仅供内部参考，请勿外传！

证券研究报告

2026年9月7日

行业：计算机

增持（维持）

上海证券 SHANGHAI SECURITIES

算力筑基，应用向实

——计算机行业2026年秋季投资策略

分析师：杨佳伟　　SAC编号：S0870526050001

仅供内部参考，请勿外传！

主要观点

中美模型双极进化，中国从追赶到局部反超。中美大模型竞争已从参数竞赛转向架构效率、推理能力与生态渗透力的全方位比拼，2026年将是关键分水岭，模型能力的核心突破体现在Agentic AI和Physical AI的进步上。GPT-5.6、Gemini 3、Claude Fable 5等美国大模型以闭源为主，聚焦通用能力突破，在长上下文推理、实时处理和自主工具使用方面处于领先梯队。同时，开源模型的性能和生态影响力显著增强，DeepSeek、Qwen等国产大模型在多个评测基准上已能与海外顶尖模型媲美，在某些特定能力（如中文理解、性价比）上实现超越。算力成本的结构性下降与模型效率的显著提升，驱动全球AI产业从“技术军备竞赛”阶段正式迈入“商业化兑现”的关键转折点。

Token经济激发算力需求，国产算力链机遇空前。国家数据局统计显示，中国日均Token调用量已从2024年初的1000亿跃升至2026年3月的140万亿，两年增长超千倍。截至今年6月，豆包大模型日均Token调用量突破180万亿，QuestMobile数据显示豆包月活达3.82亿。智谱AI的MaaS平台API在一季度涨价83%后，调用量仍增长400%，呈现供不应求。国内外科技大厂持续加码，算力资本开支攀升。在美国芯片出口管制背景下，算力自主可控愈加凸显，华为昇腾384超节点、中科曙光scaleX640均已实现规模部署，标志着国产算力从“可用”走向“好用”。

AI应用落地提速，商业进程演进。AI Agent市场规模广阔，根据亿欧智库，2026年全球AI智能体市场规模预计达到107亿美元，其中中国16.4亿美元，占比15.3%。预计2030年市场总规模有望达471亿美元，中国份额稳步提升。在通用场景中，智能体在客户服务、市场营销、IT开发三大场景的应用比例均超过50%。我们认为，伴随大模型技术持续迭代演进，叠加开源生态进一步压低技术准入门槛与调用成本，AI 产业正经历核心范式切换，竞争主线由过去的“模型能力竞赛”转向“场景落地变现”。AI Agent 产品持续涌现，推动 AI 从工具辅助角色向实质生产力跃迁。

投资建议：建议关注（1）国产AI芯片：寒武纪、海光信息、芯原股份等；（2）超节点整机：中科曙光、浪潮信息、中兴通讯、紫光股份、华勤技术等；（3）AI 应用：金山办公、科大讯飞、福昕软件、万兴科技、鼎捷数智等。

风险提示：政策推进不及预期；技术创新与商业化不及预期；行业竞争加剧。

仅供内部参考，请勿外传！

目录

Content

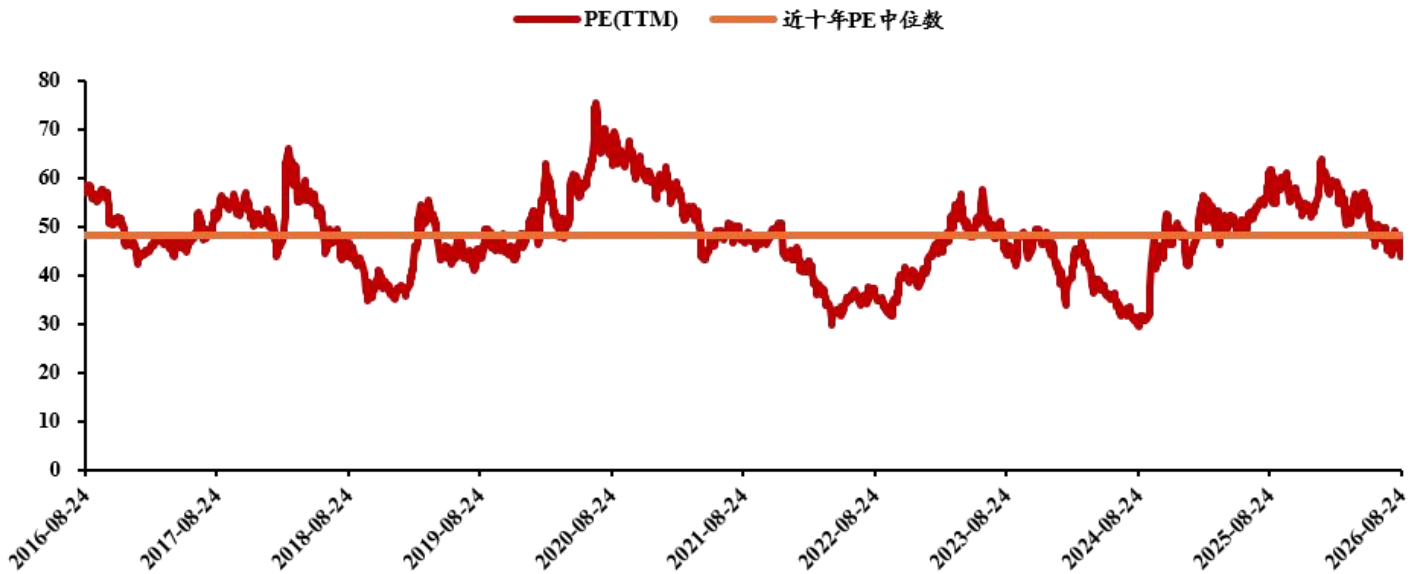
一、计算机板块行情回顾
二、大模型：技术持续演进，能力迭代升级
三、AI Agent：应用落地提速，商业进程演进
四、算力：国产算力崛起，超节点夯实智算根基
五、投资建议与核心标的
六、风险提示

仅供内部参考，请勿外传！

一、板块行情回顾

- 2026年截至8月24日，年初至今沪深300指数下跌1.44%，计算机（申万）指数下跌17.72%，跑输沪深300指数16.28个百分点。
- **当前计算机板块整体处于历史低估值水平。**截至2026年8月24日，计算机板块市盈率-TTM为43.93，近十年历史中位数为48.11，分位点处于近十年的24.42%。计算机指数（申万）过去10年市盈率-TTM低点为29.49，发生于2024年8月；历史高点为75.29，发生于2020年7月。

