



计算机行业研究

买入（维持评级）

行业周报
证券研究报告

计算机组

分析师：刘高畅（执业 S1130525120005）

liugaochang@gjzq.com.cn

AI 产业三线共振：模型迭代、Agent 落地与国模出海

前沿模型密集迭代，专业任务能力持续提升。OpenAI 同时推出 GPT-6 Sol 和 Luna：Sol 面向复杂编程和 Agent 工作流，这类任务需要规划、调用工具并验证结果；Luna 则更偏向边界清楚、调用频繁的轻量任务。Anthropic 发布 Claude Opus 5.5，进一步强化长程编程和知识工作。三款模型均提供百万级上下文与 12.8 万 Token 最大输出，并支持网页搜索、文件搜索、代码执行和电脑操作等工具。异步工具调用和执行中追加指令，使模型能够在工具运行期间处理独立工作，并根据用户反馈及时调整要求。Opus 5.5 在编程和电脑操作评测中较前代提升，代码迁移与季度业绩报告测试也体现了其处理长工程任务和资料密集型工作的潜力。Sol、Luna 的标准 Token 价格低于前代，Opus 5.5 典型工作负载成本也有所下降，能力与成本的共同改善有望提升 Agent 工作流的经济性。

个人 Agent 走向终端，C 端应用进展突破。Meta Muse 运行于配有浏览器的独立安全虚拟电脑，用户可通过 Muse 应用或 WhatsApp 以自然语言委托任务。在返校准备场景中，Muse 能够查看学校邮件和学区网站，将日期加入家庭日历，把用品放入购物车，并完成餐厅查找及报名提醒。它还可根据食谱整理采购清单，结合饮食限制提出建议，并按需生成支出追踪工具、学习指南和睡眠看板。Muse 支持在用户关闭应用后继续工作，按照计划或相关事件推进任务，在出现重要变化或需要输入时通知用户。读取与写入权限的区分、Sentinel 检查及敏感操作前的用户确认，为 Agent 进入邮件、日历和交易等场景提供了可控基础。Meta 将 Muse 引入 AI 眼镜，国内豆包手机助手消费者版也随努比亚机型发布，个人 Agent 正在获得更贴近日常习惯的终端入口。若任务完成率和用户信任持续提升，消费级 AI 的付费依据有望进一步落到节省时间和完成事务上。

Kimi K3 登陆 AWS，国产模型拓展全球分发渠道。AWS 宣布月之暗面 Kimi K3 在 Amazon Bedrock 正式可用，面向编程和知识工作提供托管调用，并被纳入开放权重模型服务体系。海外云平台能够提供企业客户触达、部署和调用渠道，帮助国产模型降低独立拓展海外市场的接入门槛。企业客户对接入便利和服务稳定的需求，也可能为托管调用带来付费空间。若 kimi 的海外调用与合作模式得到验证，其他具备模型能力、推理成本及服务交付优势的国产开放权重模型，也有望通过云平台托管及分成合作拓展海外商业化空间，形成国内应用与海外企业服务共同发展的收入来源。

相关标的

大模型&云：智谱、MiniMax、阿里巴巴、腾讯控股、金山云、百度集团、优刻得、网宿科技等。

AI 应用：深信服、德才股份、迈富时、汇量科技、卓易信息、鼎捷数致、达梦数据、中控技术、金蝶国际、用友网络、海致科技集团、汉朔科技、金山办公、合合信息、福昕软件、金斯瑞生物科技、晶泰控股-P、英矽智能等。

国内算力：寒武纪、海光信息、长鑫科技、东阳光、利通电子、中芯国际、华虹半导体、华达科技、亿田智能、摩尔线程、天数智芯、沐曦股份、壁仞科技、协创数据、禾盛新材、华丰科技、京基智农、中科曙光、浪潮信息、胜蓝股份、华勤技术、盛视科技、神州数码、润泽科技、大位科技、科华数据、潍柴重机、欧陆通、杰创智能、奥尼电子等。

海外算力：工业富联、中际旭创、胜宏科技、生益科技、东山精密、新易盛、海鸥股份、方正科技、天孚通信、联瑞新材、蓝思科技、光智科技、江海股份、鹏鼎控股、中钨高新、奥士康、光迅科技、华盛昌、联讯仪器、领益智造、飞龙股份、大元泵业、德福科技、南亚新材、景旺电子、生益电子、深南电路、沪电股份、海亮股份、英维克、鼎泰高科、欧科亿、先导基电、云南锗业、源杰科技、唯科科技、海川智能、冰轮环境、麦格米特、三环集团、火炬电子、海星股份、新锐股份、杰美特、欧陆通、四方股份、阳光电源、天岳先进等；SK 海力士、英特尔、英伟达、Lumentum、闪迪、铠侠、美光、Bloom Energy。

