

2024年10月27日



华鑫证券
CHINA FORTUNE SECURITIES

高通推出 Snapdragon 移动 GPU, Anthropic 发布升级版大模型—计算机行业周报

推荐(维持)

投资要点

分析师：宝幼琛 S1050521110002

baoyc@cfsc.com.cn

行业相对表现

表现	1M	3M	12M
计算机(申万)	23.0	42.4	3.6
沪深300	6.8	16.0	11.1

市场表现



资料来源：Wind，华鑫证券研究

相关研究

- 1、《计算机行业周报：安卓首款3nm AI 芯片天玑 9400 发布，零一万物正式对外发布新旗舰模型》2024-10-21
- 2、《计算机行业周报：英特尔发布新一代处理器，ByteDance Research 提出机器人模型》2024-10-14
- 3、《计算机行业周报：板块有望迎来估值强修复，OpenAI 宣布推出 canvas》2024-10-08

算力：算力租赁价格平稳，高通推出 Snapdragon 8 Elite

高通本周在毛伊岛举行的年度 Snapdragon 峰会上表示，包括华硕、荣耀、iQOO、OnePlus、OPPO、RealMe、三星、Vivo、小米等在内的领先制造商和智能手机品牌准备在未来几周内推出搭载 Snapdragon 8 Elite 的设备。该 CPU 采用高通的第二代定制高通 Oryon CPU，旨在无缝处理多模态 AI 的复杂性，同时优先考虑隐私。高通公司产品营销高级总监 Cisco Chen 发言总结如下：

- 1) Hexagon NPU 的性能是上一代的 12 倍。Oryon CPU 的性能是其三倍。Adreno GPU 的性能是其三倍。器件在片上系统设计中嵌入了 40 多个组件，客户可以在最终设计中混合和匹配他们想要的功能。
- 2) Oryon 的 Prime 内核是一种全新的微架构，效率比上一代高 46%。运行频率为 4.32GHz。能比过去更快地获取数据，使内核能够更快地执行下一条指令。除了两个 Prime 核心之外，还有六个性能核心。
- 3) 该平台首次采用了行业技术，例如最新的 Qualcomm Adreno 图形处理单元 (GPU) 和增强的 Qualcomm Hexagon NPU，所有这些都提供了改变游戏规则的性能改进。
- 4) Qualcomm 还针对应用内体验、多任务处理、生成式 AI、视频渲染和流媒体以及游戏进行了优化。将以更高的帧速率提供更流畅的游戏体验，并将游戏时间延长至 2.5 小时，电池效率提高 40%，GPU 性能提高 40%。它支持 Unreal Engine 的 Chaos Physics 系统，在设备上实时运行游戏物理模拟。
- 5) 新的 Qualcomm AI 引擎将利用 Oryon CPU 执行延迟关键型 AI 任务。Hexagon NPU 在加速器之间具有更高的吞吐量，可实现更快的推理性能。

AI 应用：Anthropic 推出升级版 Claude 3.5 Sonnet 以及新模型 Claude 3.5 Haiku

Anthropic 推出了升级版的 Claude 3.5 Sonnet 以及一款新模型 Claude 3.5 Haiku。其中，升级版 Claude 3.5 Sonnet 的各项能力全面胜过之前版本，其中代码能力提升显著。Claude 3.5 Haiku 的性能则与之前最大模型 Claude 3 Opus 的性能相当，同时在成本和速度上与上一代 Haiku 相近。最新版本的 Claude 3.5 Sonnet 能够根据用户指令移动光标、点击相应位置以及通过虚拟键盘输入信息，模仿人类与计算机的交互方式。

在公测中，Anthropic 引入一项突破性的新功能：计算机使用能力。从

今天起，开发者可以通过 API，指导 Claude 像人类一样使用计算机了。Claude 3.5 Sonnet 是首个在公测中提供此功能的模型。当然，这项功能仍处于实验阶段，使用起来还有些笨拙，可能出错。而 Anthropic 选择提前发布此功能，也是为了获得开发者反馈，将之快速改进。在 OSWorld 测试中，它在仅基于屏幕截图的任务类别中得分为 14.9%，明显超越了排名第二的 AI 系统（7.8%）。当允许更多操作步骤来完成任务时，Claude 得分提高到了 22.0%。这表明模型与环境的多次交互，能够优化任务性能。虽然这一结果比之前有了大幅提升，但仍然远低于人类 72.36% 的表现。这也暗示了，Claude 3.5 Sonnet 未来还有很大的改进空间。