图 计算机指数（申万）近10年PE（TTM）



资料来源：iFinD，上海证券研究所

仅供内部参考，请勿外传！

目录

Content

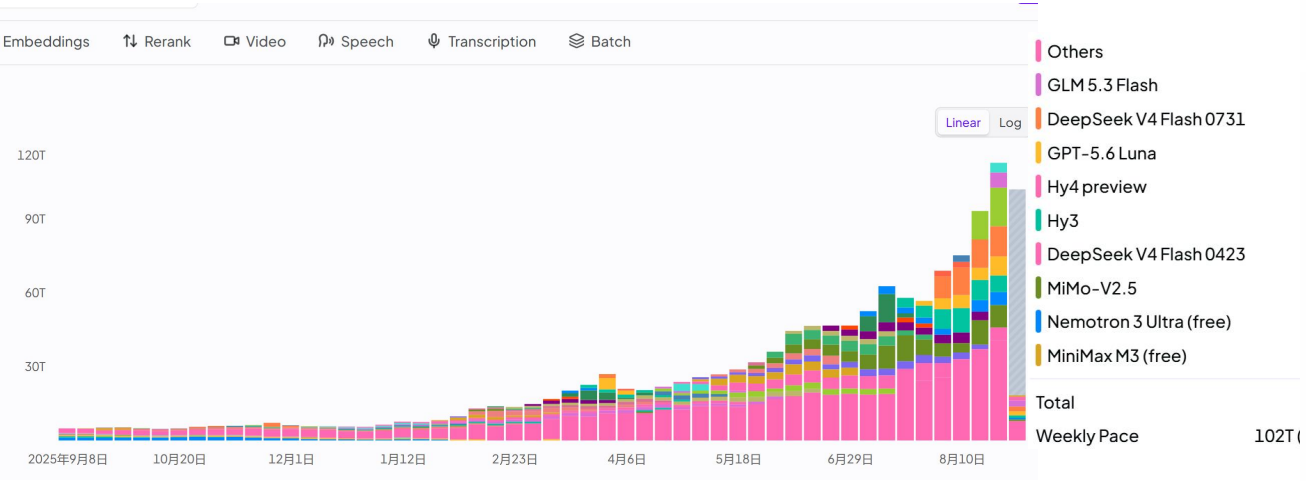
一、计算机板块行情回顾
二、大模型：技术持续演进，能力迭代升级
三、AI Agent：应用落地提速，商业进程演进
四、算力：国产算力崛起，超节点夯实智算根基
五、投资建议与核心标的
六、风险提示

仅供内部参考，请勿外传！

二、大模型格局重塑：中美模型双极进化，开源逼近闭源

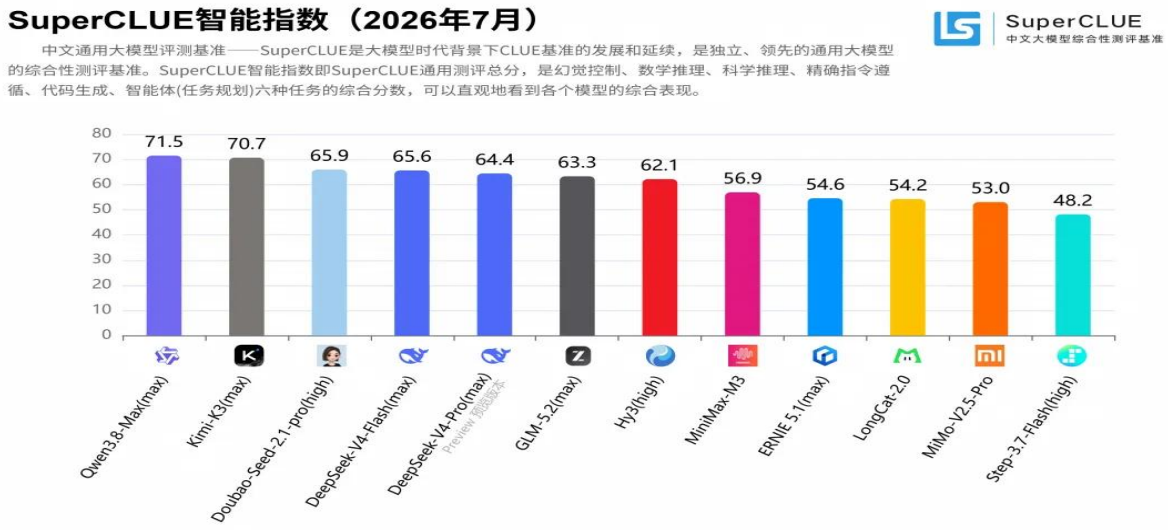
- 2026年，全球大模型竞争已从早期的参数规模竞赛，全面转向架构效率、推理能力与生态渗透力的全方位比拼。美国以OpenAI（GPT-5.6）、Google（Gemini 3）、Anthropic（Claude Fable 5）为代表的闭源模型，持续在长上下文理解、实时多模态处理、自主工具调用等通用能力上处于领先梯队。同时，中国开源模型的性能与生态影响力已实现显著跃升，以DeepSeek、Qwen 为代表的国产模型，通过创新的混合专家（MoE）架构和多头潜在注意力（MLA）等技术，在中文理解、代码生成、性价比等特定维度上与海外顶尖模型媲美，甚至实现局部超越。开源生态的繁荣正加速技术扩散与成本下降。
- 我们认为，2026年是关键分水岭，竞争核心将聚焦于Agentic AI 与Physical AI 的落地能力与商业生态构建。