风险提示

行业竞争加剧的风险；技术研发进度不及预期的风险；特定行业下游资本开支周期性波动的风险



内容目录

一、前沿模型密集迭代，专业任务能力持续提升.....	3
1.1 海外前沿模型同日更新.....	3
1.2 模型调用价格同步下调.....	5
1.3 能力与成本共振拓宽 Agent 任务边界.....	5
二、个人 Agent 走向终端，C 端应用进展突破.....	5
2.1 Muse 切入 C 端，个人 Agent 走向持续执行.....	5
2.2 手机助手落地，个人 Agent 加速进入大众消费场景.....	7
三、Kimi K3 登陆 AWS，国产模型拓展全球分发渠道.....	7
四、相关标的.....	8
风险提示.....	8

图表目录

图表 1: GPT-6 Sol 侧重复杂任务.....	3
图表 2: Opus 5.5 在 Coding 与 Computer Use 评测中领先.....	4
图表 3: Opus 5.5 默认投入时比 Opus 5 最大努力更好，但成本大约只有五分之一.....	4
图表 4: GPT-6 模型 API 降价过半.....	5
图表 5: Muse 使用场景示例.....	6
图表 6: Meta AI 眼镜: Muse 个人 Agent 的随身交互入口.....	6
图表 7: 全球首款 AI 智能体手机.....	7



一、前沿模型密集迭代，专业任务能力持续提升

1.1 海外前沿模型同日更新

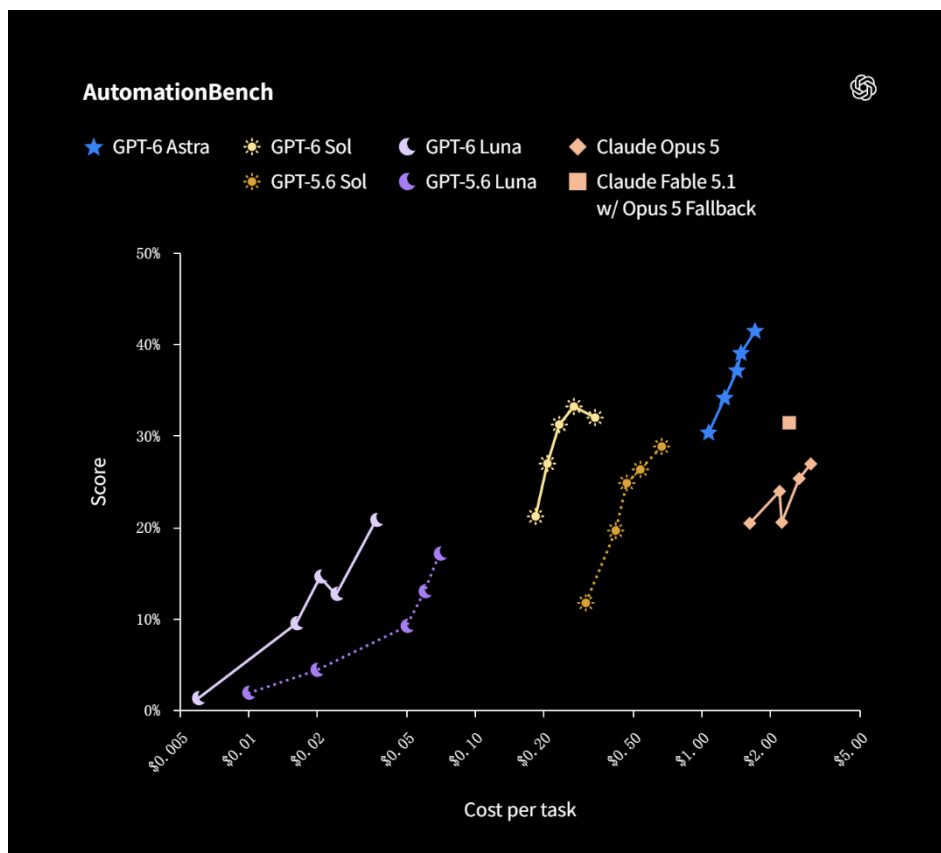
9月22日，OpenAI 推出 GPT-6 Sol、Luna，并开始向 Codex 及 ChatGPT Work 用户开放；Anthropic 同日发布 Claude Opus 5.5。OpenAI 将 Sol 定位于复杂编程和 Agent workflow，Luna 侧重范围明确、高频调用的任务；Anthropic 则将 Opus 5.5 定位于长程编程与知识工作。