Claude 3.5 Haiku 是 Anthropic 最快的模型的下一代。与 Claude 3 Haiku 的成本相同，速度相似，Claude 3.5 Haiku 在每项技能上都有改进，并且在许多智能基准测试上甚至超过了 Anthropic 上一代最大的模型——Claude 3 Opus。Claude 3.5 Haiku 在编码任务上尤其强大。例如，它在 SWE-bench Verified 上的得分为 40.6%，超过了使用公开可用的 SOTA 模型的许多智能体——包括原始的 Claude 3.5 Sonnet 和 GPT-4o。凭借低延迟、改进的指令遵循和更准确的工具使用，Claude 3.5 Haiku 非常适合面向用户的产品、专门的 sub-agent 任务以及从大量数据（如购买历史、定价或库存记录）中生成个性化体验。

Claude 3.5 Haiku 将于本月早些时候通过 Anthropic 的第一方 API、Amazon Bedrock，和谷歌云的 Vertex AI 提供——最初作为仅限文本的模型，随后将支持图像输入。

■ AI 融资动向：Lightmatter/驯鹿 AI 分别获 4 亿美元/6000 万元人民币融资

本期 AI 初创公司的融资中，Lightmatter/驯鹿 AI 分别获 4 亿美元/6000 万元人民币融资。Lightmatter 已经将芯片互连速度提高到 30Tbps，100Tbps 的版本也在路上。它有光子计算单元，有光学芯片封装和传输技术，能够系统化提升整个 AI 计算集群的计算力，计算效率，并降低功耗。Lightmatter 声称其产品有能力支持万亿级参数的大语言模型训练，并有望帮助 AI 实验室们探索通用人工智能。驯鹿 AI 采用多模型协作技术，根据任务类型和难度智能选择最佳模型。此技术确保任务的精准与高效完成。在功能方面，驯鹿 AI 的客服系统不仅覆盖了售前、售中、售后的全流程，还能够主动进行客户开发，通过对企业业务流程的深入分析，定制专属的 AI 模型。在简单易用的基础上，驯鹿 AI 能够在一个月内完成机器人的部署。

■ 投资建议

人工智能需求促使高速光模块需求量剧增，光芯片供应小于需求，目前缺口较大。在此情况下，涨价潮、投资潮一齐涌现。近期，美国网通及光通信芯片大厂 Marvell 宣布全产品线将于 2025 年 1 月 1 日起涨价；广东、无锡、英伟达、中际旭创等有关部门、国资、大厂等各方先后宣布投资或支持光芯片发展。我们认为 AI 需求高涨，继续看好 AI 软硬件机会。

建议关注以 AI 为核心的龙头厂商科大讯飞（002230.SZ）、有望迎来需求爆发的 AI 应用金桥信息（603918.SH）、高速通信连接器业务或显著

受益于 GB200 放量的鼎通科技（688668.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。

公司代码	名称	2024-10-27 股价	EPS			PE			投资评级
			2023	2024E	2025E	2023	2024E	2025E	
002230.SZ	科大讯飞	45.48	0.28	0.40	0.56	162.43	113.70	81.21	买入
002368.SZ	太极股份	25.39	0.79	1.01	1.28	32.14	25.14	19.84	买入
688256.SH	寒武纪-U	421.89	-2.04	-1.21	-0.50	-206.81	-348.67	-843.78	买入
688668.SH	鼎通科技	43.64	0.67	1.04	1.41	65.13	41.96	30.95	买入

资料来源：Wind，华鑫证券研究

正文目录

1、 算力动态：算力租赁价格平稳，高通推出 SNAPDRAGON 8 ELITE	5
1.1、 数据跟踪：算力租赁价格平稳	5
1.2、 产业动态：高通推出 Snapdragon 8 Elite，处理速度显著提升.....	5
2、 AI 应用动态：KIMI 搜索访问量环比+32.36%，ANTHROPIC 推出升级版 CLAUDE 3.5 SONNET	7
2.1、 流量跟踪：Perplexity 搜索访问量环比+7.23%.....	7
2.2、 产业动态：Anthropic 推出升级版 Claude 3.5 Sonnet 以及新模型 Claude 3.5 Haiku.....	8
3、 AI 融资动向：LIGHTMATTER/驯鹿 AI 分别获 4 亿美元/6000 万元人民币融资	12
4、 行情复盘.....	12
5、 投资建议.....	14
6、 风险提示.....	15