图 大模型调用量持续增长



资料来源：openrouter官网（最后更新：2026年8月31日），上海证券研究所

图 SuperCLUE通用榜大模型榜单（2026年7月）



资料来源：SuperCLUE，上海证券研究所

仅供内部参考，请勿外传！

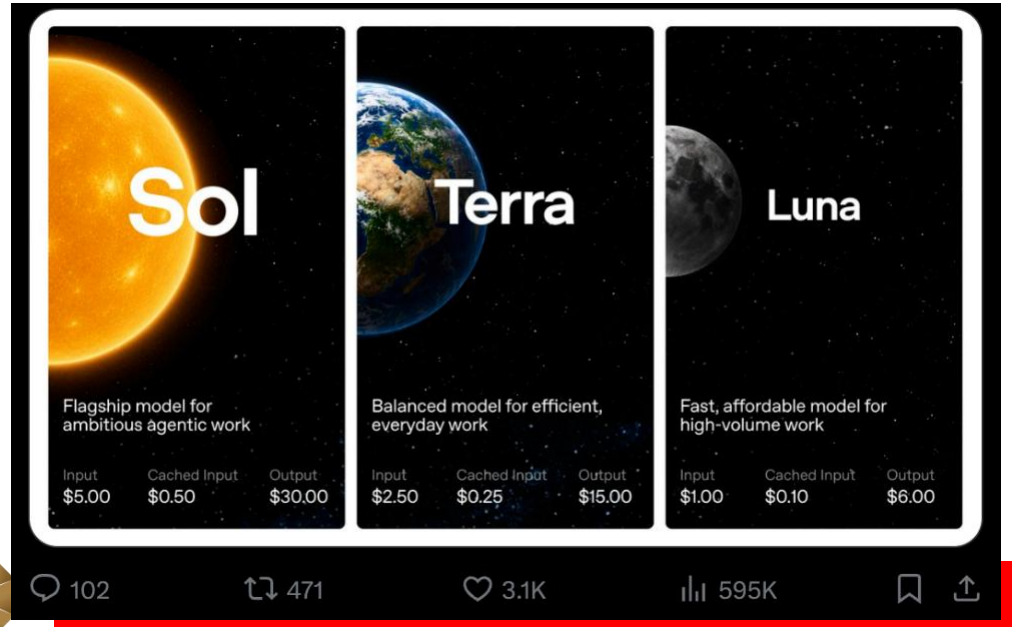
二、国内外大模型持续升级

- **OpenAI：**7月9日，OpenAI正式推出GPT-5.6系列模型，包括旗舰模型Sol、适配日常办公的均衡型模型Terra，以及速度最快、成本最低的模型Luna。

✓ **性能方面，**根据新华网微信公众号，模型能够更好地处理工作流中的复杂上下文，并生成可分享、可编辑的内容，在应用场景方面可用于软件开发、前端设计、文档和演示文稿生成、电子表格处理、科研辅助及网络安全防御等任务。其中，GPT-5.6 Sol 是 OpenAI 迄今最强的模型，模型在编程、生物学和网络安全等方向上具有更强的 Agent 能力。

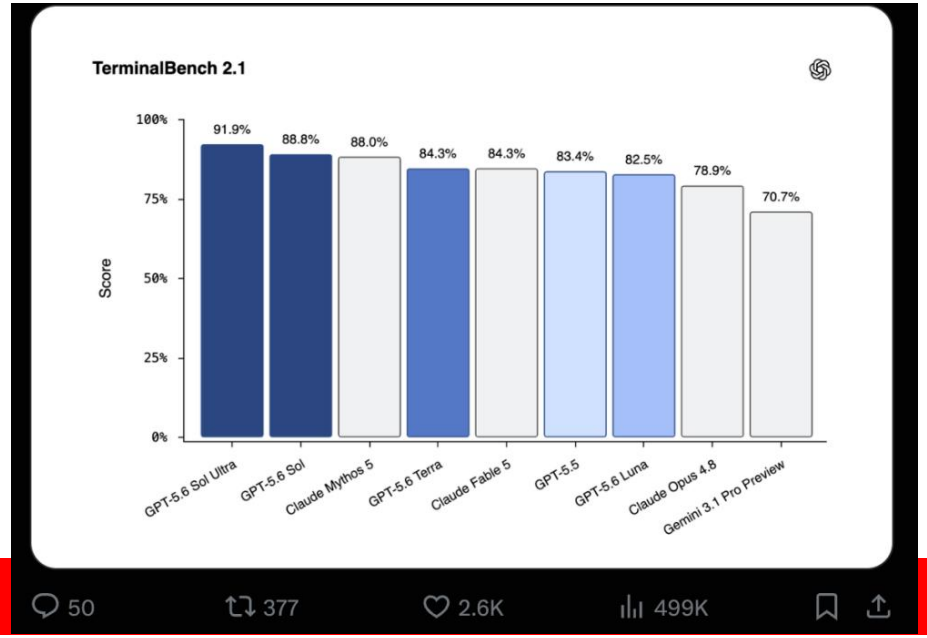
✓ **价格方面，**根据新华网微信公众号，每百万词元Sol输入为5美元、输出为30美元；Terra输入为2.5美元、输出为15美元；Luna输入为1美元、输出为6美元。

图 OpenAI推出GPT-5.6系列模型



资料来源：机器之心微信公众号，上海证券研究所

图 GPT-5.6 Sol表现



资料来源：机器之心微信公众号，上海证券研究所

仅供内部参考，请勿外传！

二、国内外大模型持续升级

- **Anthropic：**6月10日，Anthropic 发布Claude Fable 5 和 Claude Mythos 5，发布后短暂被美国商务部列为出口管制，7月1日正式解禁。7月25日，Anthropic上新大模型Opus 5。

✓ **性能方面，**根据CSDN微信公众号，**Claude Fable 5和Claude Mythos 5**底层架构一致，核心能力持平，Claude Fable 5 面向多数人开放，是首款正式向普通用户、开发者与企业提供的 Mythos 级模型。Claude Mythos 5 是“满血版”的旗舰模型，能力与 Fable 5 完全一致，但移除了所有安全分类限制。根据第一财经微信公众号，**Claude Opus 5**综合性能接近其旗舰顶配模型Fable 5，但调用成本仅为后者一半，适配日常办公、编程开发类通用任务。

✓ **收入亮眼，资本化加速。**根据第一财经微信公众号，5月29日，Anthropic宣布完成H轮融资，募资总额650亿美元，投后估值9650亿美元，超过竞对OpenAI上一轮8520亿美元的估值，成为全球身价最高的大模型初创公司。业绩表现方面，根据CNBC报道，Anthropic第二季度营收超过115亿美元，较2025年同期的7.87亿美元实现约14倍的强劲增长。

图 Claude Fable 5 和 Claude Mythos 5表现

	Claude Mythos 5 / Fable 5	Claude Mythos Preview	Claude Opus 4.8	GPT 5.5	Gemini 3.1 Pro
Agentic coding SWE-Bench Pro	80.3%	77.8%	69.2%	58.6%	54.2%
Agentic coding FrontierCode (Diamond)	29.3% xhigh	–	13.4% xhigh	5.7% xhigh	–
Knowledge work GDPval-AA	1932	–	1890	1769	1314
Knowledge work vision GDP.pdf	29.8% no tools	–	22.5% no tools	24.9% no tools	16.7% no tools
Spatial reasoning Blueprint-Bench 2	38.6%	–	14.5%	36.2%	26.5%
Tool use AutomationBench	17.4%	–	15.5%	12.9%	9.6%
Computer use OSWorld-Verified	85.0%	85.4%	83.4%	78.7%	76.2%
Legal Legal Agent Benchmark	13.3%	–	10.4%	2.1%	0.0%
Multidisciplinary reasoning Humanity's Last Exam	59.0%* no tools	56.8% no tools	49.8% no tools	41.4% no tools	44.4% no tools
	64.5%* with tools	64.7% with tools	57.9% with tools	52.2% with tools	51.4% with tools
Biology BioMysteryBench	46.1%* hard	29.6% hard	40.0% hard	–	–
	83.9%* human solvec	82.6% human solved	80.4% human solved	–	–
Agentic coding Terminal-Bench 2.1	88.0%*	–	82.7%	83.4% Codex CLI	70.7% Gemini CLI
Cybersecurity ExploitBench (Cap%)	78.0%*	69.0%	40.0%	34.0%	–
Health HealthBench Professional	66.0%*	64.7%	56.9%	51.8%	–