GPT-6 Sol 侧重需要规划、工具调用和结果验证的复杂任务。OpenAI 的模型选择说明将 Sol 用于需求存在一定模糊性、需要多步骤执行并在较大上下文中持续推进的 Agent 任务。这一定位更贴近实际软件工程中的复杂需求：模型需要先理解已有项目，再拆解任务、修改代码并检查结果，任务价值取决于能否完成整个过程。Sol 有望成为兼顾任务质量与调用成本的主力模型，为复杂开发及专业 workflow 提供更具经济性的能力选择。

GPT-6 Luna 侧重快速、边界清晰且可重复执行的工作。OpenAI 在子 Agent 配置说明中建议，将 Luna 用于快速扫描及较轻量的子任务，并将其作为更快、成本更低的选择。我们认为，这种分层有利于在同一 workflow 中按任务难度分配模型，将资料初筛、结构化整理等高频环节交给轻量模型，复杂判断交由能力更强的模型处理，在保留关键环节质量的同时控制整体成本。

工具支持方面，Sol、Luna 均提供 105 万 Token 上下文窗口及 12.8 万 Token 最大输出，支持文本、图像输入，以及网页搜索、文件搜索、代码执行和电脑操作等工具。GPT-6 系列的 API 文档还介绍了异步工具调用与执行中追加指令：模型可在工具运行期间继续处理其他独立工作，用户也可在执行过程中修正要求。我们认为，长上下文、视觉输入与多种工具的结合，有望使模型围绕同一任务处理代码、网页和文档；执行中调整要求，则更接近实际工作中边推进、边沟通的协作方式。

图表1：GPT-6 Sol 侧重复杂任务



来源：OpenAI、国金证券研究所

Claude Opus 5.5 强化长程执行与专业知识工作。根据 Anthropic 模型文档，Opus 5.5 提供 100 万 Token 上下文窗口、常规 12.8 万 Token 最大输出，并持续启用自适应思考，可通过推理强度设置调节思考深度。对于持续时间较长、需要反复读取资料和调用工具的工作，模型能否保持目标、处理任务间依赖并检查中间结果，将直接影响最终交付质量。



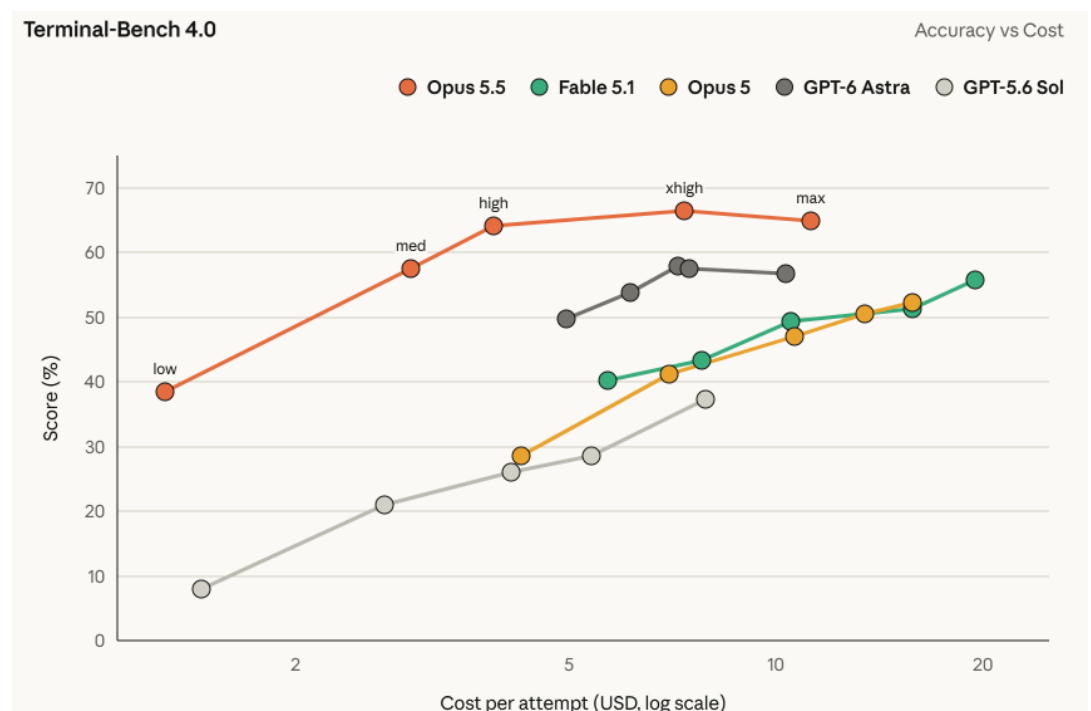
从 Anthropic 公布的同表评测看, Opus 5.5 在编程测试 Terminal-Bench 4.0 中的成绩为 66.4%, Opus 5 为 52.3%; 电脑操作测试 OSWorld 2.0 的部分得分口径分别为 81.8% 和 74.0%。其中, Opus 5.5 的 Terminal-Bench 成绩采用 xhigh 推理强度。

