

2024年09月01日



华鑫证券
CHINA FORTUNE SECURITIES

AMD RDNA4 GPU 架构显卡曝光，智谱发布基座大模型 GLM-4-Plus

—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)	-6.5	-13.4	-34.8
沪深300	-2.9	-7.2	-12.4

市场表现



资料来源：Wind，华鑫证券研究

相关研究

- 《计算机行业周报：Akeana 推出新的 RISC-V 芯片设计，xAI 发布 Grok-2 大模型》2024-08-18
- 《计算机行业深度报告：政策东风至，车路云发展正当时》2024-08-13
- 《计算机行业周报：Nvidia 推迟芯片出货但预计影响有限，GPT-4o 新版本上线》2024-08-11

算力：算力租赁价格平稳，AMD RDNA4 GPU 架构显卡曝光

AMD 计划在 2025 年 1 月正式发布 RDNA4GPU 架构显卡。根据 Geekbench 数据库的信息，这款显卡的设备标识符为 GFX1201，预计属于 RDNA4GPU 架构。目前曝光的显卡拥有 28 个计算单元，但实际应为 56 个 CUs，配备了 3,584 个流处理器。

显卡的时钟频率为 2,101MHz，目前只是工程样品的频率，最终产品的频率可能会有所不同。此外，显卡配备了 16GB 的 GDDR6 内存，与 RDNA4GPU 架构的预期相符。

AI 应用：智谱发布基座大模型 GLM-4-Plus，性能全面提升

在 8 月 29 日的 KDD 国际数据挖掘与知识发现大会上，智谱正式发布了新一代基座大模型 GLM-4-Plus。该模型在语言理解、指令遵循、长文本处理等方面性能得到全面提升。GLM-4-Plus 通过引入大量模型辅助构造高质量合成数据，显著提升了模型的整体性能，同时，采用 PPO 有效增强了模型在解决数学、编程算法等推理任务上的表现，使其更贴近人类的偏好和决策。

此外，智谱还发布了文生图模型 CogView-3-Plus 和图像/视频理解模型 GLM-4V-Plus。前者具备与当前最优的 MJ-V6 和 FLUX 等模型接近的性能，后者具备卓越的图像理解能力和基于时间感知的视频理解能力。与 GLM-4-Plus 共同构成了其自主原创的全栈大模型谱系。

GLM-4-Plus 现已在智谱大模型开放平台部署，企业和开发者即日起可以通过智谱开放平台 bigmodel.cn 上的 API 调用智谱最新的基座大模型。最新推出的基座大模型，和此前发布的 CogVideoX 等模型一道，完善了智谱自主原创的全栈大模型谱系，推动智谱实现面向世界先进水平的全面对标。

AI 融资动向：Magic/Codeium 分别获 3.2 亿/1.5 亿美元融资

本期 AI 初创公司的融资中，Magic/Codeium 融资额前二，分别为 3.2 亿/1.5 亿美元。Codeium 平台主要依托于在公共代码库中训练的生成式 AI 模型而运行。该平台支持约 70 种编程语言，并集成了许多流行的开发环境，其中包括 Microsoft Visual Studio 和 IDE。Magic 最新的模型 LTM-2-mini 能够处理一亿个 token，相当于大约 1000 万行代码，或者 750 本小说。官方表示，它在解码 token 时的计算效率是传统模型的 1000 倍。相较于当前市场上的主流模型，如 Llama3.1，LTM-2-mini 的内存需求小得多，展示了强大的性能优势。

