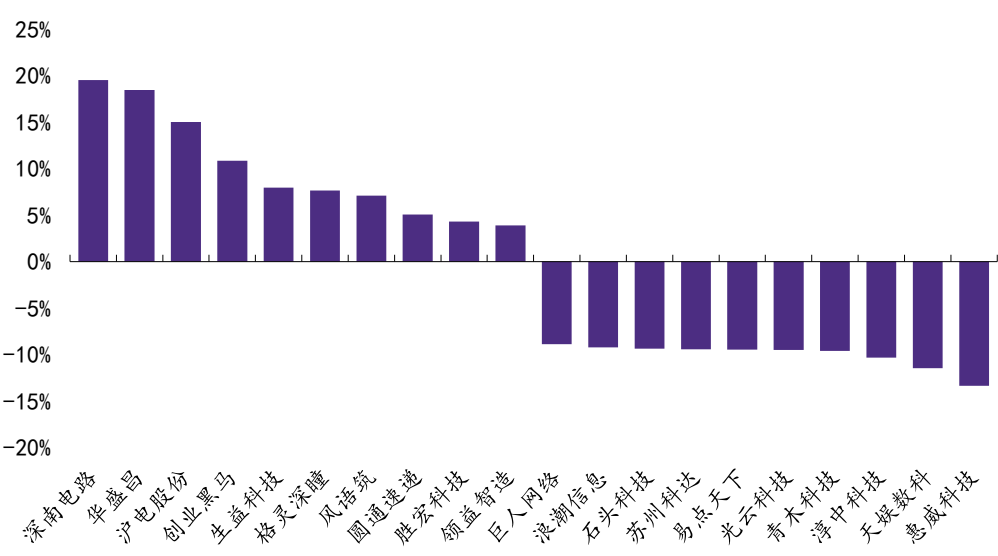
资料来源：wind, 华鑫证券研究

图表 11：上周（2026.9.7-2026.9.11日）AI算力指数内部涨跌幅度排名



资料来源：wind, 华鑫证券研究

图表 12：上周（2026.9.7-2026.9.11 日）AI 应用指数内部涨跌幅度排名



资料来源：wind, 华鑫证券研究

5、投资建议

近日，甲骨文发布其 2027 财年第一季度财报。当季总收入 193 亿美元，同比增长 30%；其中云收入 116 亿美元，同比增长 62%，云基础设施（IaaS）收入 74 亿美元、同比增长 121%，云应用（SaaS）收入 42 亿美元、同比增长 10%；GAAP 营业收入 67 亿美元，同比增长 57%，Non-GAAP EPS 1.92 美元，同比增长 30%；当季经营现金流 230 亿美元、同比增长 184%，自由现金流-50 亿美元，主因公司持续投入云基础设施建设。AI 订单与交付同步放量：当季新签 AI 云合同超 300 亿美元，剩余履约义务（RPO）同比增加 2,090 亿美元至 6,640 亿美元，公司表示 AI 云训练与推理需求持续快于供给；当季新增交付 850MW 数据中心容量，自上季末以来已向 AI 云客户交付超 30 万块 GPU，交付算力接近 4QFY26 的三倍。指引方面，2QFY27 总收入预期同比增长 30 - 34%、云收入预期同比增长 65 - 71%；FY2027 全年总收入预期至少 900 亿美元。

本次财报的核心信号在于 AI 云“订单—交付—收入”的传导链条全面提速：订单端，RPO 已达 6,640 亿美元且单季再增超 300 亿美元 AI 合同；交付端，单季 850MW 新增容量、GPU 交付量环比近三倍，IaaS 收入 121% 的增速说明产能爬坡正直接转化为收入，而非纸面订单；投入端，自由现金流阶段性转负、单季 200 亿美元股权增发落地，公司处于以资产负债表换取 AI 基建份额的投入高峰期。我们认为，甲骨文的财报与此前英伟达、Snowflake 的指引形成互证：英伟达在供给受限口径下仍给出环比约 12% 的收入指引，Snowflake 产品收入连续三个季度加速，甲骨文则完成了订单向收入的加速转化——AI 算力需求正从芯片、云平台到企业数据层全链条兑现；在订单已不再是瓶颈的背景下，头部云厂商的产能交付能力正成为 AI 云竞争格局的核心变量。

中长期，建议关注专注于功率半导体及模拟芯片测试系统研发的联动科技（301369.SZ）、专注于半导体等高端制造业的罗博特科（300757.SZ）、新能源业务高增并供货科尔摩根等全球电机巨头的唯科科技（301196.SZ）、AI 智能文字识别与商业大数据领域巨头的合合信息（688615.SH）、深耕工业 AI 与软件并长期服务高端装备等领域头部客户的能科科技（603859.SH）。

图表 13：ficonTEC2025 年年中至今公告订单

签约日期	客户/描述	业务类型	金额	折合人民币
2025/6/20	美国某头部公司 A 及其子公司	光电子封测设备	约 1,710 万欧元	约 1.36 亿元
2025/7/11	美国某头部公司 B 及其子公司	光电子封测设备	约 1,418 万美元	约 0.98 亿元
2025/9/3	瑞士某头部公司 C 的子公司	全自动硅光子封装整线设备或服务	约 946.50 万欧元	约 0.75 亿元
2025/10/21	武汉驿路通科技股份有限公司	光纤预制及组装线相关自动化设备	约 900 万美元	约 0.62 亿元
2026/1/6	瑞士某头部公司 C 的子公司	第二条全自动 OCS（光交换机）封装整线设备及服务	约 770.00 万欧元	约 0.61 亿元