图表2: Opus 5.5 在 Coding 与 Computer Use 评测中领先

	Opus 5.5	Fable 5.1	Opus 5	GPT-6 Astra	GPT-5.6 Sol
Agentic coding Terminal-Bench 4.0 ¹	66.4%	55.8%	52.3%	57.9%	37.3%
Agentic coding FrontierCode v1.1 (Main)	54.4%	50.3%	48.0%	53.3%	47.5%
Agentic coding CursorBench 4.0	57.8%	51.8%	46.6%	—	41.7%
Knowledge work GDPval-AA v2.1	1846	1735	1708	1542	1588
Business workflows AutomationBench ²	40.0%	31.4%	26.9%	41.4%	28.8%
Multidisciplinary reasoning Humanity's Last Exam	67.7% with tools	65.6% with tools	63.6% with tools	57.2% with tools	—
Agentic scientific research Terminal-Bench-Science 0.1 ³	58.7%	52.6%	29.0%	64.6%	22.4%
Computer use OSWorld 2.0	81.8% partial	80.7% partial	74.0% partial	—	—
Visual chart recognition Chartography	89.0% with tools	88.4% with tools	83.4% with tools	—	—

来源: Anthropic, 国金证券研究所

图表3: Opus 5.5 默认投入时比 Opus 5 最大努力更好, 但成本大约只有五分之一



来源: Anthropic, 国金证券研究所

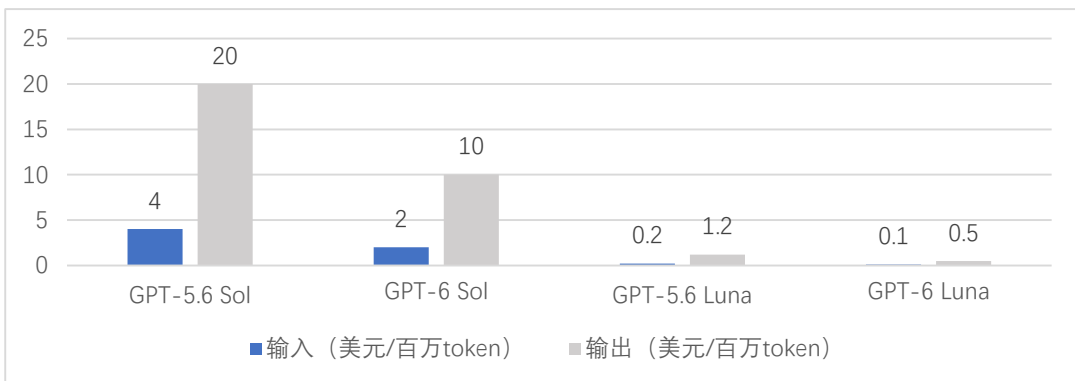


具体任务案例方面，Anthropic 披露，一名早期测试者使用 Opus 5.5 在不到一天内完成了涉及 68 万行代码的迁移。在公司季度业绩报告的内部测试中，模型需从较难检索到财报的网页环境中查找资料并撰写报告，18 次输出中有 16 次通过了对数字及引用真实性的质量检查。这些案例体现了模型承担较长工程任务和资料密集型工作的潜力，编程之外的研究、财务及办公任务有望成为能力扩展的重要方向。