投资建议

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	2024-09-01 股价	EPS			PE			投资评级
			2023	2024E	2025E	2023	2024E	2025E	
002230.SZ	科大讯飞	34.39	0.28	0.40	0.56	122.82	85.98	61.41	买入
002368.SZ	太极股份	14.86	0.79	1.01	1.28	18.81	14.71	11.61	买入
603918.SH	金桥信息	8.93	0.33	0.49	0.80	27.06	18.22	11.16	买入
688668.SH	鼎通科技	38.82	0.67	1.04	1.41	57.94	37.33	27.53	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，AMD RDNA4 GPU 架构显卡曝光.....	4
1.1、 数据跟踪：算力租赁价格平稳.....	4
1.2、 产业动态：AMD RDNA4 GPU 架构显卡曝光，配备 56 个计算单元和 16GBGDDR6 内存.....	4
2、 AI 应用动态：文心一言访问时长环比增长 23.74%，智谱发布基座大模型 GLM-4-PLUS.....	6
2.1、 流量跟踪：文心一言访问时长环比增长 23.74%.....	6
2.2、 产业动态：智谱发布基座大模型 GLM-4-Plus，性能全面提升.....	6
3、 AI 融资动向：CODE IUM/MAGIC 分别获 12.5 亿/3.2 亿美元融资.....	9
4、 行情复盘.....	10
5、 投资建议.....	12
6、 风险提示.....	12

图表目录

图表 1：本周算力租赁情况.....	4
图表 2：AMD RDNA 4GPU 配置.....	4
图表 3：2024.8.14-2024.8.20AI 相关网站流量.....	6
图表 4：GLM-4-Plus 与其他模型逻辑推理能力比较.....	7
图表 5：GLM-4-Plus 与其他模型语言文本处理能力比较.....	8
图表 6：本周 AI 初创公司的融资动态.....	9
图表 7：本周指数日涨跌幅.....	10
图表 8：本周 AI 算力指数内部涨跌幅度排名.....	10
图表 9：本周 AI 应用指数内部涨跌幅度排名.....	11
图表 10：重点关注公司及盈利预测.....	12

1、算力动态：算力租赁价格平稳，AMD RDNA4 GPU 架构显卡曝光

1.1、数据跟踪：算力租赁价格平稳

本周算力租赁价格环比持平。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 8.50 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 13 核+128G 价格为 8.50 元/时；恒源云 16 核+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	8.50/元	-
	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	-