Methodology: Reported scores are within a 1-3 percentage point difference for Claude Mythos 5 and Claude Fable 5. This table shows the higher score of the two. Starred (*) benchmarks show a larger difference due to our blocking safeguards for cybersecurity and biology-related questions. For these benchmarks, Claude Fable 5 performs closer to Claude Opus 4.8 due to fallbacks. See the system card for details.

资料来源：CSDN微信公众号，上海证券研究所

仅供内部参考，请勿外传！

二、国内外大模型持续升级

- **DeepSeek：**8月12日，DeepSeek宣布上线DeepSeek V4 Pro正式版，并提及新版本增强了Agent能力，8月13日，发布DeepSeek Harness开发者预览版。此前7月底DeepSeek-V4-Flash 正式版 API 开启公测，4月上线DeepSeek V4 双版本。
  - ✓ **性能方面，**根据第一财经微信公众号，V4-Pro正式版上下文长度为1M，最大输出长度达到384K。根据官方发布的模型评测数据图显示，新模型的智能体性能接近海外前沿模型Claude的Fable 5，与Opus-4.8水平相当。根据CSDN微信公众号，在DeepSWE 基准测试中，模型得分从预览版的 12.8 直接提升至 62.7，反映出模型处理复杂软件工程任务、连续调用工具并自主完成多轮修改的能力提升。Harness则是面向开发者的工具，以MIT协议开放源代码，采用插件式开放架构来构建Agent Harness：模型、工具、技能、会话、存储、UI等所有Agent能力均由插件组合而成，可自由替换、灵活重组。
  - ✓ **价格方面，**采用峰谷定价，V4-Pro高峰时段百万tokens输出最高价格达27元。

图 DeepSeek-V4-Pro性能对比

Benchmarks	DeepSeek-V4-Pro-0813	DeepSeek-V4-Flash-0731	DeepSeek-V4-Pro-Preview	DeepSeek-V4-Flash-Preview	Opus-4.8	Fable 5 (w/ fallback)
HLE (wo/w tools)	42.7/60.0	37.8/51.5	37.7/48.2	34.8/45.1	49.8/57.9	53.3/63.0
Terminal Bench 2.1	87.9	82.7	72.1	61.8	85.0	88.0
NL2Repo	61.5	54.2	38.5	39.4	69.7	-
Cybergym	83.3	76.7	52.7	38.7	78.3	83.1
DeepSWE	62.7	54.4	12.8	7.3	58.0	70.0
Toolathlon-Verified	74.1	70.3	55.9	49.7	76.2	77.9
Agents' Last Exam	25.7	25.2	16.5	15.8	25.7	-
AutomationBench (Public)	31.8	25.1	12.8	10.8	27.2	29.1
DSBench-FullStack	71.1	68.7	41.8	37.0	71.6	77.2
DSBench-Hard	67.2	59.6	31.1	25.8	71.7	68.3

DeepSeek V4 Pro 正式版模型在 Agent 相关评测集上的表现以及与其他前沿模型的对比

资料来源：第一财经微信公众号，上海证券研究所

图 DeepSeek-V4-Pro定价情况

DeepSeek-V4 API 新定价

（单位：¥ / 百万Tokens）

模型		输入（缓存命中）	输入（缓存未命中）	输出
DeepSeek-V4-Flash	空闲时段	0.05 元	1.5 元	4.5 元
	高峰时段	0.10 元	3.0 元	9.0 元
DeepSeek-V4-Pro	空闲时段	0.15 元	4.5 元	13.5 元
	高峰时段	0.30 元	9.0 元	27.0 元

高峰时段：北京时间 9:00 - 12:00，14:00 - 18:00 （其余为空闲时段）

北京时间 2026 年 8 月 17 日 00:00 开始生效

资料来源：DeepSeek微信公众号，上海证券研究所

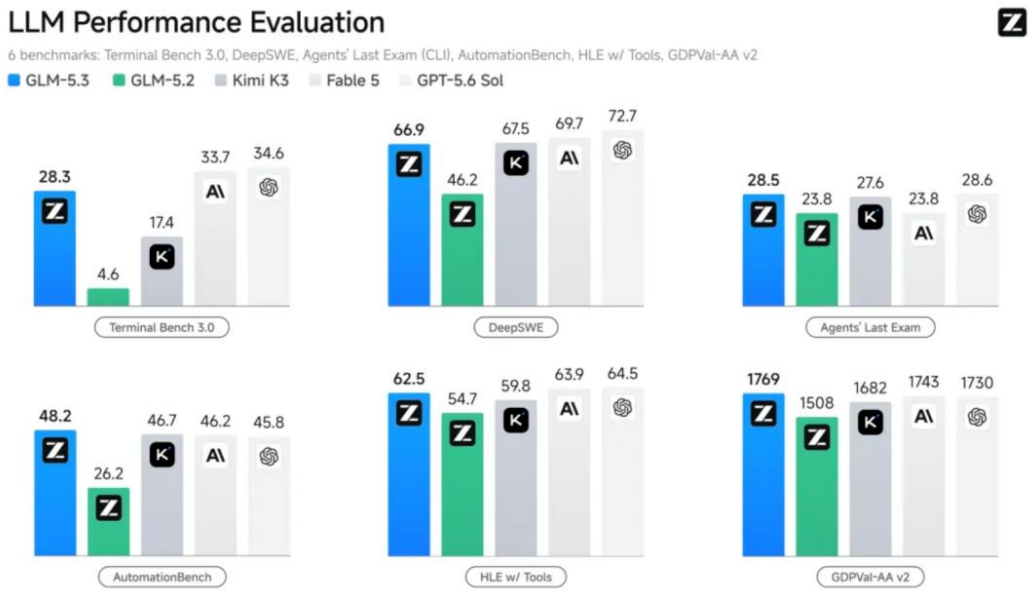
仅供内部参考，请勿外传！

二、国内外大模型持续升级

◆ **智谱：8月14日，智谱GLM-5.3发布，侧重编程能力与网络安全。**

✓ **性能方面，**根据第一财经微信公众号，GLM-5.3未更换基座，参数规模为7530亿，与上一代基本一致，能力跃升均来自后训练阶段：更长的训练时间、更丰富的任务环境、更大规模的强化学习。据智谱官方技术报告，在衡量真实终端环境任务完成的Terminal-Bench 3.0中，GLM-5.3得分从GLM-5.2的4.6跃升至28.3；在考察长程软件工程能力的DeepSWE v1.1中，从46.2提升至66.9。智谱表示，在上述两项基准中GLM-5.3均位列开源模型第一。在覆盖44种职业知识工作的GDPval-AA v2中，GLM-5.3得分1769，超过GLM-5.2的1508和Kimi K3的1682。

