

2024年06月02日



华鑫证券
CHINA FORTUNE SECURITIES

Mistral 发布首个 AI 代码模型 Codestral，有望成为开源编程大模型先锋

—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)	-7.0	-13.6	-32.6
沪深300	-0.7	1.2	-7.3

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：微软推出 COPILOT+PC，“主动式 AI”或将开启个人电脑新时代》2024-05-26
- 2、《计算机行业周报：GPT-4o 场景理解力阶跃式提升，“贾维斯时刻”或即将到来》2024-05-20
- 3、《计算机行业周报：月之暗面发布 Kimi+，新一轮 AI 应用军备竞赛进行时》2024-05-13

算力：算力租赁价格平稳，Arm 发布最强核心 Cortex-X925 CPU 和 Immortalis G925 GPU

本周算力租赁价格环比持平。在产业动态方面，5月29日，芯片设计公司 Arm 发布了针对旗舰智能手机的下一代 CPU 和 GPU 设计：Cortex-X925 CPU 和 Immortalis G925 GPU。Cortex-X925 实现了 Cortex-X 系列至今最高的性能提升，AI 算力 TOPS 增加了 50%。在 Geekbench 中，其单线程性能提升了 36%，达到了历史上最大的同比性能提升；在 AI 性能方面，Cortex-X925 基于 Phi-3 模型的首个 token 生成时间与上一代相比缩短了 46%，大型语言模型的初始响应速度实现大幅改善。而 Immortalis-G925 刷新 GPU 性能、效率双纪录。与 Immortalis G720 相比，Immortalis G925 在图形性能方面提升了 37%，相同性能下功耗可降低 30%，使用体验大幅提升。

AI 应用：Mistral 推出首个代码生成 AI 模型 Codestral，有望成为开源编程大模型先锋

5月30日，Mistral 推出了首个代码生成模型 Codestral，支持超过 80 种编程语言，具有编写特定功能的函数、编写测试以及代码填充等功能。Codestral 基于 Llama 架构，其长上下文文本规模达 32K，是 Llama 3-70B 的四倍。目前，Mistral 开放了两个 API 供开发者调用 Codestral，分别是 codestral.mistral.ai 和 api.mistral.ai。Codestral 在 Python、SQL、FIM 等代码测试方面表现优异，有望成为开源编程大模型先锋。

AI 融资动向：马斯克的 xAI 获 60 亿美元 B 轮融资

本期 AI 初创公司的融资中，xAI/智谱 AI 融资额前二，分别为 60 亿/4 亿美元。xAI 专注于回答更深层次的科学问题，期望未来可以用 AI 去帮助人们解决复杂的科学和数学问题并且“理解”宇宙。本次融资将用于推出 xAI 的首个产品、建设先进基础设施，并加速未来技术的研发。智谱 AI 致力于推动人工智能技术的边界，尤其是认知智能领域。它的技术突破主要集中在大规模语言模型的开发上，此外它也构建了高精度的通用知识图谱并自主开发了科技大数据检索平台 Aminer。

投资建议

Arm 发布的针对旗舰智能手机的下一代 CPU 和 GPU 设计，预示着端侧 AI 的加速落地。端侧 AI 作为推动智能化转型的关键驱动力，其在边缘计算、数据隐私保护、实时性工业需求等场景中有出独特的技术优势，AI 的商业化落地应用将被大幅拓展，如安防、工业自动化、智能家