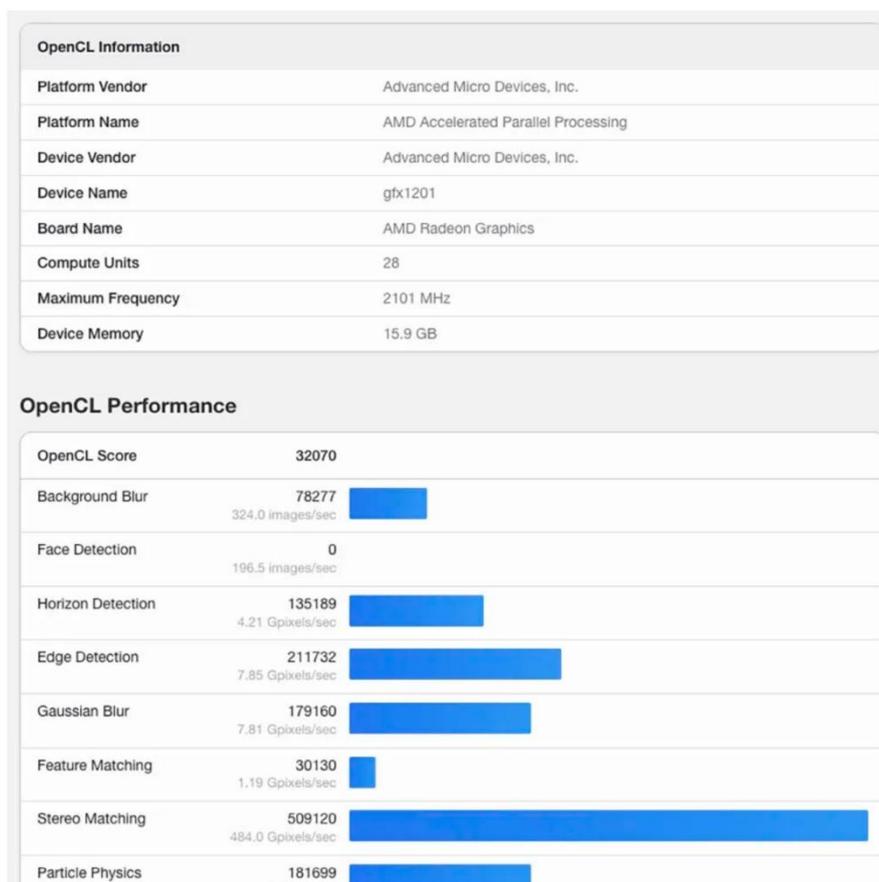
资料来源：腾讯云，阿里云，恒源云，华鑫证券研究

1.2、产业动态：AMD RDNA4 GPU 架构显卡曝光，配备 56 个计算单元和 16GBGDDR6 内存

AMD 计划在 2025 年 1 月正式发布 RDNA 4GPU 架构显卡根据。根据 Geekbench 数据库的信息，这款显卡的设备标识符为 GFX1201，预计属于 RDNA 4GPU 架构。目前曝光的显卡拥有 28 个计算单元，但实际应为 56 个 CUs，配备了 3,584 个流处理器。

显卡的时钟频率为 2,101MHz，目前只是工程样品的频率，最终产品的频率可能会有所不同。此外，显卡配备了 16GB 的 GDDR6 内存，与 RDNA4 GPU 架构的预期相符。

图表 2：AMD RDNA 4GPU 配置



资料来源：硅基 LIFE，华鑫证券研究

2、 AI 应用动态：文心一言访问时长环比增长 23.74%，智谱发布基座大模型 GLM-4-Plus

2.1、 流量跟踪：文心一言访问时长环比增长 23.74%

本期（2024.8.14-2024.8.20）AI 相关网站流量数据：访问量前三位分别为 ChatGPT（633.8M）、Bing（411.7M）和 Canva（163.0M）；访问量环比增速前三位分别为 Canva（4.82%）、ChatGPT（4.48%）和文心一言（4.36%）；平均停留时长前三位分别为 Discord（00:12:47）、Character.AI（00:12:33）和 NotionAI（00:09:08）；平均停留时长环比增速前三位分别为文心一言（23.74%）、QuillBot（0.84%）和 NotionAI（0.37%）。

图表 3：2024.8.14-2024.8.20AI 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	633.8	4.48%	6:02	-0.55%
Bing	搜索	微软	411.7	0.81%	6:50	-0.73%
Discord	游戏社区	微软	130.6	0.15%	12:47	-0.39%
Canva	在线设计	Canva	163.0	4.82%	8:18	0.00%
Github	代码托管	微软	112.2	-1.06%	6:45	-1.22%
Gemini	聊天机器人	谷歌	64.3	5.53%	5:00	-4.46%
Character.AI	聊天机器人	Character.AI	44.54	-1.68%	12:33	0.14%
NotionAI	文本/笔记	Notion	36.4	2.45%	9:08	0.37%
QuillBot	释义工具	QuillBot	15.15	1.75%	4:01	0.84%
Kimi	聊天机器人	MoonshotAI	5.268	-3.57%	2:56	-2.22%
DeepL	翻译工具	DeepL	46.36	2.32%	8:56	1.13%
文心一言	聊天机器人	百度	4.641	4.36%	2:52	23.74%
Perplexity	AI 搜索	Perplexity	13.91	4.19%	6:29	-3.95%

资料来源：similarweb, 华鑫证券研究

2.2、 产业动态：智谱发布基座大模型 GLM-4-Plus，性能全面提升

在 8 月 29 日的 KDD 国际数据挖掘与知识发现大会上，智谱正式发布了新一代基座大模型 GLM-4-Plus。该模型在语言理解、指令遵循、长文本处理等方面性能得到全面提升。在

各项指标上，与 GPT-4o 等第一梯队模型持平。

GLM-4-Plus 通过引入大量模型辅助构造高质量合成数据，显著提升了模型的整体性能，同时，采用 PPO 有效增强了模型在解决数学、编程算法等推理任务上的表现，使其更贴近人类的偏好和决策。

此外，智谱还发布了文生图模型 CogView-3-Plus 和图像/视频理解模型 GLM-4V-Plus。前者具备与当前最优的 MJ-V6 和 FLUX 等模型接近的性能，后者具备卓越的图像理解能力和基于时间感知的视频理解能力。与 GLM-4-Plus 共同构成了其自主原创的全栈大模型谱系。