图 GLM5.3模型表现



资料来源：第一财经微信公众号，上海证券研究所

仅供内部参考，请勿外传！

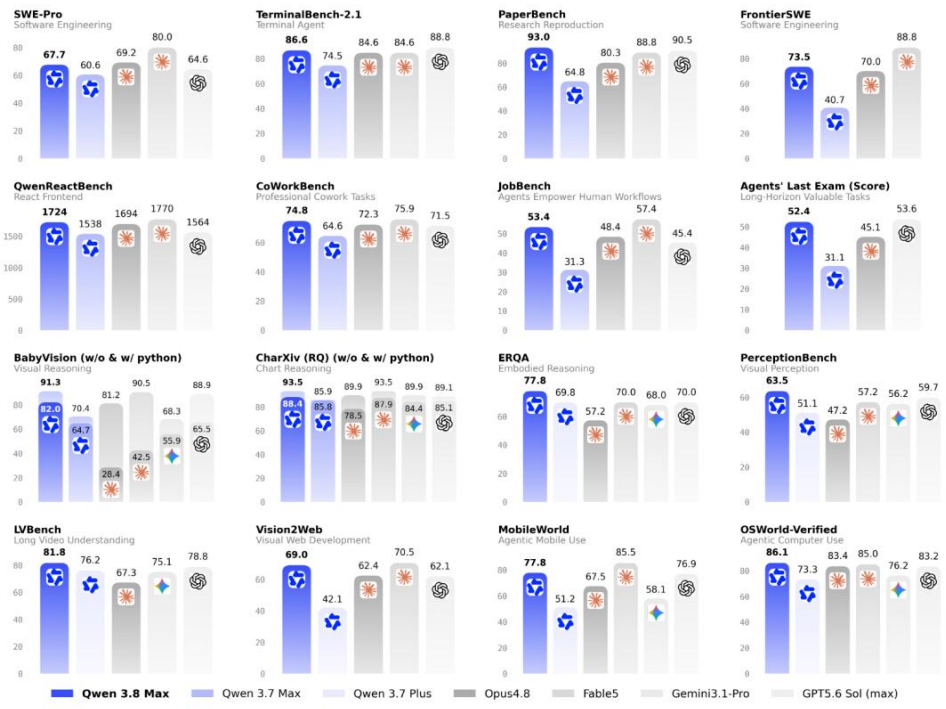
二、国内外大模型持续升级

◆ **阿里、Kimi：大参数规模模型加速迭代。**

✓ 阿里巴巴：根据第一财经微信公众号，8月3日，阿里巴巴发布新一代基座大模型Qwen3.8-Max。模型总参数量达2.4万亿，采用稀疏混合专家（MoE）架构，单次仅激活约950亿参数，兼顾旗舰级“脑容量”与极低推理延迟。模型支持最长100万Token上下文窗口，原生具备多模态视觉能力。

✓ 月之暗面：根据AIGC开放社区微信公众号，7月28日，Kimi K3 正式开源，是全球第一个开源的3万亿参数级别模型，编程和Agent能力逼近 Claude Fable 5 和 GPT-5.6 Sol，而推理成本仅为后者的几分之一。

图 Qwen3.8-Max基准测试结果



资料来源：阿里巴巴，智东西微信公众号，上海证券研究所

图 Kimi K3评测结果



资料来源：AIGC开放社区微信公众号，上海证券研究所

仅供内部参考，请勿外传！

目录

Content

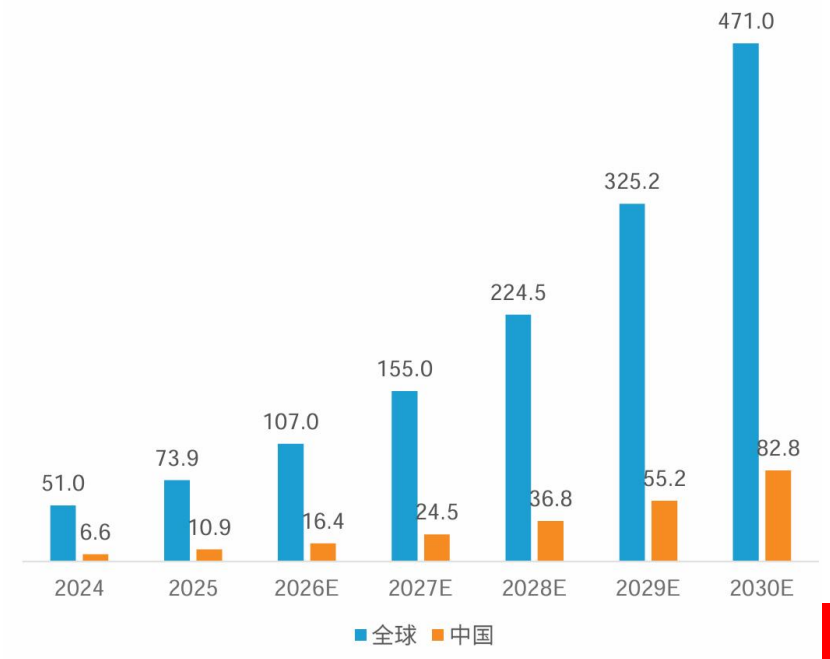
一、计算机板块行情回顾
二、大模型：技术持续演进，能力迭代升级
三、AI Agent：应用落地提速，商业进程演进
四、算力：国产算力崛起，超节点夯实智算根基
五、投资建议与核心标的
六、风险提示