居、医疗健康等多个垂类行业。在深度学习、自然语言处理（NLP）、计算机视觉等核心技术领域具有深厚积累的行业龙头厂商将核心受益。

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）等。

■ 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	2024-06-02 股价	EPS			PE			投资评级
			2023	2024E	2025E	2023	2024E	2025E	
002230.SZ	科大讯飞	42.20	0.28	0.40	0.56	150.71	105.50	75.36	买入
002368.SZ	太极股份	21.35	0.79	1.01	1.28	27.03	21.14	16.68	买入
603918.SH	金桥信息	12.18	0.33	0.49	0.80	36.91	24.86	15.23	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，ARM 发布最强核心 CORTEX-X925 CPU 和 IMMORTALIS G925 GPU.....	4
1.1、 数据跟踪：租赁价格稳定	4
1.1、 产业动态：Arm 发布最强核心 Cortex-X925 CPU 及 Immortalis-G925 GPU，手机性能再上新台阶	4
2、 AI 应用动态：MISTRAL 推出首个代码生成 AI 模型 CODESTRAL，有望成为开源编程大模型先锋.....	7
2.1、 流量跟踪：Kimi 访问量环比+15.41%.....	7
2.2、 产业动态：Mistral 推出首个代码生成 AI 模型 Codestral，支持 80+编程语言	8
3、 AI 融资动向：马斯克的 XAI 获 60 亿美元 B 轮融资	10
4、 行情复盘.....	11
5、 投资建议.....	13
6、 风险提示.....	13

图表目录

图表 1：本周算力租赁情况	4
图表 2：Arm 核心 CPU 图	4
图表 3：Cortex-X925 性能改进	5
图表 4：Armv9 CPU 产品组合的新增和更新.....	5
图表 5：Immortalis-G925 GPU 架构	6
图表 6：Immortalis-G925 整体性能提升.....	6
图表 7：Immortalis-G925 AI 性能改进.....	6
图表 8：2024.5.22-2024.5.28 AI 相关网站流量.....	7
图表 9：使用各项测试评估 Codestral 能力	8
图表 10：使用 Spider 测试评估 Codestral SQL 能力	9
图表 11：系统级搜图.....	9
图表 12：通过自然语言描述寻找此前浏览过的一条裙子	9
图表 13：本周 AI 初创公司的融资动态	10
图表 14：本周指数日涨跌幅	11
图表 15：本周 AI 算力指数内部涨跌幅度排名	12
图表 16：本周 AI 应用指数内部涨跌幅度排名	12
图表 17：重点关注公司及盈利预测	13

1、算力动态：算力租赁价格平稳，Arm 发布最强核心 Cortex-X925 CPU 和 Immortalis G925 GPU

1.1、数据跟踪：租赁价格稳定

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 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 的价格为 9.00 元/时。

图表 1：本周算力租赁情况

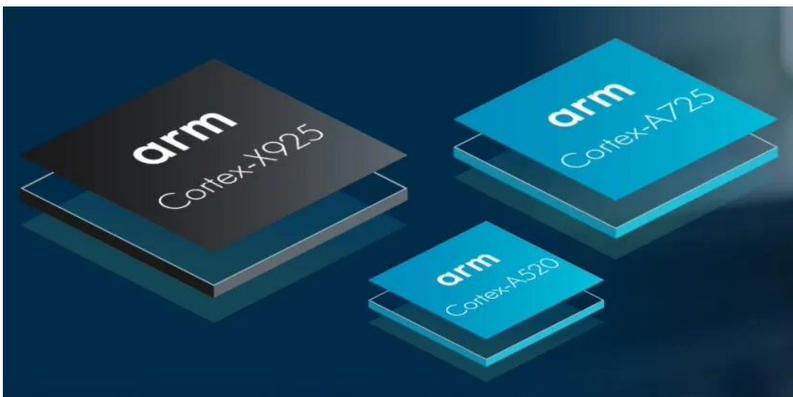
显卡配置	CPU	内存	磁盘大小 (G)	平台名称	价格 (每小时)	价格环比上周
A100-40G	16	96	可自定, 额外收费	腾讯云	28.64/元	0.00%
	12 核	94G	可自定, 额外收费	阿里云	31.58/元	0.00%
A100-80G	13	128	系统盘: 20G 数据盘: 50GB	恒源云	9.00/元	0.00%
	16 核	125G	可自定, 额外收费	阿里云	34.74/元	0.00%
A800-80G	16	256	系统盘: 20G 数据盘: 50GB	恒源云	9.00/元	-