其中部分 GLM-4-Plus 功能及性能测评如下：

- 1) **通用知识掌握程度、逻辑推理能力比较：**GLM-4-Plus 在大多数任务上能与 GPT-4o 打成平手，整体优于 Gemini1.5pro 和 Claude3.5Sonnet；

图表 4：GLM-4-Plus 与其他模型逻辑推理能力比较

		GLM-4-Plus (-/GPT-4o)	GPT-4o	Claude 3.5 Sonnet	Gemini 1.5 Pro
Language	AlignBench	83.2 (99%)	83.8	80.7	74.7
	MMLU	86.8 (98%)	88.7	88.3	85.9
	MATH	74.2 (97%)	76.6	71.1	67.7
	GPQA	50.7 (99%)	51.0*	56.4*	46.2
	LiveCodeBench	45.8* (101%)	45.5*	49.8	33.6
	NaturalCodeBench	50.4 (96%)	52.3	53.1	42.3
	IFEval	79.5 (97%)	81.9	80.6	74.4
	LogicGame	25.6 (106%)	24.1	30.4	22.5
Long Context	LongBench-Chat	8.80 (98%)	9.00	8.60	8.60
	Ruler	93.0 (-)	-	-	95.8
	InfiniteBench/EN.MC	85.1 (103%)	82.5	79.5	80.9
Image Understanding	OCRBench	833 (113%)	736	788	754
	MME	2274.7 (99%)	2310.3	1920	2110.6
	MMBench-V1.1-Avg	82.4 (102%)	80.5	78.5	73.9
	MMStar	62.5 (98%)	63.9	62.2	59.1
	MMVet	69.9 (101%)	69.1	66.0	64.0
	AI2D	83.6 (99%)	84.6	80.2	79.1
	SEEDBench_IMG	77.4 (100%)	77.1	72.2	-
Video Understanding	MVBench	71.2 (149%)	47.8	-	52.6
	LVBench	38.3 (110%)	34.7	-	33.1
	Temporal QA	🟢	✖	-	🟢
	Multi-turn Dialogue	🟢	🟢	-	🟢
	Bilingual En/Zh	🟢	🟢	-	🟢

*represents reproduced results

资料来源：机器之心，华鑫证券研究

- 2) **语言文本处理能力：**GLM-4-Plus 与 GPT4o 及 405B 参数量的 Llama3.1 不相上下，它通过更精准的长短文本数据混合策略，取得了更强的长文本的推理效果。

图表 5: GLM-4-Plus 与其他模型语言文本处理能力比较

	AlignBench	MMLU	MATH	GPQA	LCB	NCB
Llama 405B	60.7	88.6	73.8	51.1 (*50.1)	*39.4	50
GPT-4o	83.8	88.7	76.6	53.6 (*51.0)	*45.5	52.3
Claude 3.5 Sonnet	80.7	88.3	71.1	59.4 (*56.4)	49.8	53.1
Claude 3Opus	79.2	85.7	60.1	50.4	33.9	46.4
Gemini 1.5Pro	74.7	85.9	67.7	46.2	33.6 (August)	42.3
GLM-4-Plus	83.2	86.8	74.2	50.7	*45.8	50.4

资料来源: APPSO, 华鑫证券研究

GLM-4-Plus 完成部署，智谱全栈大模型谱系得到完善。GLM-4-Plus 现已在智谱大模型开放平台部署，企业和开发者即日起可以通过智谱开放平台 bigmodel.cn 上的 API 调用智谱最新的基座大模型。最新推出的基座大模型，和此前发布的 CogVideoX 等模型一道，完善了智谱自主原创的全栈大模型谱系，推动智谱实现面向世界先进水平的全面对标。