图表目录

图表 1：本周算力租赁情况	5
图表 2：高通 Snapdragon Elite 8 CPU	6
图表 3：2024.10.21-2024.10.25 AI 相关网站流量.....	7
图表 4：Claude 3.5 Sonnet 电脑使用能力.....	8
图表 5：Claude 3.5 Sonnet 编码能力.....	9
图表 6：TAU-bench 测试表现.....	9
图表 7：推理测试基准 GPQA 测试表现	9
图表 8：视觉 QA、数学推理、文档视觉问答、图表问答、科学表格基准测试表现.....	10
图表 9：本周 AI 初创公司的融资动态	12
图表 10：本周指数日涨跌幅	12
图表 11：本周 AI 算力指数内部涨跌幅度排名	13
图表 12：本周 AI 应用指数内部涨跌幅度排名	13
图表 13：重点关注公司及盈利预测	15

1、算力动态：算力租赁价格平稳，高通推出 Snapdragon 8 Elite

1.1、数据跟踪：算力租赁价格平稳

本周恒源云 13 核+128G 价格环比上周上涨 5.88%。具体来看，显卡配置为 A100-40G 中，腾讯云 16 核+96G 价格为 28.64 元/时，阿里云 12 核+94GiB 价格为 31.58 元/时；显卡配置为 A100-80G 中，恒源云 13 核+128G 价格为 9.00 元/时；阿里云 16 核+125GiB 价格为 34.74 元/时；显卡配置为 A800-80G 中，恒源云 16+256G 价格为 9.00 元/时。

图表 1：本周算力租赁情况

显卡配置	CPU	内存	磁盘大小 (G)	平台名称	价格 (每小时)	价格环比上周
A100-40G	16	96	可自定, 额外收费	腾讯云	28.64/元	0.00%
	12 核	94G	可自定, 额外收费	阿里云	31.58/元	0.00%
A100-80G	13	128	系统盘: 20G 数据盘: 50GB	恒源云	9.00/元	5.88%
	16 核	125G	可自定, 额外收费	阿里云	34.74/元	0.00%
A800-80G	16	256	系统盘: 20G 数据盘: 50GB	恒源云	9.00/元	0.00%

资料来源：腾讯云，阿里云，恒源云，华鑫证券研究

1.2、产业动态：高通推出 Snapdragon 8 Elite，处理速度显著提升

高通本周在毛伊岛举行的年度 Snapdragon 峰会上表示，包括华硕、荣耀、iQOO、OnePlus、OPPO、RealMe、三星、Vivo、小米等在内的领先制造商和智能手机品牌准备在未来几周内推出搭载 Snapdragon 8 Elite 的设备。该 CPU 采用高通的第二代定制高通 Oryon CPU，旨在无缝处理多模态 AI 的复杂性，同时优先考虑隐私。高通公司产品营销高级总监 Cisco Chen 发言总结如下：

- 1) Hexagon NPU 的性能是上一代的 12 倍。Oryon CPU 的性能是其三倍。Adreno GPU 的性能是其三倍。器件在片上系统设计中嵌入了 40 多个组件，客户可以在最终设计中混合和匹配他们想要的功能。
- 2) Oryon 的 Prime 内核是一种全新的微架构，效率比上一代高 46%。运行频率为 4.32GHz。能比过去更快地获取数据，使内核能够更快地执行下一条指令。除了两个 Prime 核心之外，还有六个性能核心。

- 3) 该平台首次采用了行业技术，例如最新的 Qualcomm Adreno 图形处理单元（GPU）和增强的 Qualcomm Hexagon NPU，所有这些都提供了改变游戏规则的性能改进。
- 4) Qualcomm 还针对应用内体验、多任务处理、生成式 AI、视频渲染和流媒体以及游戏进行了优化。将以更高的帧速率提供更流畅的游戏体验，并将游戏时间延长至 2.5 小时，电池效率提高 40%，GPU 性能提高 40%。它支持 Unreal Engine 的 Chaos Physics 系统，在设备上实时运行游戏物理模拟。
- 5) 新的 Qualcomm AI 引擎将利用 Oryon CPU 执行延迟关键型 AI 任务。Hexagon NPU 在加速器之间具有更高的吞吐量，可实现更快的推理性能。