资料来源：腾讯云，阿里云，恒源云，华鑫证券研究

1.1、产业动态：Arm 发布最强核心 Cortex-X925 CPU 及 Immortalis-G925 GPU，手机性能再上新台阶

5 月 29 日，芯片设计公司 Arm 发布了针对旗舰智能手机的下一代 CPU 和 GPU 设计：Cortex-X925 CPU 和 Immortalis G925 GPU。这两款芯片分别是目前用于联发科天玑 9300 处理器中的 Cortex-X4 和 Immortalis G720 的继任者。

图表 2：Arm 核心 CPU 图



资料来源：IT之家，华鑫证券研究

Cortex-X925 提供了 Cortex-X 系列至今最高的性能提升。主要由于 Arm 对微构架进行了重大调整，使 AI 算力 TOPS 增加了 50%，并采用更大的 3MB 的私有 L2 缓存（也可以配置为 2MB），同时通过 RTL 和实体设计团队间合作，达成在 3nm 制程的最优化。在 Geekbench 中，单线程性能提升了 36%，达到了历史上最大的同比性能提升；与上一代的 Cortex-X4 CPU 相比，Cortex-X925 的 AI 性能大幅提升，基于 Phi-3 模型的首个 token 生成时间缩短了 46%，可大幅改善大型语言模型的初始响应速度。

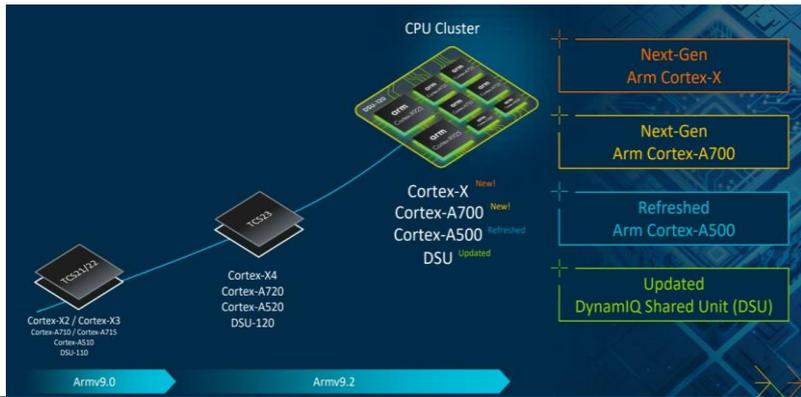
图表 3：Cortex-X925 性能改进



资料来源：Arm，华鑫证券研究

用于下一代 AI 体验的 Armv9 CPU 集群。新的 Armv9.2 CPU 集群为主流设备和日常用户带来了更高的性能和更强大的 AI 功能，在关键用户体验指标上提高了 30%。它提供了一整套实际改进，集群中的每个 CPU 组件都涵盖广泛的实际用例和工作负载。例如，Cortex-X925 可处理启动应用程序和 Web 浏览的“突发”工作负载；Cortex-A725 提供常见 AI 工作负载和 AAA 游戏所需的持续性能；Cortex-A520 的高效率最适合轻媒体以及空闲和后台任务。