3、 AI 融资动向： Magic/Codeium 分别获 3.2 亿/1.5 亿美元融资

本期 AI 初创公司的融资中， Magic/Codeium 融资额前二， 分别为 3.2 亿/1.5 亿美元。 Magic 最新的模型 LTM-2-mini 能够处理一亿个 token， 相当于大约 1000 万行代码， 或者 750 本小说。 官方表示， 它在解码 token 时的计算效率是传统模型的 1000 倍。 相较于当前市场上的主流模型， 如 Llama3.1， LTM-2-mini 的内存需求小得多， 展示了强大的性能优势。 Codeium 平台主要依托于在公共代码库中训练的生成式 AI 模型而运行。 该平台支持约 70 种编程语言， 并集成了许多流行的开发环境， 其中包括 Microsoft Visual Studio 和 IDE。

图表 6： 本周 AI 初创公司的融资动态

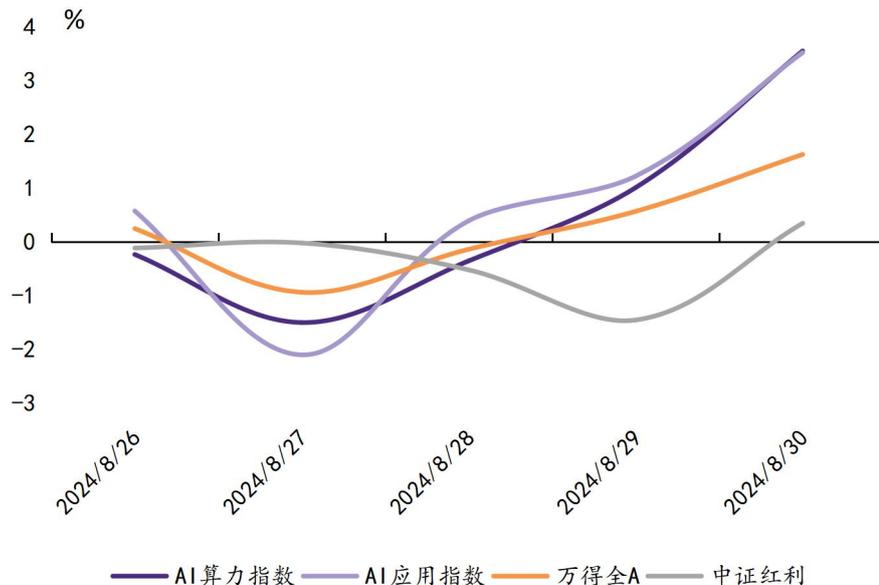
应用	应用类型	领投资方	融资轮	融资额	目前累计融资额	目前估值
Codeium	AI 编程工具	GeneralCatalyst	C 轮	1.5 亿美元	——	——
Magic	AI 编程工具	谷歌 CEO	——	3.2 亿美元	——	——
Anysphere	AI 编程工具	AndreessenHorowitz、ThriveCapital	A 轮	6000 万美元	——	——
向量方程	智能数字人开发	——	天使轮	——	——	——
Dify.AI	LLMOps 平台	CNZZ、风投侠基金	战略投资	——	——	——

资料来源： 36kr， 亿欧智库， 企名片， 凤凰新闻， 站长之家， 极客公园， 华鑫证券研究

4、行情复盘

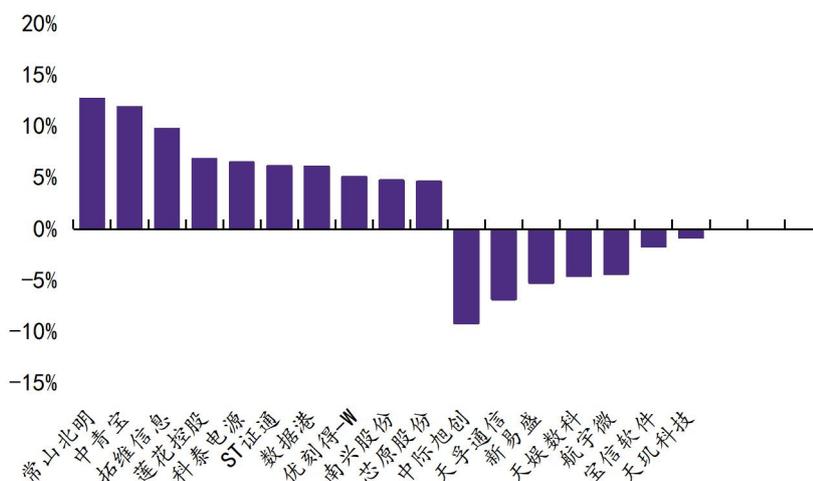
本周，AI 算力指数/AI 应用指数/万得全 A/中证红利日涨幅最大值分别为 3.55%/3.52%/1.63%/0.35%，日跌幅最大值分别为-1.50%/-2.10%/-0.93%/-1.45%。AI 算力指数内部，常山北明以+12.74%录得本周最大涨幅，天玑科技以-0.85%录得本周最大跌幅。AI 应用指数内部，领益智造以+16.71%录得本周最大涨幅，光弘科技以-1.81%录得本周最大跌幅。