2025/9/24-2026/1/26	以色列的纳斯达克上市 的头部公司 E	单面晶圆测试设备及服务	约 921.60 万 美元	约 0.64 亿元
2026/3/13	暂未披露	双面晶圆测试设备及服务	约 608.09 万 欧元	约 0.48 亿元
2026/3/19-2026/3/25	纳斯达克上市的公司 F 及其子公司	耦合设备及服务（可用于可 插拔硅光高速光模块封装制 程核心环节的量产）	约 6 亿元人民 币	约 6 亿元
2026/4/1	纳斯达克上市的公司 F	耦合设备及服务（可用于可 插拔硅光高速光模块封装制 程核心环节的量产）	约 3,570 万美 元	约 2.46 亿元
2026/4/8-2026/5/1	纽约证券交易所上市 的公司 B 的子公司	耦合设备及相关服务	约 2680 万美 元	约 1.83 亿元
2026/4/8-2026/5/1	纳斯达克上市的公司 F	视觉检测设备、高精度激光 bar 条封装设备及相关服务	约 3226 万美 元	约 2.20 亿
2026/7/23	公司 H 的子公司	光学元器件量产的自动化制 造设备及服务	约 1674 万欧 元	约 1.29 亿
总金额				约 19.22 亿元

资料来源：Wind，公司公告，华鑫证券研究

图表 14：重点关注公司及盈利预测

公司代码	名称	2026-09-14 股价	EPS			PE			投资评级
			2025	2026E	2027E	2025	2026E	2027E	
300757.SZ	罗博特科	612.30	-0.40	2.13	4.49	-1530.75	287.46	136.37	买入
301196.SZ	唯科科技	81.38	2.05	1.92	2.39	39.28	42.36	34.00	买入
301369.SZ	联动科技	124.72	0.48	1.08	2.62	259.83	115.48	47.60	买入
603859.SH	能科科技	38.10	0.92	1.05	1.32	41.41	36.29	28.86	买入
688615.SH	合合信息	88.19	3.24	4.22	5.25	27.22	20.90	16.80	买入

资料来源：Wind，华鑫证券研究

6、风险提示

1) AI 底层技术迭代速度不及预期。2) 政策监管及版权风险。3) AI 应用落地效果不及预期。4) 推荐公司业绩不及预期风险。

■ 中小盘&北交所组介绍

任春阳：华东师范大学经济学硕士，6 年证券行业经验，2021 年 11 月加盟华鑫证券研究所，从事计算机与中小盘行业上市公司研究

周文龙：澳大利亚莫纳什大学金融硕士

何春玉：金融学士、理学硕士，2023 年 8 月加盟华鑫证券研究所。

■ 证券分析师承诺

本报告署名分析师具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

■ 证券投资评级说明

股票投资评级说明：

	投资建议	预测个股相对同期证券市场代表性指数涨幅
1	买入	>20%
2	增持	10%—20%
3	中性	-10%—10%
4	卖出	<-10%

行业投资评级说明：

	投资建议	行业指数相对同期证券市场代表性指数涨幅
1	推荐	>10%
2	中性	-10%—10%
3	回避	<-10%

以报告日后的 12 个月内，预测个股或行业指数相对于相关证券市场主要指数的涨跌幅为标准。

相关证券市场代表性指数说明：A 股市场以沪深 300 指数为基准；新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以恒生指数为基准；美国市场以道琼斯指数为基准。

■ 免责条款

华鑫证券有限责任公司（以下简称“华鑫证券”）具有中国证监会核准的证券投资咨询业务资格。本报告由华鑫证券制作，仅供华鑫证券的客户使用。本公司不会因接收人收到本报告而视其为客户。

本报告中的信息均来源于公开资料，华鑫证券研究部门及相关研究人员力求准确可靠，但对这些信息的准确性及完整性不作任何保证。我们已力求报告内容客观、公正，但报告中的信息与所表达的观点不构成所述证券买卖的出价或询价的依据，该等信息、意见并未考虑到获取本报告人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。投资者应当对本报告中的信息和意见进行独立评估，并应同时结合各自的投资目的、财务状况和特定需求，必要时就财务、法律、商业、税收等方面咨询专业顾问的意见。对依据或者使用本报告所造成的一切后果，华鑫证券及/或其关联人员均不承担任何法律责任。本公司或关联机构可能会持有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提供投资银行、财务顾问或者金融产品等服务。本公司在知晓范围内依法合规地履行披露。

本报告中的资料、意见、预测均只反映报告初次发布时的判断，可能会随时调整。该等意见、评估及预测无需通知即可随时更改。在不同时期，华鑫证券可能会发出与本报告所载意见、评估及预测不一致的研究报告。华鑫证券没有将此意见及建议向报告所有接收者进行更新的义务。

本报告版权仅为华鑫证券所有，未经华鑫证券书面授权，任何机构和个人不得以任何形式刊载、翻版、复制、发布、转发或引用本报告的任何部分。若华鑫证券以外的机构向其客户发放本报告，则由该机构独自为此发送行为负责，华鑫证券对此等行为不承担任何责任。本报告同时不构成华鑫证券向发送本报告的机构之客户提供的投资建议。如未经华鑫证券授权，私自转载或者转发本报告，所引起的一切后果及法律责任由私自转载或转发者承担。华鑫证券将保留随时追究其法律责任的权利。请投资者慎重使用未经授权刊载或者转发的华鑫证券研究报告。