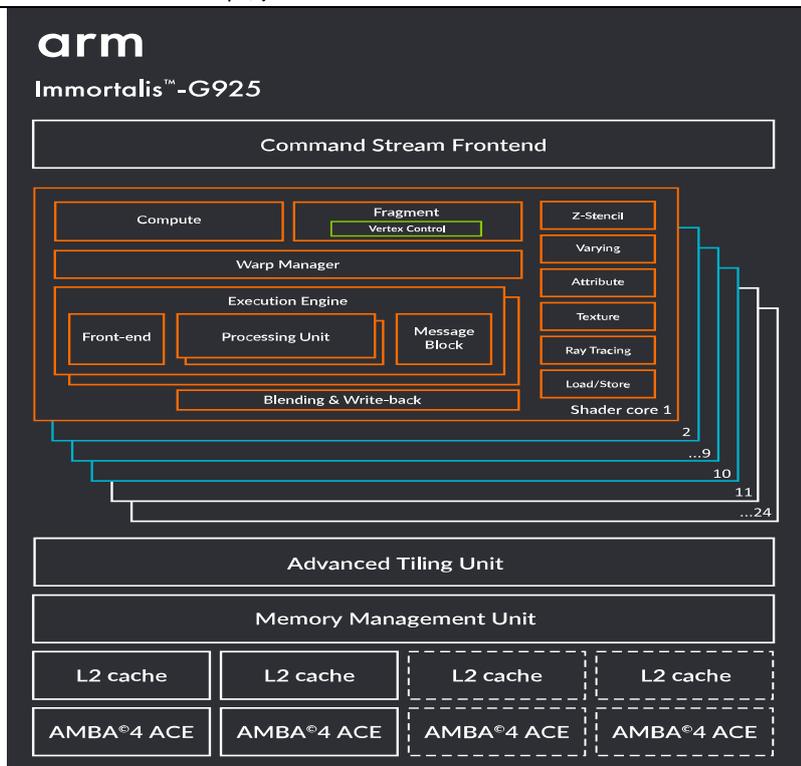
图表 4：Armv9 CPU 产品组合的新增和更新



资料来源：Arm，华鑫证券研究

迄今为止性能最高、效率最高的 GPU。Immortalis-G925 是专为旗舰智能手机设计的最新旗舰 GPU，基于第 5 代 Arm GPU 架构构建。Immortalis-G925 提供片段预传递、双平铺和移位转换单元吞吐量，从而实现更好、更持久的帧速率和更长的游戏时间。与 Immortalis G720 相比，Immortalis G925 在图形性能方面提升了 37%，相同性能下功耗可降低 30%，在一系列流行的手机游戏中支持以平均每秒 120 帧的帧率运行。面对复杂对象的光线追踪性能提升 52%。在 AI 和机器学习工作负载上 Immortalis G925 性能平均提升了 34%，例如 AI 推理速度比当前 GPU 要快 36%；自然语言处理速度最高可提升 50%；图像分割速度最高可提升 41%；语音转文本速度最高可提升 32%。

图表 5: Immortalis-G925 GPU 架构



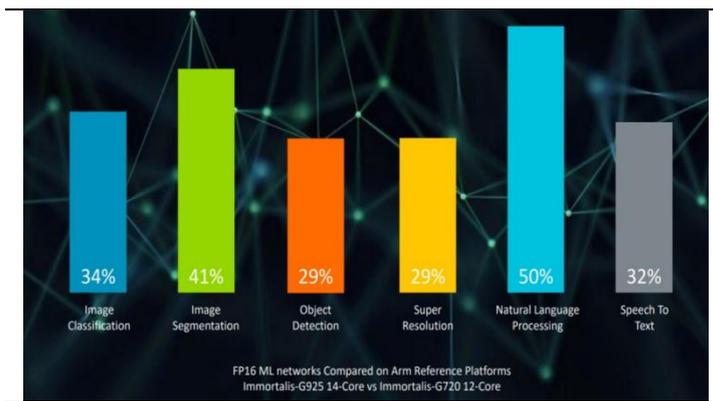
资料来源：Arm，华鑫证券研究

图表 6: Immortalis-G925 整体性能提升

图表 7: Immortalis-G925 AI 性能改进



资料来源：Arm，华鑫证券研究



资料来源：Arm，华鑫证券研究

2、AI 应用动态：Mistral 推出首个代码生成 AI 模型 Codestral，有望成为开源编程大模型先锋

2.1、流量跟踪：Kimi 访问量环比+15.41%

本期（2024. 5. 22-2024. 5. 28）AI 相关网站流量数据：访问量前三位分别为 Bing（310.9M）、Discord（262.6M）和 Canva（144.4M）；访问量环比增速前三位分别为 Kimi（15.41%）、Discord（3.02%）和 Midjourney（1.74%）；平均停留时长前三位分别为 Character.AI（0:15:03）、Canva（0:10:03）和 Midjourney（0:10:03）；平均停留时长环比增速前三位分别为 Character.AI（1.92%）、QuillBot（0.56%）和 NotionAI（0.20%）。