1.2 模型调用价格同步下调

OpenAI 披露，Sol、Luna 的 Token 价格低于各自 GPT-5.6 前代产品；按标准处理、输入不超过 27.2 万 Token 的 API 价格，Sol 每百万输入、输出 Token 分别为 2 美元、10 美元，Luna 分别为 0.1 美元、0.5 美元。Opus 5.5 的常规输入、输出价格分别为每百万 Token 4 美元、20 美元，缓存读取为 0.2 美元。Anthropic 表示，在默认设置下，其典型工作负载测试中 Opus 5.5 的运行成本较 Opus 5 降低 40%。

图表4：GPT-6 模型 API 降价过半



来源：OpenAI、国金证券研究所

1.3 能力与成本共振拓宽 Agent 任务边界

我们认为，单次任务成本下降，有望提高包含多轮推理、工具调用及结果校验的 Agent 工作流的经济性，使更多原本成本较高的任务具备应用条件。前沿模型竞争正进一步延伸至持续执行任务，Sol、Luna 扩大不同成本预算下的任务覆盖，Opus 5.5 强化复杂任务执行，能力与效率的共同改善有望拓宽 Agent 可承担的工作范围。

二、个人 Agent 走向终端，C 端应用进展突破

2.1 Muse 切入 C 端，个人 Agent 走向持续执行

9 月 8 日，Meta 发布个人 Agent 产品 Muse。根据官方介绍，Muse 运行于配有浏览器的独立虚拟电脑 Muse Secure VM，可跨日常应用执行任务；用户可在 Muse 应用或 WhatsApp 中与其沟通，无需具备技术经验。我们认为，将运行环境和工具操作整合到产品内部，有望让非专业用户通过自然语言直接委托事务，降低持续使用 Agent 的门槛。

家庭事务是 Muse 产品团队展示的具体应用场景。在返校准备案例中，团队成员让 Muse 持续查看学校邮件和学区网站，将重要日期加入家庭日历，把所需用品放入购物车，并查找孩子想要的打折卫衣、预订庆祝开学的晚餐。Muse 还从邮件中发现了一项即将截止的体育选拔报名，并主动发出提醒，家人随后完成了表格提交。这类需求包含信息筛选、日期提取、购物与预约等多个环节，个人 Agent 有望减少用户在邮件、日历和购物网站之间反复切换的时间。

Muse 可根据用户偏好协助准备聚餐。Meta 给出的场景是，Muse 可将用户在 Instagram 收藏的食谱视频整理为食材采购清单，为聚餐提出菜单建议，并记住朋友的饮食限制。我们认为，个人偏好与实际任务相结合，有望减少每次使用时重复补充背景负担，使 Agent 在持续互动中逐步适应用户的日常安排。



图表5: Muse 使用场景示例

Ideas

I can line up Kiran's steak dinner and birthday weekend

Kiran turns 11 Friday. I can pull together the steak dinner options with phone numbers, plus the September 11 to 13 events that actually fit, into one birthday-weekend plan you can approve or adjust.

I can split one Sunday cook into three plates

I can build your two-week starter around your 16:8 window and 100g protein target, with a hotel-friendly protein list for Menlo weeks and one Sunday batch cook at home that splits into three plates: yours high-protein, Kyri's low-carb, and the boys' with carb sides. If your Monday weigh-in trend stalls, I can adjust the next week's portions.

I can track India Christmas quotes live

I pull replies from email, extract per-person costs for Delhi / Jaipur / Kolkata, and hand you a running ledger with deposit deadlines.

I can block focus time for your open commitments

Each Monday I pull your Commitment Scanner ledger, match open promises against Calendar including Menlo travel, and draft time blocks to get them done, with reschedule options if you're overloaded.

来源: Meta, 国金证券研究所

日常管理工具也可由 Muse 按需生成。产品设计说明列举了支出追踪工具、交互式学习指南和用于观察睡眠规律的看板等输出形式,并介绍 Muse 可以生成文档、PDF 和网页。这意味着非专业用户有望直接提出管理目标并获得可使用的工具,例如围绕个人支出或学习计划组织信息, AI 编程能力因而能够通过消费级产品服务于更广泛的人群。

Muse 支持后台持续执行。Meta 介绍, Muse 可在用户关闭应用后继续工作,按照计划或相关事件推进任务,并在出现重要变化或需要用户输入时通知用户。我们认为,对于报名提醒、日程管理和长期计划等需求, Agent 能否持续跟进比单次回答是否完整更为重要,有望推动使用关系从临时问答延伸为长期事务委托。

