AI 事件点评

优于大市

DeepSeek 发布高性价比开源模型，有望拉平模型差距、加速 AI 云与应用发展

◆ 行业研究·行业快评 ◆ 互联网·互联网Ⅱ ◆ 投资评级：优于大市（维持）

证券分析师： 张伦可 0755-81982651 zhanglunke@guosen.com.cn 执证编码：S0980521120004
联系人： 刘子谭 liuzitan@guosen.com.cn

事项：

2024 年 12 月 26 日，DeepSeek 发布开源模型 V3，训练成本仅 557.6 万美元，性能却能对标 GPT-4o。2025 年 1 月 20 日，DeepSeek 继续发布开源模型 R1，训练周期仅两个月，在数学、代码、自然语言推理等任务上性能比肩 OpenAI o1 正式版。对比 OpenAI 与谷歌每年数十亿美元 AI 预算，以及 25 年 1 月 22 日发布的计划投资高达 5000 亿建设 AI 相关基础设施的“星际之门”项目，低成本的 Deepseek 引起海内外强烈关注与反思。2025 年 1 月 27 日，DeepSeek 事件继续发酵，并在资本市场引发强烈反应，美国主要 AI 相关科技股均遭遇股市地震，其中英伟达跌近 17%，单日市值蒸发约 6000 亿美元。伴随热度，DeepSeek 应用迅速登顶 15 个国家和地区的苹果应用商店免费 APP 下载排行榜，火爆出圈，截止目前 20 天已经实现 2000 万下载量。

国信互联网观点：1）**对 AI 模型层，Deepseek 的开源与高性价比将显著加剧大模型层竞争，降低大模型门槛、利好追赶者。**Deepseek 打破已有过度依赖算力与标注数据的训练模式，架构上的“捷径”对于利用大算力与标注数据作为护城河的领先模型是巨大的挑战，为其他模型研发者提供了新的技术思路和追赶方式。DeepSeek 不仅主打高性价比还将模型全部开源，这将极大推动开源生态的繁荣，也意味着模型层竞争更加激烈，促使模型开发者不断提升模型性能、降低成本。**2）对 AI 芯片算力层：短期降低先进算力需求预期，ASIC 和国产芯片厂商拥有了更长的时间窗口。**DeepSeek 通过创新的训练方法，如在预训练阶段加入强化学习，证实了在有限算力下实现前沿 AI 能力的可能性，部分企业预计会减少对大规模算力基础设施的激进投入，短期降低对英伟达的先进算力需求预期，也使得 ASIC 和国产芯片厂商拥有了更长的时间窗口，算力市场预计走向多元化发展。3）**对云厂商：利好云厂商下游需求增长，显著缩小了云厂 AI 前期投入与应用兑现之间的时间与资源成本，有望进一步提升国产云厂商盈利能力。**云厂商集算力供给、大模型研发与 AI 应用为一体，DeepSeek 高性价比、开源模型虽然削弱模型层竞争壁垒，但为云厂商提供了更具性价比的 API，如 R1 上线短短两周，腾讯云、华为云、微软 Azure 和亚马逊 AWS 均已上线相关服务。目前云厂商需要承受巨大 AI 前期投入与应用业绩兑现的时间差，如近期星门计划微软未参与，表明 AI 投入已经达到短期经济体投入能力的上限（今年微软 Capex 800-900 亿 vs1000 亿盈利，Meta Capex 600-650 亿 vs660 亿盈利），而 Deepseek 的技术路线使得云厂可以更加平衡 AI 的 ROI、模型的成本效益和实用性。Deepseek 拓展 AI 应用场景，激发新的算力需求，有望显著带动 AI 云增长。对国内云厂商，Deepseek 将加速企业数字化转型上云，规模效应下进一步提升云业务利润率。4）**对 AI 应用层：降低 AI 应用研发与落地的成本，加速 AI 应用发展，Agent 与端侧 AI 预期增强。**DeepSeek 模型使得开发利用大模型训练、调优的门槛降低，高性价比的模型使得 AI 应用研发和使用成本显著降低，加速垂类模型发展、利好 AI 在各行业的渗透。DeepSeek-R1 具备深度思考能力，有望成为互动场景或工作任务的“Agent 智能体”大脑。同时，Deepseek 将同等模型能力所需的算力大幅压缩，有望部署到端侧，加速端侧 AI 的落地。