图表 7：本周指数日涨跌幅



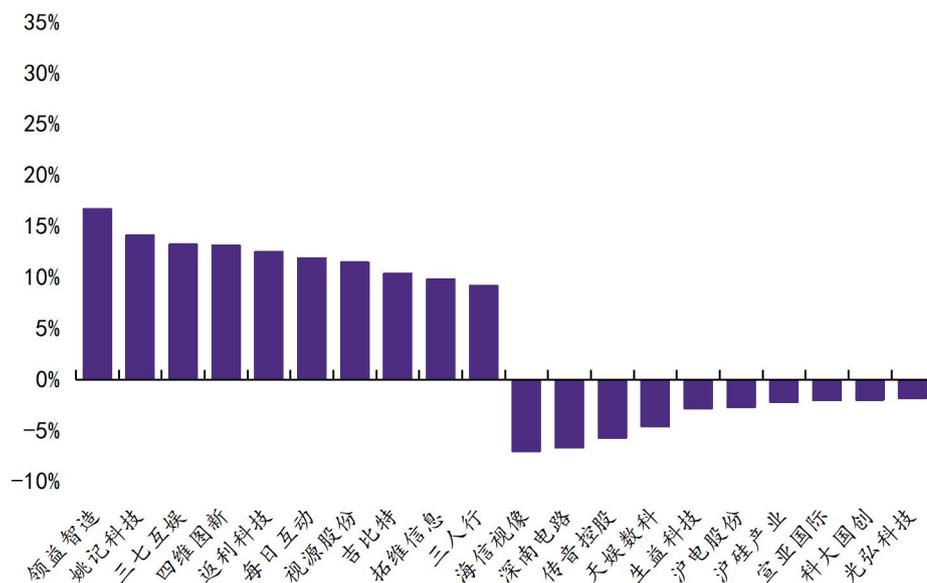
资料来源：wind, 华鑫证券研究

图表 8：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 9：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

英伟达 28 日发布了 2025 财年第二季度（自然年 2024 年二季度）财报，第二季度英伟达营收为 300.4 亿美元，同比增长 122%；净利润 165.99 亿美元，同比增长 168%。在 2025 财年第二财季财报电话会议上，英伟达驳斥了有关 Blackwell 延迟生产的谣言，并表示他们只需对制造流程进行某些更改即可帮助提高芯片的生产良率，同时将保持下个季度产能提升的正常节奏。英伟达表示，Blackwell 需求强劲，预计 Blackwell 第四财季的收入将达到数十亿美元。因此我们仍看好 AI 产业链，预计算力芯片将维持高景气度出货。

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 10：重点关注公司及盈利预测

公司代码	名称	2024-09-01 股价	EPS			PE			投资评级
			2023	2024E	2025E	2023	2024E	2025E	
002230.SZ	科大讯飞	34.39	0.28	0.40	0.56	122.82	85.98	61.41	买入
002368.SZ	太极股份	14.86	0.79	1.01	1.28	18.81	14.71	11.61	买入
603918.SH	金桥信息	8.93	0.33	0.49	0.80	27.06	18.22	11.16	买入
688668.SH	鼎通科技	38.82	0.67	1.04	1.41	57.94	37.33	27.53	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。