图表 8：2024. 5. 22-2024. 5. 28 AI 相关网站流量

应用	应用类型	归属公司	上周平均访问量 (M)	访问量环比	平均停留时长	时长环比
ChatGPT	聊天机器人	OpenAI	93.08M	-10.24%	0:03:07	-43.33%
Bing	搜索	微软	310.9M	-0.86%	0:06:57	-1.42%
Discord	游戏社区	微软	262.6M	3.02%	0:06:32	-0.25%
Canva	在线设计	Canva	144.4M	0.21%	0:10:03	0.00%
Github	代码托管	微软	103.7M	-0.38%	0:07:20	-0.68%

Gemini	聊天机器人	谷歌	91.5M	-8.63%	0:06:21	-1.04%
Character.AI	聊天机器人	Character.AI	65.2M	-1.42%	0:15:03	1.92%
NotionAI	文本/笔记	Notion	36.7M	-0.86%	0:08:13	0.20%
QuillBot	释义工具	QuillBot	13.4M	-6.74%	0:05:59	0.56%
Kimi	聊天机器人	Moonshot AI	5.8M	15.41%	0:03:21	-0.99%
DeepL	翻译工具	DeepL	63.4M	0.75%	0:08:56	-0.37%
文心一言	聊天机器人	百度	4.0M	-11.07%	0:01:32	0.00%
Midjourney	AI 绘画	Midjourney	4.3M	1.74%	0:10:03	——
天工 AI	聊天机器人	昆仑万维	1.3M	-8.15%	0:05:10	-0.76%

资料来源：similarweb, 华鑫证券研究

2.2、产业动态：Mistral 推出首个代码生成 AI 模型 Codestral，支持 80+编程语言

5月30日，Mistral 推出了首个代码生成模型 Codestral，支持超过 80 种编程语言，包含 Python, Java, C, C++, HTML, JavaScript 等前端语言。Codestral 基于 Llama 架构，尽管只有 22B 参数量，却实现了 32K 长上下文窗口，这一表现是拥有 70B 参数量的 Llama 3 的四倍。

Codestral 可以完成的任务涵盖了编写特定功能的函数，编写测试，以及代码填充等方面。开发人员可以通过 Le Chat 对话界面免费与 Codestral 进行交互，有助于减少代码错误与漏洞。

目前，Mistral 开放了两个 API 供开发者调用 Codestral，分别是 `codestral.mistral.ai` 和 `api.mistral.ai`。前者提供 8 周免费测试期，后者采用基于 token 的收费方式。

Mistral 展现的 Codestral 在各项语言中的部分表现如下：

1) **Python:** Mistral 使用 HumanEval 和 MBPP 评估 Codestral 的代码生成能力，CruxEval 评估输出预测，和 RepoBench 评估远程存储库中的代码补全能力。其中，在 RepoBench 基准中，Codestral 的能力达到了 SOTA（最佳）成绩。

图表 9：使用各项测试评估 Codestral 能力

	HumanEvalFIM Python	HumanEvalFIM JavaScript	HumanEvalFIM Java	HumanEval FIM
Codestral 22B	89.4%	95.1%	90.3%	91.6%
CodeLlama 70B	-	-	-	-
DeepSeek Coder 33B	65.4%	82.5%	86.6%	78.2%
Llama 3 70B	-	-	-	-
				Average

资料来源: Mistral 官网, 华鑫证券研究

Codestral 有望成为领先开源编程大模型, 商业化落地仍在规划。与此同时, Mistral 针对 Codestral 发布了全新“非生产”(Non-Production)许可协议, 即 MNPL (Mistral Non-Production License)。按照规定 Codestral 仅可用于研究目的, 不可用于商业用途。次协议的“非商用”界定十分严格, Mistral 公司也严禁内部使用。目前, Mistral 表示商业用途无法得到使用者对模型研发作出贡献, 但预计未来其他开源模型有望开放商业用途。

3、AI 融资动向: 马斯克的 xAI 获 60 亿美元 B 轮融资

本期 AI 初创公司的融资中, xAI/智谱 AI 融资额前二, 分别为 60 亿/4 亿美元。xAI 专注于回答更深层次的科学问题, 期望未来可以用 AI 去帮助人们解决复杂的科学和数学问题并且“理解”宇宙。本次融资将用于推出 xAI 的首个产品、建设先进基础设施, 并加速未来技术的研发。智谱 AI 致力于推动人工智能技术的边界, 尤其是认知智能领域。它的技术突破主要集中在大规模语言模型的开发上, 此外它也构建了高精度的通用知识图谱并自主开发了科技大数据检索平台 Aminer。