仅供内部参考，请勿外传！

三、AI Agent商业化加速演进，市场空间广阔

- **从概念到生产力，商业化元年开启。**我们认为，伴随大模型技术持续迭代演进，叠加开源生态进一步压低技术准入门槛与调用成本，AI 产业正经历核心范式切换，竞争主线由过去的“模型能力竞赛”转向“场景落地变现”。AI Agent 产品持续涌现，推动 AI 从工具辅助角色向实质生产力跃迁。
- **AI Agent市场规模广阔。**根据亿欧智库发布的《2026全球AI商业落地价值洞察研究报告》，2026年全球AI智能体市场规模预计达到107亿美元，其中中国16.4亿美元，占比15.3%。预计2030年市场总规模有望达471亿美元，中国份额稳步提升。AI智能体从技术探索走向规模化落地，依赖于技术、数据和场景、商业模式、安全治理等多方面的协同演进。在通用场景中，客户服务、市场营销、IT开发三大场景的应用比例均超过50%，是智能体落地的重要环节。

图 2024-2030 全球与中国AI智能体市场规模（亿美元）



资料来源：Research and Market，OMDIA，亿欧数据，亿欧智库，上海证券研究所

图 AI智能体落地核心要素与主要环节



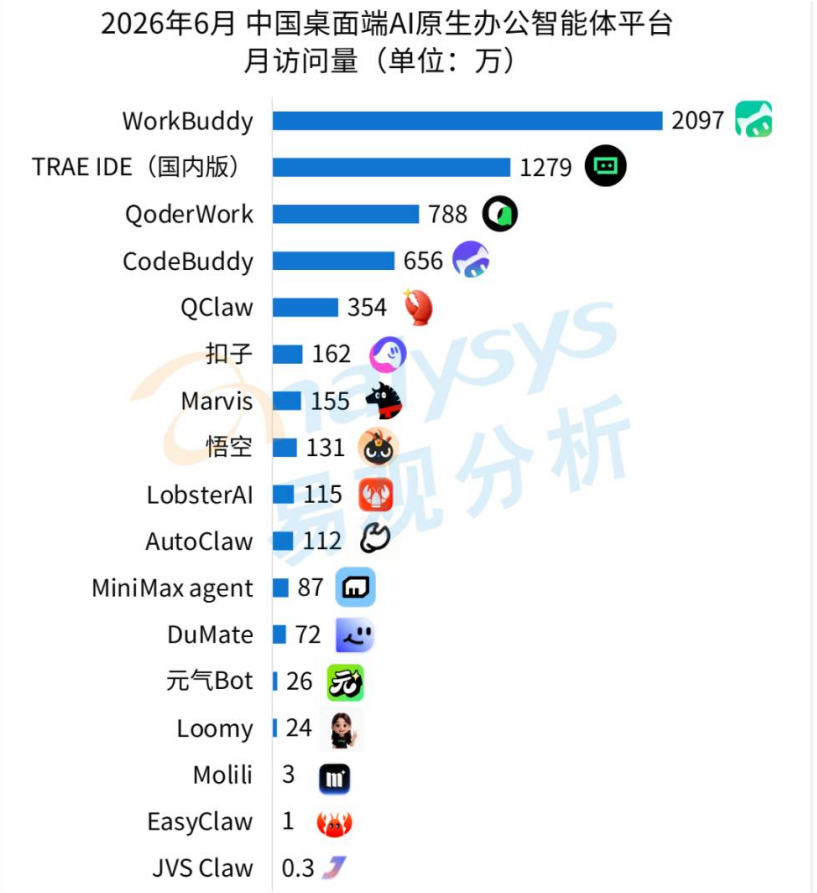
资料来源：Research and Market，OMDIA，亿欧数据，亿欧智库，上海证券研究所

仅供内部参考，请勿外传！

三、国内外AI Agent加速落地

- **B端爆发：**全球科技巨头正积极构建AI Agent生态，通过构建开放生态、完善工具链和推动场景落地，加速AI从“辅助工具”向“自主代理”的转变。谷歌推动开放标准（如A2A）与生态整合，上线AI Agent Space一站式商用市场；微软发布Copilot Studio平台，已有超10万家企业创建或编辑智能体。OpenClaw框架（开源AI Agent框架）引燃全球开发者社区，OpenRouter平台数据显示，到3月2日当周，平台周度Token调用量为14.8万亿。
- **C端涌现：**OpenAI发布Operator（可代理用户执行网页操作），智谱发布AutoGLM（手机智能体）和GLM-PC（电脑智能体），字节于2026年春节前推出“豆包手机”（AI智能体手机），Anthropic为Claude新增Computer Use功能，实现跨APP复杂任务自动化。Agent正从“聊天”进化为“执行”，成为真正的个人生产力助手。
- **腾讯、阿里、字节同步整合AI办公产品线。**年初大厂批量上线 AI 办公相关产品，7 月进入组织与产品收拢阶段，竞争焦点转向企业办公统一入口，角逐存量改造与新增市场机会。根据易观发布的《2026年Q2中国办公智能体平台市场洞察报告》，2026年6月，17款主流桌面端AI原生办公智能体合计访问量突破6000万次。腾讯WorkBuddy以2097万次月访问量位居第一，超过第二、第三名之和。

图 中国桌面端AI原生办公智能体平台访问量（2026年6月）



数据统计说明：①“访问量”指在统计周期内用户与PC端AI原生办公智能体平台的交互次数（已剔除异常访问数据及无实际交互行为发生的数据）。②核心数据来源包括运营商数据、部分平台披露及抽样监测数据、易观用户调研数据、网络公开数据等，通过易观自有模型测算得出。③仅统计通过电脑访问产品桌面端应用的用户行为数据，不包括通过Web页面/浏览器插件等浏览器端、手机/平板等移动端、办公软件平台内嵌智能体功能的数据。

资料来源：易观数据，上海证券研究所

仅供内部参考，请勿外传！

目录

Content

一、计算机板块行情回顾
二、大模型：技术持续演进，能力迭代升级
三、AI Agent：应用落地提速，商业进程演进
四、算力：国产算力崛起，超节点夯实智算根基
五、投资建议与核心标的
六、风险提示

仅供内部参考，请勿外传！

四、Token经济爆发，推动算力需求范式转移

- **Token调用量持续增高，推理需求井喷。**国家数据局数据显示，中国日均Token调用量已从2024年初的1000亿跃升至2026年3月的140万亿，两年增长超千倍，呈现指数级爆发。8月，国内首个词元经济专项金融产品“Token贷”在广州海珠区发布。截至今年6月，豆包大模型日均Token调用量突破180万亿，QuestMobile数据显示豆包月活达3.82亿，超过第二名通义千问与第三名DeepSeek的总和；智谱AI的MaaS平台API在2026年一季度涨价83%后，调用量仍逆势增长400%，呈现供不应求态势。我们认为，这标志着市场需求已从“尝鲜”转向“刚需”。
- **算力需求范式转变：**随着AI Agent在B端规模化部署和C端普及，算力需求正从“训练主导”转向“推理与智能体主导”。IDC相关预测显示，到2027年，推理将占智能算力需求的70%以上；Gartner指出，2027年中国云上AI推理负载占比将从当前的20%跃升至80%。