图表 2: 高通 Snapdragon Elite 8 CPU



资料来源：硅基 LIFE、华鑫证券研究

2、AI 应用动态：Kimi 搜索访问量环比 +32.36%，Anthropic 推出升级版 Claude 3.5 Sonnet

2.1、流量跟踪：Perplexity 搜索访问量环比+7.23%

本期（2024.10.21-2024.10.25）AI 相关网站流量数据：访问量前三位分别为 ChatGPT（844.5M）、Bing（404.1M）和 Canva（176.6M）；访问量环比增速前三位分别为 Perplexity（7.23%）、ChatGPT（3.49%）和 Character.AI（2.37%）；平均停留时长前三位分别为 Discord（00:12:26）、Character.AI（00:12:16）和 DeepL（00:08:58）；平均停留时长环比增速前三位分别为 Kimi（1.61%）、Character.AI（0.68%）和 QuillBot（0.40%）。

图表 3：2024.10.21-2024.10.25 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	844.5	3.49%	6:11	0.00%
Bing	搜索	微软	404.1	-1.05%	6:52	-0.48%
Discord	游戏社区	微软	124.5	-1.35%	12:26	0.00%
Canva	在线设计	Canva	176.6	1.03%	7:58	-0.62%
Github	代码托管	微软	115.8	-2.03%	6:40	0.25%
Gemini	聊天机器人	谷歌	66.99	0.34%	4:49	-2.03%
Character.AI	聊天机器人	Character.AI	42.80	2.37%	12:16	0.68%
NotionAI	文本/笔记	Notion	35.67	-0.78%	8:52	-0.19%
QuillBot	释义工具	QuillBot	18.23	-1.67%	4:08	0.40%
Kimi	聊天机器人	Moonshot AI	6.760	-7.83%	3:09	1.61%
DeepL	翻译工具	DeepL	49.15	0.80%	8:58	-0.19%
文心一言	聊天机器人	百度	4.893	-14.35%	3:02	0.00%
Perplexity	AI 搜索	Perplexity	20.16	7.23%	6:09	0.00%

资料来源：similarweb, 华鑫证券研究

2.2、产业动态：Anthropic 推出升级版 Claude 3.5 Sonnet 以及新模型 Claude 3.5 Haiku

Anthropic 推出了升级版的 Claude 3.5 Sonnet 以及一款新模型 Claude 3.5 Haiku。其中，升级版 Claude 3.5 Sonnet 的各项能力全面胜过之前版本，其中代码能力提升显著。Claude 3.5 Haiku 的性能则与之前最大模型 Claude 3 Opus 的性能相当，同时在成本和速度上与上一代 Haiku 相近。最新版本的 Claude 3.5 Sonnet 能够根据用户指令移动光标、点击相应位置以及通过虚拟键盘输入信息，模仿人类与计算机的交互方式。

在公测中，Anthropic 引入一项突破性的新功能：计算机使用能力。从今天起，开发者可以通过 API，指导 Claude 像人类一样使用计算机了。Claude 3.5 Sonnet 是首个在公测中提供此功能的模型。当然，这项功能仍处于实验阶段，使用起来还有些笨拙，可能出错。而 Anthropic 选择提前发布此功能，也是为了获得开发者反馈，将之快速改进。在 OSWorld 测试中，它在仅基于屏幕截图的任务类别中得分为 14.9%，明显超越了排名第二的 AI 系统（7.8%）。当允许更多操作步骤来完成任务时，Claude 得分提高到了 22.0%。这表明模型与环境的多次交互，能够优化任务性能。虽然这一结果比之前有了大幅提升，但仍然远低于人类 72.36% 的表现。这也暗示了，Claude 3.5 Sonnet 未来还有很大的改进空间。

图表 4：Claude 3.5 Sonnet 电脑使用能力

Category	Claude 3.5 Sonnet (New) - 15 steps		Claude 3.5 Sonnet (New) - 50 steps		Human Success Rate [3]
	Success Rate	95% CI	Success Rate	95% CI	
OS	54.2%	[34.3, 74.1]%	41.7%	[22.0, 61.4]%	75.00%
Office	7.7%	[2.9, 12.5]%	17.9%	[11.0, 24.8]%	71.79%
Daily	16.7%	[8.4, 25.0]%	24.4%	[14.9, 33.9]%	70.51%
Professional	24.5%	[12.5, 36.5]%	40.8%	[27.0, 54.6]%	73.47%
Workflow	7.9%	[2.6, 13.2]%	10.9%	[4.9, 17.0]%	73.27%
Overall	14.9%	[11.3, 18.5]%	22%	[17.8, 26.2]%	72.36%

Table 1 Results of OSWorld[3] tests with screenshot only inputs. For models, we report the average success rate and the 95% Confidence Interval (CI).