图表 13: 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
HoneyNaps	AI 睡眠技术	韩国兴业银行 嗨投资合伙人 QUAD 投资管理	B 轮	1160 万美 元	约 1620 万美元	——
赛福基因	基因大数据 AI 解析	锦江发展集团 中喜基金	C 轮	亿元人民币	超亿元人 民币	——
xAI	AI 系统开发	Andreessen Horowitz 红杉资本	B 轮	60 亿美元	——	240 亿美 元

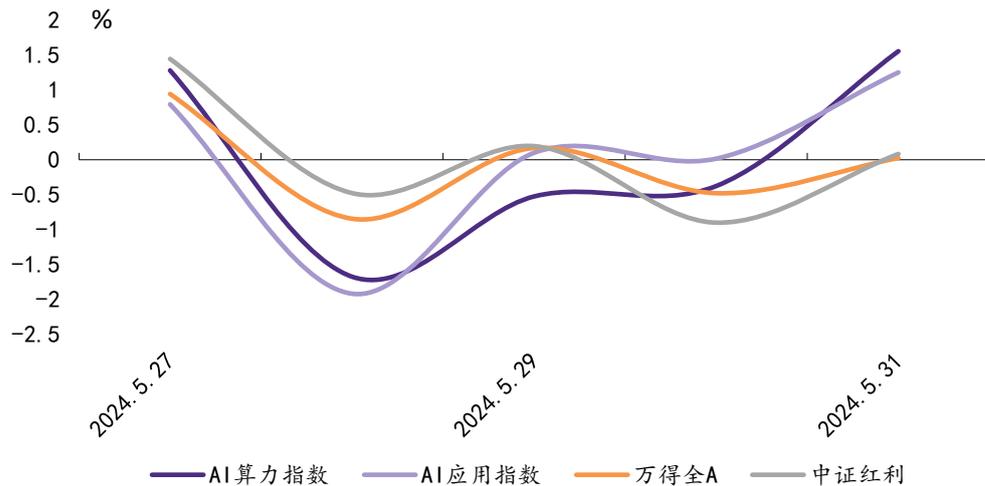
Transcend	数据隐私和 AI 治理平台	StepStone Group HighlandX	B 轮	4000 万美 元	——	——
爱设计	个人与组织的 PPT 工作站	视觉中国 星连资本 36 氪	B1 轮	数千万元 人民币	——	——
Iris.ai	AI 学术研究工具	Silverline Capital EIC 加速器基金	A 轮	764 万欧元	——	——
Perplexity	AI 搜索	BVP Databricks NEA AIX Ventures	第四轮	2.5 亿美元	——	30 亿美元
智谱 AI	AI 大模型	Prosperity7	C 轮	4 亿美元	超 5 亿美 元	约 30 亿美 元
Orca AI	商业船舶自动 驾驶	OCV Partners MizMaa Ventures	——	2300 万美 元	近 4000 万美元	——

资料来源：百度百科，AI 工具集 (ai-bot.cn)，搜狐网，币界网，36 氪官网，站长之家，阿尔法公社，数据学习网，企业预警通，businesswire，华鑫证券研究