投资建议：Deepseek 有望加速国内云厂商大模型追赶速度、拉平模型层差距。同时加速国内企业上云、利好云厂商下游需求增长。显著缩小云厂 AI 前期投入与应用兑现之间的时间与资源成本，规模效应下有望进一步提升国产云厂商利润率。因此，我们推荐国内云厂商龙头阿里巴巴，具备云业务与优质社交场景生态的腾讯控股，以及海外云厂商龙头亚马逊。

评论：

◆ Deepseek 模型介绍

1）**DeepSeek-V3：**2024 年 12 月 26 日发布，在多项测试中达到了与 GPT-4 和 Claude 3.5 等顶级模型相当的性能水平。其采用多头潜在注意力（MLA）机制，通过压缩注意力机制中的键和值，有效减少推理阶段的计算量，提高模型运行效率。根据官网介绍训练成本仅 557.6 万美元，性能却与 GPT - 4o 媲美，对比 OpenAI 训练 GPT-4 花费约 1 亿美元性价比显著。DeepSeek V3 使用的 token 数量约为 14.8 万亿（1480B），对比 GPT-4 MoE 使用了 13 万亿（1300B）token，数量相当。

图1：**DeepSeek-V3 训练成本**

- **Training Cost**: $5.6 million
- **Training Duration**: 57 days
- **GPU Usage**: 2.788 million H800 GPU hours

Training Costs	Pre-Training	Context Extension	Post-Training	Total
in H800 GPU Hours	2664K	119K	5K	2788K
in USD	$5.328M	$0.238M	$0.01M	$5.576M

Table 1 | Training costs of DeepSeek-V3, assuming the rental price of H800 is $2 per GPU hour.

资料来源：Deepseek 官网、国信证券经济研究所整理

2）**DeepSeek-R1：**训练基于 DeepSeek-V3 的基座模型，通过强化学习从 V3 进化而来，推理过程包含大量反思和验证，思维链长度可达数万字。DeepSeek-R1 在后训练阶段大规模使用了强化学习技术，在仅有极少标注数据的情况下，极大提升了模型推理能力。遵循 MIT License，即允许用户通过蒸馏技术借助 R1 训练其他模型。在基准测试中表现与 OpenAI 的 o1 模型相当，但价格却显著低于 o1，性价比更具优势。

表1：DeepSeek 模型情况介绍

模型名称	DeepSeek-V2	DeepSeek-V3	DeepSeek-R1
发布时间	2024 年 5 月	2024 年 12 月 26 日	2025 年 1 月 20 日
参数量	总参数 2360 亿，激活参数 210 亿	拥有 6710 亿参数(约为 GPT-4 MoE 的 1/3)，激活参数为 370 亿(约为 GPT-4 MoE 的 1/7)	DeepSeek-R1-Zero 和 DeepSeek-R1 均为 6710 亿参数(MoE 架构，每个 token 激活 370 亿参数）。 同时还蒸馏了 6 个小模型，参数范围从 15 亿到 700 亿不等。
性能	综合性能达 GPT-4 级别。 ① 中文综合能力在众多开源模型中最强，超过 GPT-4，与 GPT-4-Turbo、文心 4.0 等闭源模型在评测中处于同一梯队； ② 英文综合能力与最强的开源模型 LLaMA3-70B 处于同一梯队，超过最强 MoE 开源模型 Mixtral8x22B。 ③ 在 8 卡 H800 机器上，输入吞吐量超过每秒 10 万 tokens，输出超过每秒 5 万 tokens。	① 在知识类任务、算法类代码场景、工程类代码场景、中文能力、数学能力等方面有优势。 ② 在多语言编程测试排行榜中，已超越 Anthropic 的 Claude 3.5 Sonnet 模型，仅次于 OpenAI o1 大模型； ③ 在数学能力方面超过了所有开源闭源模型； ④ 在编程任务中通过率达到 40%，高于 Llama3.1 的 31%和 Claude3.5 的 33%； ⑤ 中文多语言理解测试得分为 89 分，远超 Llama3.1 的 74 分。 ⑥ 生成速度相比 V2.5 模型实现了 3 倍的提升，达到每秒吞吐量 60token(V2.5 为 20TPS)；	① 在数学、代码、自然语言推理等任务上，性能比肩 OpenAI o1 正式版。 ② 在 AI ME2024 基准测试中，DeepSeek-R1-Zero 准确率为 71%，DeepSeek-R1 准确率提升至 79.8%； ③ 在 MATH-500 测试中，DeepSeek-R1 准确率为 97.3%，OpenAI o1 为 96.4%。