资料来源：新智元，华鑫证券研究

其次就是在 Canvas 界面中，如何决定让 AI 全局重写还是局部更改，OpenAI 目前的策略是，如果想要局部更改，就可以通过类似划词的方式进行；而右下角的所有功能，则是倾向于全局重写。

在各项行业基准测试中，升级版 Claude 3.5 Sonnet 性能得到了全方位提升。特别是，智能体编码、工具使用任务中取得显著突破。

1) 在编码能力方面，它在 SWE-bench Verified 测试中，性能从 33.4% 大幅提升至 49.0%。这超越了所有公开可用的模型——包括 OpenAI o1-preview 等推理模型和专为智能体编码设计的专门系统。

图表 5: Claude 3.5 Sonnet 编码能力

Model	% Resolved	Date	Logs	Trajs	Site
New Claude 3.5 Sonnet	49%				
Gru(2024-08-24)	45.20	2024-08-24	🔗	🔗	🔗
Honeycomb	40.60	2024-08-20	🔗	🔗	🔗
Amazon Q Developer Agent (v20240719-dev)	38.80	2024-07-21	🔗	🔗	🔗
AutoCodeRover (v20240620) + GPT 4o (2024-05-13)	38.40	2024-06-28	🔗	-	🔗
Factory Code Droid	37.00	2024-06-17	🔗	-	🔗
👍 SWE-agent + Claude 3.5 Sonnet	33.60	2024-06-20	🔗	🔗	-
👍 AppMap Navie + GPT 4o (2024-05-13)	26.20	2024-06-15	🔗	-	🔗
Amazon Q Developer Agent (v20240430-dev)	25.60	2024-05-09	🔗	-	🔗
EPAM AI/Run Developer Agent + GPT4o	24.00	2024-08-20	🔗	🔗	🔗
👍 SWE-agent + GPT 4o (2024-05-13)	23.20	2024-07-28	🔗	🔗	🔗
👍 SWE-agent + GPT 4 (1106)	22.40	2024-04-02	🔗	🔗	🔗
👍 SWE-agent + Claude 3 Opus	18.20	2024-04-02	🔗	🔗	-
👍 RAG + Claude 3 Opus	7.00	2024-04-02	🔗	-	🔗
👍 RAG + Claude 2	4.40	2023-10-10	🔗	-	-
👍 RAG + GPT 4 (1106)	2.80	2024-04-02	🔗	-	-
👍 RAG + SWE-Llama 7B	1.40	2023-10-10	🔗	-	-
👍 RAG + SWE-Llama 13B	1.20	2023-10-10	🔗	-	-
👍 RAG + ChatGPT 3.5	0.40	2023-10-10	🔗	-	-

资料来源: 新智元, 华鑫证券研究

2) 在 TAU-bench (一项评估智能体工具使用能力的基准测试) 中, Claude 3.5 Sonnet 也表现出色: 在零售领域的得分从 62.6% 提高到 69.2%, 在更具挑战性的航空领域则从 36.0% 跃升至 46.0%。

图表 6: TAU-bench 测试表现

	Claude 3.5 Sonnet (New)	Claude 3.5 Haiku	Claude 3.5 Sonnet	Claude 3 Opus	Claude 3 Haiku
Retail	69.2%	51.0%	62.6%	45.1%	18.2%
Airline	46.0%	22.8%	36.0%	34.5%	16.0%