在操作权限方面, Meta 的安全说明指出, 用户可区分邮件等服务的读取与写入权限, Sentinel 负责判断连接器操作及网络请求是否允许执行;需要审批时,操作会暂停并由用户确认。我们认为,个人 Agent 要进入日历、邮件及交易等场景,需要让用户能够理解并控制其行为,权限管理和结果可检查性将影响用户愿意委托的任务范围。

终端入口仍在扩展。9月23日, Meta 在 Connect 期间宣布将 Muse 引入 AI 眼镜,使用户能够以免手持方式与个人 Agent 交互,官方列举的方向包括健康习惯管理和繁忙日程安排。眼镜等随身设备有望降低发起任务的操作成本,让个人 Agent 更自然地参与出行、家庭和日常事务,也为模型能力转化为消费级服务提供更多入口。

图表6: Meta AI 眼镜: Muse 个人 Agent 的随身交互入口





来源：Meta，国金证券研究所

相较于 Codex、Claude Code 等产品所代表的专业任务使用路径，Muse 的核心产业意义在于，将跨应用操作、持续执行及个性化记忆整合为面向非专业用户的服务。我们认为，家庭事务、采购准备与个人管理工具等案例说明，Agent 的潜在需求并不限于编程和专业办公，普通用户日常生活中分散、重复、需要跟进的事务，同样可能形成稳定的使用场景。若实际任务完成率及用户信任持续提升，C 端 AI 应用的付费依据有望进一步落到节省时间和完成事务上。

2.2 手机助手落地，个人 Agent 加速进入大众消费场景

国内方面，据红星资本局 9 月 14 日报道，豆包手机助手消费者版正式发布，首款搭载机型为努比亚 NaviX Ultra；据《科创板日报》9 月 16 日报道，该机型已正式上市开售。我们认为，系统级手机助手为个人 Agent 提供了更贴近日常使用习惯的入口，有望推动相关能力向大众消费场景渗透。

图表7：全球首款 AI 智能体手机



来源：努比亚手机，国金证券研究所

三、Kimi K3 登陆 AWS，国产模型拓展全球分发渠道

9 月 18 日，AWS 宣布月之暗面 Kimi K3 在 Amazon Bedrock 正式可用，面向编程和知识工作。AWS 将 K3 纳入其开放权重模型服务体系，用户可通过平台调用。

我们认为，海外云平台能够提供企业客户触达、部署及调用渠道，帮助国产模型降低独立拓展海外市场的接入门槛。开放权重与商业化并不矛盾，企业客户对部署便利、服务稳定性和运维支持的需求，有望为托管调用服务带来付费空间，模型厂商亦有机会通过相应合作安排获取收入。



K3 登陆 Bedrock 为观察国产模型海外商业化提供了新的样本,后续应关注实际调用需求、客户留存及合作收益能否形成持续贡献。若其海外调用与合作模式得到验证,其他具备模型能力、推理成本及服务交付优势的国产开放权重模型,也有望通过云平台托管及分成合作拓展海外商业化空间,形成国内应用与海外企业服务共同发展的收入来源。

四、相关标的

大模型&云: 智谱、MiniMax、阿里巴巴、腾讯控股、金山云、百度集团、优刻得、网宿科技等。

AI 应用: 深信服、德才股份、迈富时、汇量科技、卓易信息、鼎捷数致、达梦数据、中控技术、金蝶国际、用友网络、海致科技集团、汉朔科技、金山办公、合合信息、福昕软件、金斯瑞生物科技、晶泰控股-P、英矽智能等。

国内算力: 寒武纪、海光信息、长鑫科技、东阳光、利通电子、中芯国际、华虹半导体、华达科技、亿田智能、摩尔线程、天数智芯、沐曦股份、壁仞科技、协创数据、禾盛新材、华丰科技、京基智农、中科曙光、浪潮信息、胜蓝股份、华勤技术、盛视科技、神州数码、润泽科技、大位科技、科华数据、潍柴重机、欧陆通、杰创智能、奥尼电子等。