训练花费	未明确提及具体训练时长，但训练参数量高达 8.1 万亿个 token，计算量仅为 Meta Llama 3 70B 的 1/5。	完整训练消耗了 278.8 万个 GPU 小时，**训练成本为 557.6 万美元**。	预训练在 2048 块英伟达 H800 GPU 集群上运行 55 天完成。

资料来源：Deepseek 官网、华尔街见闻，国信证券经济研究所整理

◆ Deepseek 与 GPT、Gemini、Llama 等竞品对比

优势：

1）**训练与推理的高性价比：**DeepSeek-V3 的训练成本仅为 557.6 万美元，远低于 GPT-4o 等模型所需的数亿美元。其 API 调用价格也显著低于 GPT-4o，推理成本低至每百万 tokens 0.014 美元。

2）**响应速度快：**DeepSeek-V3 与 R1 采用 MoE 架构，每个 Token 仅激活 370 亿参数，显著减少了计算量，提高了推理速度，低延迟和高扩展性使其在需要快速响应的应用场景中表现优异。

3）**数学推理和编程任务表现优异：**在数学推理和编程任务中，DeepSeek-V3 表现出色，中文能力突出，更适合中文语境下的任务处理。DeepSeek-R1 在高难度推理任务中表现突出，例如在 AIME 2024 和 MATH-500 等基准测试中，得分高于 OpenAI 的 o1 模型。

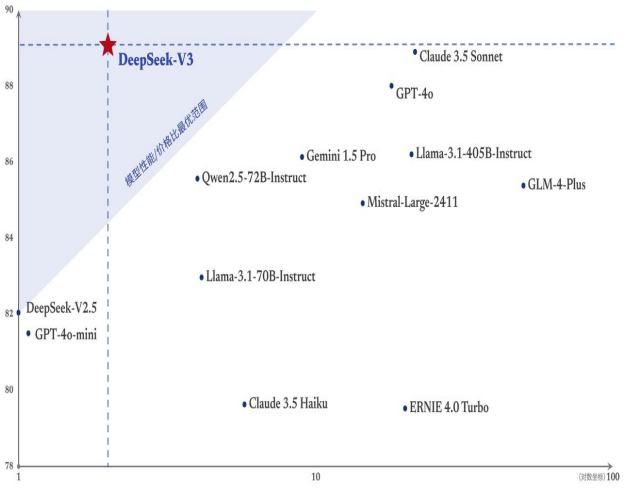
4）**开源与灵活性高：**DeepSeek-V3 开源，允许开发者自行部署、训练、微调和应用模型，提供了更多的自由和灵活性。

劣势：

1）**通用综合性以及多模态能力稍弱：**在通用文本生成和创意应用中，DeepSeek-V3 与 GPT-4o 等模型相比还需要更多实际应用场景的验证。在多模态处理能力方面，如对图像、音频的处理能力，相较于 Gemini Ultra 存在差距。

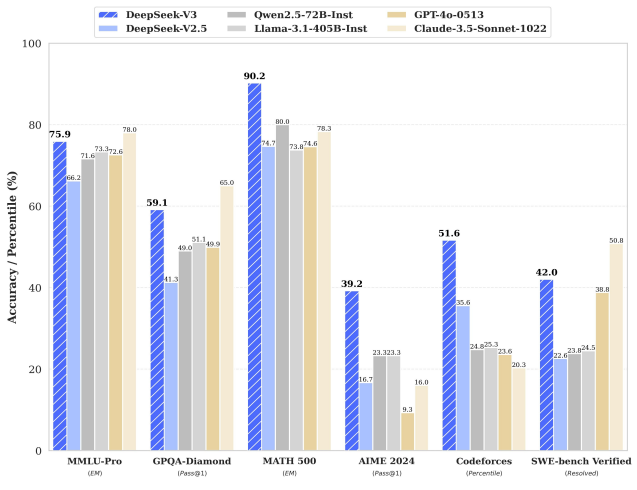
2）**上下文窗口较小：**DeepSeek-V3 的上下文窗口最多为 128K 的上下文窗口，与 Gemini Ultra 的 1000K 超长上下文窗口相比，DeepSeek 模型的上下文窗口长度仍显不足，在处理一些极端长文本任务以及多任务时表现不足。