4、行情复盘

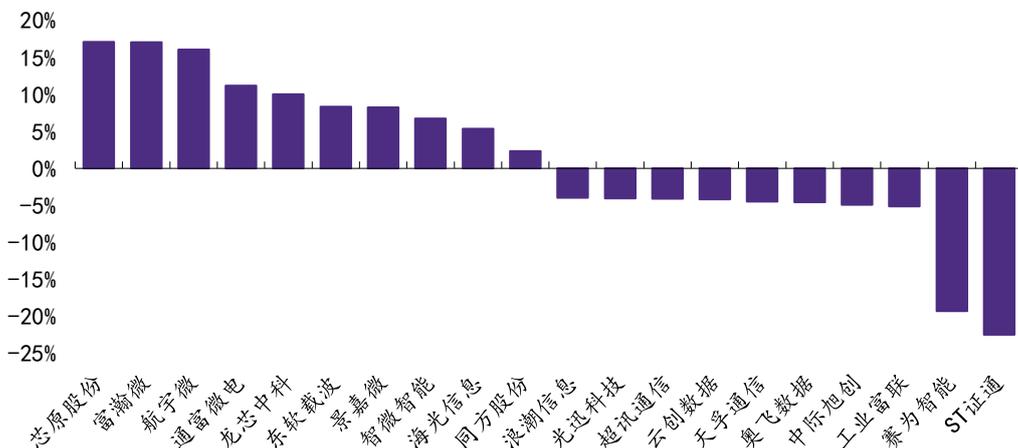
本周，AI 算力指数/AI 应用指数/万得全 A/中证红利的单日涨幅最大值分别为 +1.55%/+1.25%/+0.94%/+1.44%，单日跌幅最大值分别为 -1.67%/-1.92%/-0.84%/-0.90%。AI 算力指数内部，芯原股份以 +17.08% 录得本周最大涨幅，ST 证通以 -22.50% 录得本周最大跌幅。AI 应用指数内部，芯原股份以 +17.08% 录得本周最大涨幅，赛为智能以 -19.29% 录得本周最大跌幅。

图表 14：本周指数日涨跌幅



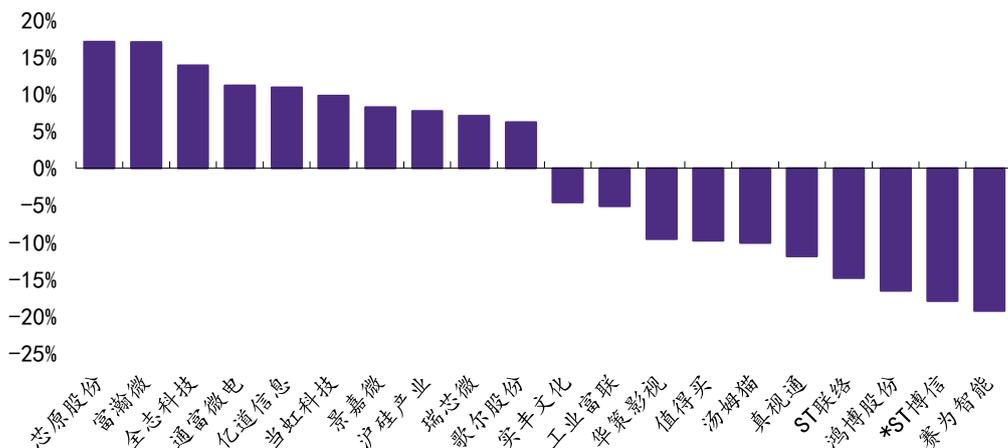
资料来源: wind, 华鑫证券研究

图 15: 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 16: 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

Arm 发布的针对旗舰智能手机的下一代 CPU 和 GPU 设计，预示着端侧 AI 的加速落地。端侧 AI 作为推动智能化转型的关键驱动力，其在边缘计算、数据隐私保护、实时性工业需求等场景中有出独特的技术优势，AI 的商业化落地应用将被大幅拓展，如安防、工业自动化、智能家居、医疗健康等多个垂类行业。在深度学习、自然语言处理（NLP）、计算机视觉等核心技术领域具有深厚积累的行业龙头厂商将核心受益。

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）等。

图表 17：重点关注公司及盈利预测

公司代码	名称	2024-06-02 股价	EPS			PE			投资评级
			2023	2024E	2025E	2023	2024E	2025E	
002230.SZ	科大讯飞	42.20	0.28	0.40	0.56	150.71	105.50	75.36	买入
002368.SZ	太极股份	21.35	0.79	1.01	1.28	27.03	21.14	16.68	买入
603918.SH	金桥信息	12.18	0.33	0.49	0.80	36.91	24.86	15.23	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

许思琪：澳大利亚国立大学硕士。

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。