Table 3 pass^1 TAU-bench results 公众号 · 新智元

资料来源: 新智元, 华鑫证券研究

3) 推理测试基准 GPQA (Diamond) 上, 新版 Claude 3.5 Sonnet 大幅超越 GPT-4o。

图表 7: 推理测试基准 GPQA 测试表现

Claude 3.5 Sonnet (new) benchmarks

	Claude 3.5 Sonnet (new)	Claude 3.5 Sonnet	GPT-4o	Gemini 1.5 Pro
Graduate level reasoning <i>GPQA (Diamond)</i>	65.0% 0-shot CoT	59.4% 0-shot CoT	53.6% 0-shot CoT	59.1% 0-shot CoT
Undergraduate level knowledge <i>MMLU Pro</i>	78.0% 0-shot CoT	75.1% 0-shot CoT	—	75.8% 0-shot CoT
Agentic coding <i>SWE-bench Verified</i>	49.0%	33.4%	—	—
Code <i>HumanEval</i>	93.7% 0-shot	92.0% 0-shot	90.2% 0-shot	—
Math problem-solving <i>MATH</i>	78.3% 0-shot CoT	71.1% 0-shot CoT	76.6% 0-shot CoT	86.5% 4-shot CoT
Multilingual math <i>MGSM</i>	92.5% 0-shot CoT	91.6% 0-shot CoT	90.5% 0-shot CoT	—
Reasoning over text <i>DROP, F1 Score</i>	88.3 3-shot	87.1 3-shot	83.4 3-shot	—
Agentic tool use <i>TAU-bench</i>	Retail 69.2% Airline 46.0%	Retail 62.6% Airline 36.0%	—	—

公众号 · 新智元

资料来源：新智元，华鑫证券研究

4) 在视觉 QA、数学推理、文档视觉问答、图表问答、科学表格基准测试中，Claude 3.5 Sonnet 性能成为业界新标杆。

图表 8：视觉 QA、数学推理、文档视觉问答、图表问答、科学表格基准测试表现

	Claude 3.5 Sonnet (new)	Claude 3.5 Sonnet	GPT-4o	Gemini 1.5 Pro
Visual question answering <i>MMMU (val)</i>	70.4% 0-shot CoT	68.3% 0-shot CoT	69.1% 0-shot CoT	65.9% 0-shot CoT
Visual math reasoning <i>MathVista (testmini)</i>	70.7% 0-shot CoT	67.7% 0-shot CoT	63.8% 0-shot CoT	68.1% 0-shot CoT
Science diagrams <i>A12D, test</i>	95.3% 0-shot	94.7% 0-shot	94.2% 0-shot	—
Chart Q&A <i>Relaxed accuracy (test)</i>	90.8% 0-shot CoT	90.8% 0-shot CoT	85.7% 0-shot CoT	—
Document visual Q&A <i>ANLS score, test</i>	94.2% 0-shot	95.2% 0-shot	92.8% 0-shot	—

公众号 · 新智元

资料来源：新智元，华鑫证券研究

值得一提的是，新版 Claude 3.5 Sonnet 性能突破同时，仍保持了与前代模型相同的价格和运行速度。一些早期测试用户的反馈，进一步印证了升级后 Claude 3.5 Sonnet，在

AI 驱动编码领域实现“质”的飞跃。

1) GitLab: 在 DevSecOps 任务测试中，发现 Claude 3.5 Sonnet 在不增加延迟的前提下，推理能力显著提升（各用例最高提升 10%），使其成为驱动复杂软件开发流程的理想选择

2) Cognition: 将新版 Claude 3.5 Sonnet 应用于自主 AI 评估，在编码、规划和问题解决等方面，相较前代模型均取得了实质性进步

3) The Browser Company: 在使用该模型自动化网络工作流程时发现，Claude 3.5 Sonnet 的表现超越了他们此前测试过的所有模型。

Claude 3.5 Haiku 是 Anthropic 最快的模型的下一代。与 Claude 3 Haiku 的成本相同，速度相似，Claude 3.5 Haiku 在每项技能上都有改进，并且在许多智能基准测试上甚至超过了 Anthropic 上一代最大的模型——Claude 3 Opus。Claude 3.5 Haiku 在编码任务上尤其强大。例如，它在 SWE-bench Verified 上的得分为 40.6%，超过了使用公开可用的 SOTA 模型的许多智能体——包括原始的 Claude 3.5 Sonnet 和 GPT-4o。凭借低延迟、改进的指令遵循和更准确的工具使用，Claude 3.5 Haiku 非常适合面向用户的产品、专门的 sub-agent 任务以及从大量数据（如购买历史、定价或库存记录）中生成个性化体验。

Claude 3.5 Haiku 将于本月晚些时候通过 Anthropic 的第一方 API、Amazon Bedrock，和谷歌云的 Vertex AI 提供——最初作为仅限文本的模型，随后将支持图像输入。