图2：**MMLU ReduxZeroEval 得分与输入 API 价格变化（￥/1M Tokens）**



资料来源：Deepseek 官网、国信证券经济研究所整理

图3：**Deepseek 与其他大模型测评得分比较**



资料来源：Deepseek 官网、国信证券经济研究所整理

◆ **Deepseek 高性价比开源模型带来的影响**

1）对 AI 模型层：开源与高性价比特点显著加剧大模型层竞争，降低大模型门槛、利好追赶者

打破已有过度依赖算力与标注数据的训练模式，显著降低大模型准入壁垒，利好模型追赶者：在此之前，AI 大模型发展遵循堆算力和数据的模式，DeepSeek 打破了这种传统路径依赖，展示了通过改进模型架构和训练方法，如大规模使用强化学习技术，即使在数据标注量少的情况下，也能极大提升模型推理能力。架构上的“捷径”对于利用大算力与标注数据作为护城河的领先模型是巨大的挑战，为其他模型研发者提供了新的技术思路和追赶方式，预计将引发一波模仿、探索高效训练方法和创新模型架构，从而加速追赶的趋势。

造成模型层同质化，加剧大模型从能力、迭代周期到性价比全面竞争，促进开源生态发展：DeepSeek 不仅主打高性价比还将模型全部开源，多家团队已宣布复现其训练过程，这将极大推动开源生态的繁荣，也意味着模型层竞争更加激烈，闭源模型不再拥有绝对优势，促使模型开发者不断提升模型性能、降低成本，以在市场中拥有更多客户和使用量。

2）对 AI 算力层：短期降低对先进算力需求预期，ASIC 和国产芯片厂商拥有了更长的时间窗口

DeepSeek 通过创新的训练方法，如在预训练阶段加入强化学习，用较少的计算资源就达到了接近 GPT-o1 的性能，这使业界开始反思大算力在 AI 发展尤其是大模型训练过程中的必要性，部分企业预计会减少对大规模算力基础设施的激进投入。短期内可能会局部缓解算力压力，但长期来看，随着 AI 能力的边界扩展（如多模态、复杂推理、通用人工智能）以及应用场景的爆发式扩展，算力需求仍将增长。

另一方面，也为国产显卡和 ASIC 芯片带来了机会。因为 DeepSeek 的 RL 策略对并行计算需求下降 40%，这使得国产算力硬件有机会凭借成本和服务优势在市场中占据一席之地。客户可以根据实际应用场景灵活进行定制化芯片开发，算力市场预计走向多元化发展。

3）对云厂商：利好云厂商下游需求增长，显著缩小了云厂 AI 前期巨大投入与应用兑现之间的时间与资源成本，有望进一步提升国产云厂商利润率

目前云厂商自身集算力供给、大模型研发与 AI 应用为一体，DeepSeek 高性价比、开源模型的发布虽然削弱模型层竞争壁垒，加大 AI 云格局的不确定性，但为云厂商提供了更具性价比的 AI 方案。DeepSeek 高性价比、开源模型的发布削弱了云厂商/大模型厂商在 AI 模型服务层面的壁垒，让大模型差距更小。但 DeepSeek 利好国内外大模型向 OpenAI 等一流模型追赶。同时，DeepSeek 的高性价比开源模型为云厂商提供了更高效、低成本的 API 调用方案/AI 解决方案，如 R1 上线短短两周，腾讯云、华为云、微软 Azure 和亚马逊 AWS 均已上线 DeepSeek-R1 相关服务，并提供了便捷的部署和调用方式。