图 2024-2030年中国词元调用量增长趋势预测



资料来源：iiMedia Research（艾媒咨询），上海证券研究所

图 豆包大模型日均Token使用量突破180万亿



资料来源：新皮层NewNewThing，火山引擎FORCE2026原动力大会，上海证券研究所

仅供内部参考，请勿外传！

四、科技大厂持续加码，算力资本开支攀升

◆ **海外方面，北美云厂商披露财报，持续加码资本支出。**

✓ 微软：根据科创板日报微信公众号，微软第四财季资本支出达到410亿美元，同比增长超过70%，2027财年的资本支出预计为1750亿美元。

✓ 谷歌：根据第一财经微信公众号，谷歌Q2继续上调资本支出预期至1950亿至2050 亿美元之间。

✓ Meta：根据第一财经微信公众号，Meta Q2资本支出飙升至310.8亿美元，研发支出同比增长67%至217亿美元，全年资本支出指引升至1300亿-1450亿美元，主要用于全球数据中心和AI基础设施建设。

✓ 亚马逊：根据第一财经微信公众号，亚马逊Q2资本支出为531亿美元，主要与AWS和生成式人工智能相关。预计2026年的现金资本支出将达到约2200亿美元。

◆ **国内方面，国内互联网大厂及运营商同步加大资本投入。**

✓ 腾讯：根据21世纪经济报道微信公众号，Q2腾讯资本开支达到527.84亿元，同比增长176%、环比增长65%，上半年资本开支累计达到847.20亿元，超过2025年全年792亿元的规模。腾讯表示，在基础设施层，公司大幅增加了算力采购，将助力公司把应用及模型的使用转化为收入。

✓ 阿里：根据科创板日报微信公众号，阿里云计划今年将模块化数据中心的全球产能提升两倍以上。其已将大型AIDC数据中心的交付工期缩短至100天，同时阿里云数据中心整体建设成本降低10%以上。

✓ 字节：根据科创板日报微信公众号，字节计划今年资本支出将超过2000亿元人民币，较此前的初步计划增加了25%。

✓ 中国移动：根据21世纪经济报道微信公众号，中国移动上半年资本开支610亿元，同比增长4.5%，其中算力网投资增长71%，智能网投资增长85%。算间平均时延降至12毫秒以下，算网大脑已全面落地国家级、枢纽级、区域级算力平台。

仅供内部参考，请勿外传！

四、国产算力：超节点成舞台，从“可用”到“好用”

- **外部压力**：2024年底至2025年初，美国BIS多次加码对华半导体管制，将140余家中国实体列入清单，并限制台积电、三星等为大陆客户代工先进制程芯片，加速了自主可控的紧迫性。
- **内部突破**：国产AI芯片在大型训练场景中取得里程碑式进展。华为昇腾384超节点、中科曙光scaleX640等国产算力方案已实现大规模部署，成功支撑了多个万亿参数级别的大模型训练任务，我们认为这标志着国产算力在性能、规模和稳定性上已从“可用”走向“好用”，具备了与海外主流方案竞争的实力。
- **产业机遇**：随着强化学习成为后训练标配，推理计算占比持续提升，对能效比和定制化要求更高。根据TrendForce报告显示，2026年中国高端AI芯片市场中，国产解决方案的份额将接近90%。我们认为，这为ASIC及国产AI芯片（如海光DCU、寒武纪思元）提供了弯道超车的机会，有望逐步抢占英伟达GPU在推理市场的份额。

图 2025年中国市场AI加速卡出货量及市场份额

厂商	出货量（万块）	整体市场份额
英伟达	220.0	55.0%
华为	81.2	20.3%
平头哥	26.5	6.6%
AMD	16.0	4.0%
昆仑芯	11.6	2.9%
寒武纪	11.6	2.9%
海光信息	8.3	2.1%
沐曦	6.6	1.7%
天数智芯	4.9	1.2%

资料来源：IDC，云头条，上海证券研究所

图 华为昇腾384超节点



资料来源：华为计算公众号，上海证券研究所

仅供内部参考，请勿外传！

四、WAIC超节点集中亮相，万卡集群筑牢算力底座

◆ **2026世界人工智能大会（WAIC 2026）召开，多家厂商集中展出超节点算力方案。**H2 算力主题展区汇聚多款重磅产品，包括华为昇腾950超节点真机、中科曙光全国产十万卡AI超集群曙光8000、中兴通讯的OEX超节点、沐曦股份的曦景S600超节点等。我们认为，在先进制程工艺受限、推理算力需求旺盛的背景下，算力自主可控愈加凸显，超节点将打开算力产业成长新空间，带动智算产业从“单点突破”迈入“系统级竞争”，加速规模化布局。

表 WAIC超节点算力方案集中亮相

企业	超节点方案
华为	昇腾950超节点主要面向下一代万亿级大模型训练，以单机柜集成64张昇腾计算卡为基本单元，基于灵衢互联协议，实现了业界最大规模的1024卡集群，提供1 EFLOPS FP8、2 EFLOPS FP4算力与256TB全局统一内存编址。超节点技术研发采取全栈自研路线：从芯片（昇腾NPU）到互联协议（灵衢），再到系统软件与散热方案，均实现了自主研发。
中兴通讯	联合曦智科技、壁仞科技、沐曦股份、燧原科技、天数智芯等多家合作伙伴，基于OEX+dOCS架构打造了国产高性能Matrix超节点，该方案是国内少数同时满足128卡高密度、多芯片混插、dOCS算网一体化液冷的系统。
新华三	UniPoD S80000系列覆盖32卡到1024卡，最高可扩展至16384卡。
联想	问天超节点算力解决方案，单节点可搭载40张GPU，FP8算力超28 PFLOPS，HBM显存突破5.76 TB，可满足万亿参数大模型训练与推理极致需求。
海光、中科曙光	海光提供CPU+DCU芯片，中科曙光负责系统集成，形成“芯片+系统”的深度协同，发布ScaleX640+ScaleX40超节点，中国首个全国产十万卡AI超集群“曙光8000（登峰）”正式落成。
沐曦股份	首发曦景S600超节点，采用自研全精度GPGPU架构+MXMACA软件栈覆盖智算、通算全场景
摩尔线程	MTT C256超节点首创一层Scale-up网络，实现标准单宽机柜128卡全互联，并柜扩展可达256卡极速互联，攻克传统多层网络的带宽损耗与高延迟痛点
燧原科技	首发云燧ESL64-O超节点，实现“芯片—模组—系统—模型—应用”五层高效协同设计，性能与能效大幅提升。
东方算芯	超节点方案单柜配置64卡自研AI加速卡DF1000系列OAM模组，可提供33P@BF16算力，单卡支持896GB/s高速互联。
阿里云	阿里平头哥真武M890 AI芯片搭配阿里云磐久AL128超节点服务器，磐久AL128单柜集成128颗芯片，通过自研ALink互连架构实现Pb/s级带宽与百纳秒级延迟。真武M890采用自研并行计算架构，内置144GB高带宽显存，片间互联带宽达到800GB/s。
百度	旗下昆仑芯所支持的超节点产品已实现量产落地。