3、AI 融资动向：Lightmatter/驯鹿 AI 分别获 4 亿美元/6000 万元人民币融资

本期 AI 初创公司的融资中，Lightmatter/驯鹿 AI 分别获 4 亿美元/6000 万元人民币融资。Lightmatter 已经将芯片互连速度提高到 30Tbps，100Tbps 的版本也在路上。它有光子计算单元，有光学芯片封装和传输技术，能够系统化提升整个 AI 计算集群的计算力，计算效率，并降低功耗。Lightmatter 声称其产品有能力支持万亿级参数的大语言模型训练，并有望帮助 AI 实验室们探索通用人工智能。驯鹿 AI 采用多模型协作技术，根据任务类型和难度智能选择最佳模型。此技术确保任务的精准与高效完成。在功能方面，驯鹿 AI 的客服系统不仅覆盖了售前、售中、售后的全流程，还能够主动进行客户开发，通过对企业业务流程的深入分析，定制专属的 AI 模型。在简单易用的基础上，驯鹿 AI 能够在一个月内完成机器人的部署。

图表 9：本周 AI 初创公司的融资动态

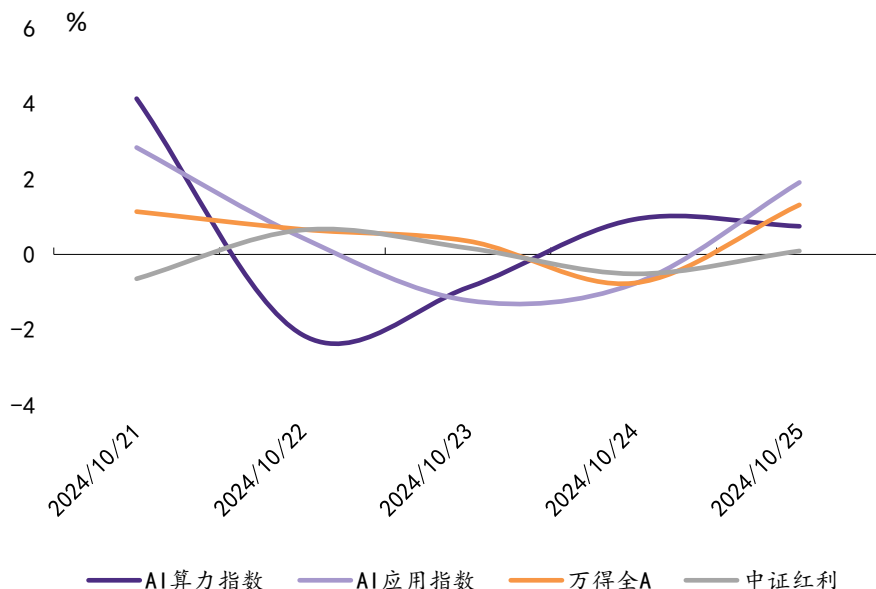
应用	应用类型	领投方	融资轮	融资额	目前累计融资额	目前估值
Lightmatter	光子计算	T. Rowe Price Associates Inc.	D 轮	4 亿美元	8.5 亿美元	44 亿美元
驯鹿 AI	大语言模型	IDG、盛大资本	战略融资	6000 万元	——	——

资料来源：阿尔法公社，搜狐网，华鑫证券研究

4、行情复盘

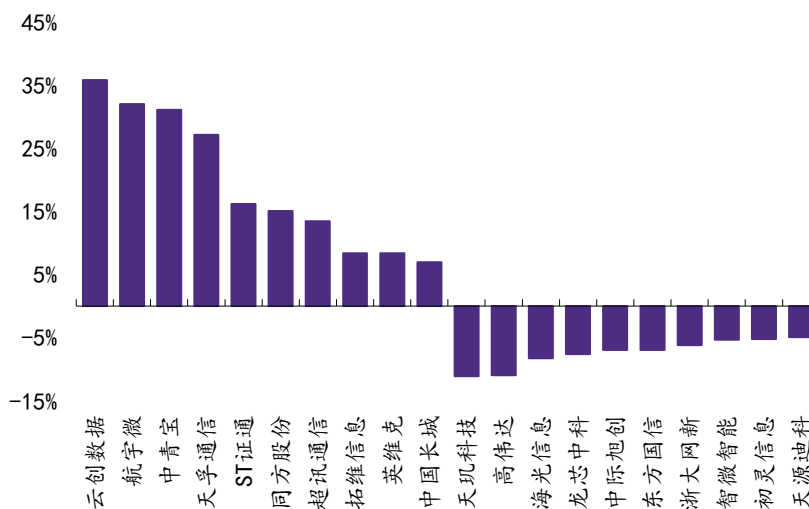
本周，AI 算力指数/AI 应用指数/万得全 A/中证红利日涨幅最大值分别为 4.13%/2.84%/1.31%/0.65%，日跌幅最大值分别为-2.13%/-1.22%/-0.76%/-0.65%。AI 算力指数内部，云创数据以+35.77%录得本周最大涨幅，天玑科技以-11.11%录得本周最大跌幅。AI 应用指数内部，鸿博股份以+38.03%录得本周最大涨幅，苏州科达以-8.50%录得本周最大跌幅。