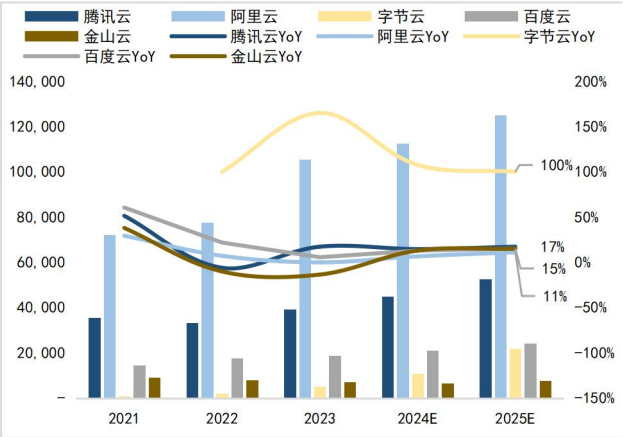
Deepseek 缩小了云厂 AI 前期投入与应用兑现之间的时间与资源成本。近期星门计划微软未参与，表明 AI 投入已经达到短期经济体投入能力的上限（今年微软 Capex 800-900 亿 vs1000 亿盈利，Meta Capex 600-650 亿 vs660 亿盈利），而 Deepseek 的技术路线使得云厂在高额前期投入的重压下有了喘息之机，更好地去评估 AI 板块的 ROI，更加注重模型的成本效益和实用性，加大在模型部署、优化和管理的投入，加强对 AI 应用场景的拓展和落地。

Deepseek 拓展 AI 应用场景，激发新的算力需求，带动 AI 云增长。从 Deepseek AI 对话助手 20 天已经实现 2000 万下载量，预计 DeepSeek 模型的普及将赋能更多应用场景，从而推动了云服务厂商的业务增长，云服务厂商既是技术降本的受益者，也是放大降本效应的推动者。

对国内云厂商，Deepseek 将加速企业数字化转型上云，规模效应下进一步提升云业务利润率。AI 背景下数字化和云化是必然的趋势，且 AI 云的技术壁垒、相关配套服务的利润空间和整体市场空间显著高于传统云，而 Deepseek 模型的出现加速了各行业的数字化转型进程。预计将带动国内云厂商利润率向海

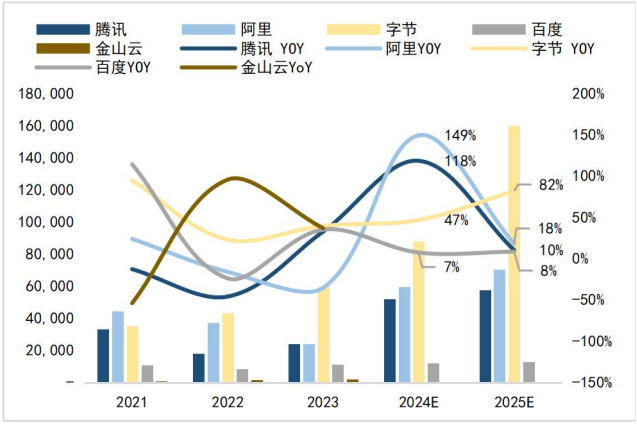
外云厂商靠拢。

图4：国内云厂商云收入与同比变化预测(百万元/%)



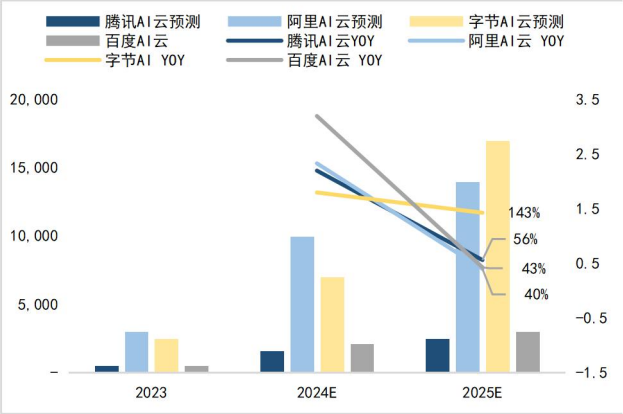
资料来源：各公司财报、国信证券经济研究所整理

图5：国内云厂商资本开支与同比变化预测(百万元/%)