海外算力: 工业富联、中际旭创、胜宏科技、生益科技、东山精密、新易盛、海鹏股份、方正科技、天孚通信、联瑞新材、蓝思科技、光智科技、江海股份、鹏鼎控股、中钨高新、奥士康、光迅科技、华盛昌、联讯仪器、领益智造、飞龙股份、大元泵业、德福科技、南亚新材、景旺电子、生益电子、深南电路、沪电股份、海亮股份、英维克、鼎泰高科、欧科亿、先导基电、云南锗业、源杰科技、唯科科技、海川智能、冰轮环境、麦格米特、三环集团、火炬电子、海星股份、新锐股份、杰美特、欧陆通、四方股份、阳光电源、天岳先进等; SK 海力士、英特尔、英伟达、Lumentum、闪迪、铠侠、美光、Bloom Energy。

风险提示

行业竞争加剧的风险:

云厂商、模型厂商及芯片企业降价或推出替代方案,可能压缩份额与利润率。

技术研发进度不及预期的风险:

模型能力、Agent 应用开发若迭代放缓可能导致商业化进展速度不及预期。

特定行业下游资本开支周期性波动的风险:

互联网云厂商、运营商、数据中心及金融、政府客户算力基础设施投资若放缓可能导致上游环节增长不及预期。



行业投资评级的说明：

买入：预期未来 3—6 个月内该行业上涨幅度超过大盘在 15%以上；

增持：预期未来 3—6 个月内该行业上涨幅度超过大盘在 5%—15%；

中性：预期未来 3—6 个月内该行业变动幅度相对大盘在 -5%—5%；

减持：预期未来 3—6 个月内该行业下跌幅度超过大盘在 5%以上。



特别声明：

国金证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

任何形式的复制、转发、转载、引用、修改、仿制、刊发，或以任何侵犯本公司版权的其他方式使用。经过书面授权的引用、刊发，需注明出处为“国金证券股份有限公司”，且不得对本报告进行任何有悖原意的删节和修改。

本报告的产生基于国金证券及其研究人员认为可信的公开资料或实地调研资料，但国金证券及其研究人员对这些信息的准确性和完整性不作任何保证。本报告反映撰写研究人员的不同设想、见解及分析方法，故本报告所载观点可能与其他类似研究报告的观点及市场实际情况不一致，国金证券不对使用本报告所包含的材料产生的任何直接或间接损失或与此有关的其他任何损失承担任何责任。且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，在不作事先通知的情况下，可能会随时调整，亦可因使用不同假设和标准、采用不同观点和分析方法而与国金证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。

本报告仅为参考之用，在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告提及的任何证券或金融工具均可能含有重大的风险，可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可能会受汇率影响而波动。过往的业绩并不能代表未来的表现。

客户应当考虑到国金证券存在可能影响本报告客观性的利益冲突，而不应视本报告为作出投资决策的唯一因素。证券研究报告是用于服务具备专业知识的投资者和投资顾问的专业产品，使用时必须经专业人士进行解读。国金证券建议获取报告人员应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。报告本身、报告中的信息或所表达意见也不构成投资、法律、会计或税务的最终操作建议，国金证券不就报告中的内容对最终操作建议做出任何担保，在任何时候均不构成对任何人的个人推荐。

在法律允许的情况下，国金证券的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供多种金融服务。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。国金证券并不因收件人收到本报告而视其为国金证券的客户。本报告对于收件人而言属高度机密，只有符合条件的收件人才能使用。根据《证券期货投资者适当性管理办法》，本报告仅供国金证券股份有限公司客户中风险评级高于 C3 级（含 C3 级）的投资者使用；本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具，本报告的收件人须保持自身的独立判断。使用国金证券研究报告进行投资，遭受任何损失，国金证券不承担相关法律责任。

若国金证券以外的任何机构或个人发送本报告，则由该机构或个人为此发送行为承担全部责任。本报告不构成国金证券向发送本报告机构或个人的收件人提供投资建议，国金证券不为此承担任何责任。

此报告仅限于中国境内使用。国金证券版权所有，保留一切权利。

上海	北京	深圳
电话：021-80234211	电话：010-85950438	电话：0755-86695353
邮箱：researchsh@gjzq.com.cn	邮箱：researchbj@gjzq.com.cn	邮箱：researchsz@gjzq.com.cn
邮编：201204	邮编：100005	邮编：518000
地址：上海浦东新区芳甸路 1088 号 紫竹国际大厦 5 楼	地址：北京市东城区建国内大街 26 号 新闻大厦 8 层南侧	地址：深圳市福田区金田路 2028 号皇岗商务中心 18 楼 1806



【小程序】
国金证券研究服务



【公众号】
国金证券研究