图表 10：本周指数日涨跌幅



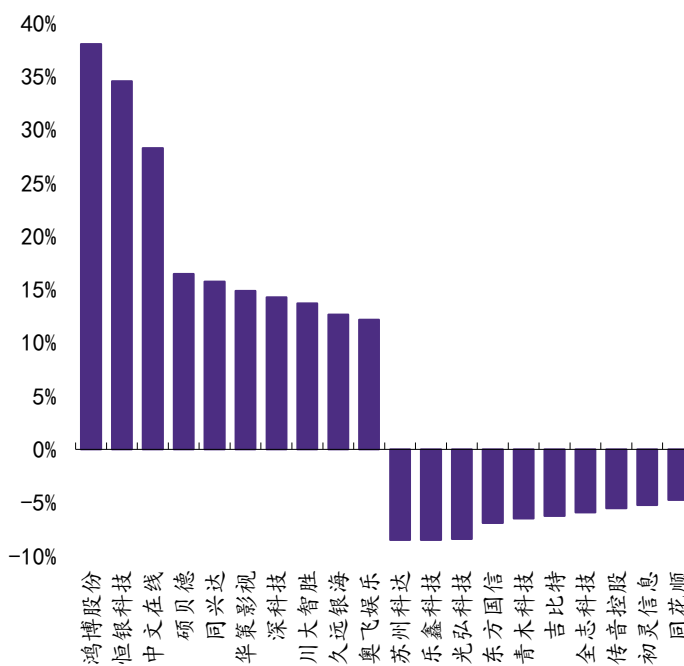
资料来源: wind, 华鑫证券研究

图表 11: 本周 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: 本周 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

AI 需求高涨近期持续被验证。人工智能需求促使高速光模块需求量剧增，光芯片供应小于需求，目前缺口较大。在此情况下，涨价潮、投资潮一齐涌现。近期，美国网通及光通信芯片大厂 Marvell 宣布全产品线将于 2025 年 1 月 1 日起涨价；广东、无锡、英伟达、中际旭创等有关部门、国资、大厂等各方先后宣布投资或支持光芯片发展。我们认为 AI 需求高涨，继续看好 AI 软硬件机会。

建议关注以 AI 为核心的龙头厂商科大讯飞（002230.SZ）、有望迎来需求爆发的 AI 应用金桥信息（603918.SH）、高速通信连接器业务或显著受益于 GB200 放量的鼎通科技（688668.SH）。

图表 13: 重点关注公司及盈利预测

公司代码	名称	2024-10-27 股价	EPS			PE			投资评级
			2023	2024E	2025E	2023	2024E	2025E	
002230.SZ	科大讯飞	45.48	0.28	0.40	0.56	162.43	113.70	81.21	买入
002368.SZ	太极股份	25.39	0.79	1.01	1.28	32.14	25.14	19.84	买入
688256.SH	寒武纪-U	421.89	-2.04	-1.21	-0.50	-206.81	-348.67	-843.78	买入
688668.SH	鼎通科技	43.64	0.67	1.04	1.41	65.13	41.96	30.95	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。

■ 计算机&中小盘组介绍

宝幼琛：本硕毕业于上海交通大学，多次新财富、水晶球最佳分析师团队成员，7年证券从业经验，2021年11月加盟华鑫证券研究所，目前主要负责计算机与中小盘行业上市公司研究。擅长领域包括：云计算、网络安全、人工智能、区块链等。

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

陶欣怡：毕业于上海交通大学，于2023年10月加入团队。

■ 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。

■ 免责声明

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。