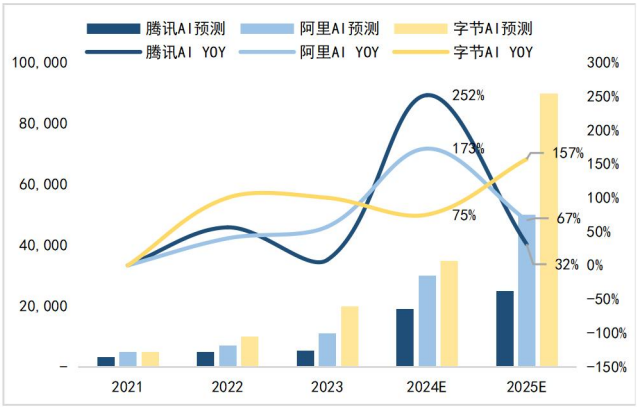
资料来源：各公司财报、国信证券经济研究所整理

图6：**国内云厂商 AI 云收入与同比变化预测(百万元/%)**



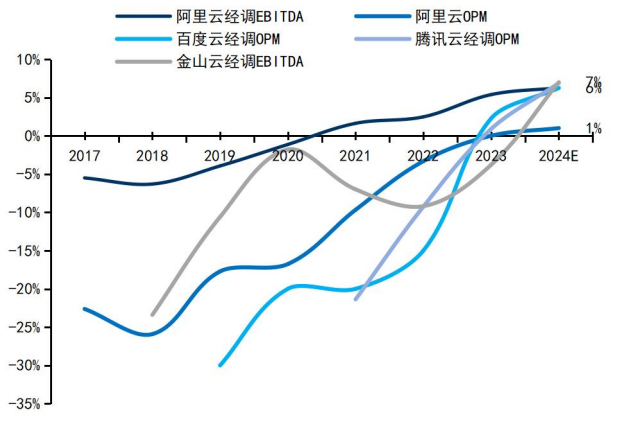
资料来源：各公司财报会、国信证券经济研究所整理预测

图7：国内云厂商 AI 资本开支与同比变化预测(百万元/%)



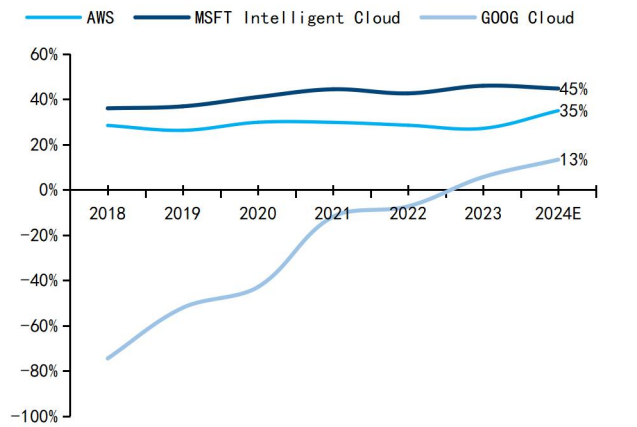
资料来源：各公司财报、国信证券经济研究所整理预测

图8：**国内云厂商经营利润率或 EBITDA 率变化**



资料来源：各公司财报会、国信证券经济研究所整理预测

图9：**海外云厂商经营利润率变化**



资料来源：各公司财报、国信证券经济研究所整理预测

4）**对 AI 应用层：降低 AI 应用研发与落地的成本，加速 AI 应用发展，Agent 与端侧 AI 预期增强**

降低垂类模型/应用开发门槛，加速 AI 应用/Agent 在各个场景落地：DeepSeek 模型的低成本优势使得开发利用大模型训练、调优的门槛降低，企业无需投入巨额资金用于模型训练就能获取高性能模型，加速垂类模型发展，利好 AI 在各行业的渗透，如医疗、教育等领域，催生出更多创新的 AI 应用场景和商业模式。且 DeepSeek-R1 具备深度思考和出色的推理能力、且成本低，有望成为互动场景或工作任务的“Agent 智能体” 大脑，利于 AI Agent 在各个场景普及。

显著降低推理成本，提升应用端盈利能力：DeepSeek 高性价比的模型使得 AI 应用研发和使用成本显著降低，从而提升企业盈利能力，应用厂商也可以有更多资源进行进行产品优化和市场拓展。

将同等模型能力所需的算力极度压缩，为 AI 端侧落地提供技术基础：Deepseek 将同等模型能力所需的算力大幅压缩，模型提供的高性价比和高效推理能力使其能够更广泛地应用于端侧设备，预计将加速了端侧 AI 应用的落地。