资料来源：C114通信网微信公众号，中国计算机报微信公众号，上海证券研究所

仅供内部参考，请勿外传！

目录

Content

一、计算机板块行情回顾
二、大模型：技术持续演进，能力迭代升级
三、AI Agent：应用落地提速，商业进程演进
四、算力：国产算力崛起，超节点夯实智算根基
五、投资建议与核心标的
六、风险提示

仅供内部参考，请勿外传！

五、投资建议与核心标的

- ◆ 我们认为，应紧抓“AI商业化”主线，聚焦算力、应用赛道：
- ◆ AI主线：把握从“技术军备”到“商业兑现”的转折
- ✓ 算力链（高确定性）：直接受益于Token需求爆发和国产替代，关注AI芯片、服务器、液冷、光模块等基础设施供应商。
- ✓ 应用层（高弹性）：关注在AI Agent、行业大模型、端侧应用有实际落地能力，并能带来业绩增量的软件企业。

表 核心标的一览表

领域	核心标的	证券代码	市值（2026年9月7日，单位：亿元）	2025EPS	2026EPS	2025PE	2026PE
算力芯片	寒武纪	688256	6869.48	4.88	8.94	277.59	122.28
	海光信息	688041	5641.40	1.09	1.93	204.96	125.86
超节点	中科曙光	603019	1219.95	1.49	2.05	57.58	40.72
	浪潮信息	000977	1097.54	1.64	3.30	40.53	22.67
AI应用	金山办公	688111	1111.45	3.96	6.04	77.45	39.71
	科大讯飞	002230	947.38	0.36	0.46	138.51	85.19

资料来源：iFinD，上海证券研究所；盈利预测为iFinD一致预期（截至2026.9.7）

仅供内部参考，请勿外传！

目录

Content

一、计算机板块行情回顾
二、大模型：技术持续演进，能力迭代升级
三、AI Agent：应用落地提速，商业进程演进
四、算力：国产算力崛起，超节点夯实智算根基
五、投资建议与核心标的
六、风险提示

仅供内部参考，请勿外传！

六、风险提示

- **政策推进不及预期：**经济环境或财政压力可能影响未来产业政策与补贴力度，延缓产业化进程。
- **技术创新与商业化不及预期：**技术研发难度大、周期长，产业化进程可能慢于市场预期；AI Agent的商业模式和盈利路径尚在探索，大规模商用存在不确定性。
- **行业竞争加剧：**中美科技竞争背景下，技术封锁与市场竞争可能加剧，影响国内企业的市场份额和盈利能力。

仅供内部参考，请勿外传！

行业评级与免责声明

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资格或相当的专业胜任能力，以勤勉尽责的职业态度，独立、客观地出具本报告，并保证报告采用的信息均来自合规渠道，力求清晰、准确地反映作者的研究观点，结论不受任何第三方的授意或影响。此外，作者薪酬的任何部分不与本报告中的具体推荐意见或观点直接或间接相关。

公司业务资格说明

本公司具备证券投资咨询业务资格。

投资评级体系与评级定义

股票投资评级：	分析师给出下列评级中的其中一项代表其根据公司基本面及（或）估值预期以报告日起6个月内公司股价相对于同期市场基准指数表现的看法。	
	买入	股价表现将强于基准指数20%以上
	增持	股价表现将强于基准指数5-20%
	中性	股价表现将介于基准指数±5%之间
	减持	股价表现将弱于基准指数5%以上
	无评级	由于我们无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，致使我们无法给出明确的投资评级
行业投资评级：	分析师给出下列评级中的其中一项代表其根据行业历史基本面及（或）估值对所研究行业以报告日起12个月内的基本面和行业指数相对于同期市场基准指数表现的看法。	
	增持	行业基本面看好，相对表现优于同期基准指数
	中性	行业基本面稳定，相对表现与同期基准指数持平
	减持	行业基本面看淡，相对表现弱于同期基准指数
相关证券市场基准指数说明：A股市场以沪深300指数为基准；港股市场以恒生指数为基准；美股市场以标普500或纳斯达克综合指数为基准。		

投资评级说明：

不同证券研究机构采用不同的评级术语及评级标准，投资者应区分不同机构在相同评级名称下的定义差异。本评级体系采用的是相对评级体系。投资者买卖证券的决定取决于个人的实际情况。投资者应阅读整篇报告，以获取比较完整的观点与信息，投资者不应以分析师的投资评级取代个人的分析与判断。

仅供内部参考，请勿外传！

行业评级与免责声明

免责声明

本报告仅供上海证券有限责任公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为客户。

本报告版权归本公司所有，本公司对本报告保留一切权利。未经书面授权，任何机构和个人均不得对本报告进行任何形式的发布、复制、引用或转载。如经过本公司同意引用、刊发的，须注明出处为上海证券有限责任公司研究所，且不得对本报告进行有悖原意的引用、删节和修改。

在法律许可的情况下，本公司或其关联机构可能会持有报告中涉及的公司所发行的证券或期权并进行交易，也可能为这些公司提供或争取提供多种金融服务。

本报告的信息来源于已公开的资料，本公司对该等信息的准确性、完整性或可靠性不作任何保证。本报告所载的资料、意见和推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值或投资收入可升可跌。过往表现不应作为日后的表现依据。在不同时期，本公司可发出与本报告所载资料、意见或推测不一致的报告。本公司不保证本报告所含信息保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本报告中的内容和意见仅供参考，并不构成客户私人咨询建议。在任何情况下，本公司、本公司员工或关联机构不承诺投资者一定获利，不与投资者分享投资收益，也不对任何人因使用本报告中的任何内容所引致的任何损失负责，投资者据此做出的任何投资决策与本公司、本公司员工或关联机构无关。

市场有风险，投资需谨慎。投资者不应将本报告作为投资决策的唯一参考因素，也不应当认为本报告可以取代自己的判断。