◆ **风险提示：**

AI 模型技术发展不及预期，行业竞争加剧，大模型幻觉、伦理等安全性问题。

相关研究报告：

《海外垂类 AI 专题(8)：AI 激发 SaaS 新一轮产品创新周期，美股软件板块反转确立》 ——2024-12-07
《AI 对巨头业务的赋能和影响：云计算、广告、AI Coding 变化最明显》 ——2024-11-20
《美股科技互联网 24Q3 财报总结：云持续供不应求，AI 促进数字广告行业增长》 ——2024-11-06
《电商行业点评-直播电商迎来规范发展新阶段，预计双 11 阿里京东 GMV 增速趋近大盘》 ——2024-09-30
《美股科技互联网 24Q2 财报总结：云周期逐步回升，AI 促进广告加速增长》 ——2024-08-07

免责声明

分析师声明

作者保证报告所采用的数据均来自合规渠道；分析逻辑基于作者的职业理解，通过合理判断并得出结论，力求独立、客观、公正，结论不受任何第三方的授意或影响；作者在过去、现在或未来未就其研究报告所提供的具体建议或所表述的意见直接或间接收取任何报酬，特此声明。

国信证券投资评级

投资评级标准	类别	级别	说明
报告中投资建议所涉及的评级（如有）分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后 6 到 12 个月内的相对市场表现，也即报告发布日后的 6 到 12 个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。A 股市场以沪深 300 指数（000300.SH）作为基准；新三板市场以三板成指（899001.CSI）为基准；香港市场以恒生指数（HSI.HI）作为基准；美国市场以标普 500 指数（SPX.GI）或纳斯达克指数（IXIC.GI）为基准。	股票投资评级	优于大市	股价表现优于市场代表性指数 10%以上
		中性	股价表现介于市场代表性指数±10%之间
		弱于大市	股价表现弱于市场代表性指数 10%以上
		无评级	股价与市场代表性指数相比无明确观点
	行业投资评级	优于大市	行业指数表现优于市场代表性指数 10%以上
		中性	行业指数表现介于市场代表性指数±10%之间
		弱于大市	行业指数表现弱于市场代表性指数 10%以上

重要声明

本报告由国信证券股份有限公司（已具备中国证监会许可的证券投资咨询业务资格）制作；报告版权归国信证券股份有限公司（以下简称“我公司”）所有。本报告仅供我公司客户使用，本公司不会因接收人收到本报告而视其为客户。未经书面许可，任何机构和个人不得以任何形式使用、复制或传播。任何有关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以我公司向客户发布的本报告完整版本为准。

本报告基于已公开的资料或信息撰写，但我公司不保证该资料及信息的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映我公司于本报告公开发布当日的判断，在不同时期，我公司可能撰写并发布与本报告所载资料、建议及推测不一致的报告。我公司不保证本报告所含信息及资料处于最新状态；我公司可能随时补充、更新和修订有关信息及资料，投资者应当自行关注相关更新和修订内容。我公司或关联机构可能会持有本报告中所提到的公司所发行的证券并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问或金融产品等相关服务。本公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中意见或建议不一致的投资决策。

本报告仅供参考之用，不构成出售或购买证券或其他投资标的要约或邀请。在任何情况下，本报告中的信息和意见均不构成对任何个人的投资建议。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。投资者应结合自己的投资目标和财务状况自行判断是否采用本报告所载内容和信息并自行承担风险，我公司及雇员对投资者使用本报告及其内容而造成的一切后果不承担任何法律责任。

证券投资咨询业务的说明

本公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询，是指从事证券投资咨询业务的机构及其投资咨询人员以下列形式为证券投资人或者客户提供证券投资分析、预测或者建议等直接或者间接有偿咨询服务的活动：接受投资人或者客户委托，提供证券投资咨询服务；举办有关证券投资咨询的讲座、报告会、分析会等；在报刊上发表证券投资咨询的文章、评论、报告，以及通过电台、电视台等公众传播媒体提供证券投资咨询服务；通过电话、传真、电脑网络等电信设备系统，提供证券投资咨询服务；中国证监会认定的其他形式。

发布证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。

国信证券经济研究所

深圳
深圳市福田区福华一路 125 号国信金融大厦 36 层
邮编：518046　总机：0755-82130833

上海
上海浦东民生路 1199 弄证大五道口广场 1 号楼 12 层
邮编：200135

北京
北京西城区金融大街兴盛街 6 号国信证券 9 层
邮编：100032