

2024年05月19日



华鑫证券
CHINA FORTUNE SECURITIES

GPT-4o 场景理解力阶跃式提升，“贾维斯时刻”或即将到来

—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)	4.3	3.3	-26.2
沪深300	3.8	8.1	-6.8

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：月之暗面发布 Kimi+，新一轮 AI 应用军备竞赛进行时》2024-05-13
- 《计算机行业周报：商汤日日新 5.0 大模型综合性能或超越 GPT-4Turbo，端侧 AI 加速推进进行时》2024-04-29
- 《计算机行业周报：Adobe 携手 OpenAI 重塑视频编辑生态，Gaudi3 芯片引领 AI 算力新突破》2024-04-22

算力：A100-80G 租赁价格下降，谷歌发布第六代 TPU 芯片

本周跟踪算力租赁平台中，显卡配置为 A100-80G 中，恒源云 13 核+128G 价格为 9.00 元/时，价格环比下降 10.00%。产业动态上，谷歌公司于 5 月 15 日召开的 I/O 2024 开发者大会上，发布了第六代数据中心 AI 芯片 Tensor 处理器单元 (TPU) Trillium，并将于今年推出交付。与上一代 TPU v5e 相比，Trillium TPU 的每芯片峰值计算性能提高近 5 倍，为下一代 AI 模型的研发提供强大动力。

AI 应用：OpenAI 推出新一代旗舰 AI 模型 GPT-4o，反应速度大幅提升

5 月 13 日，OpenAI 正式发布 GPT-4o，实现实时多模态对话。用户可免费使用 GPTs, browsing, 图像理解, 代码解释器等此前付费功能。GPT-4o 在 GPT-4 的基础上改进了多模态能力，实现从模型到硬件 Infra 到推理工程软件的全面优化。其次，一月前英伟达交付 H200 与 OpenAI，其推理速度为 H100 的 2 倍，几乎达到人类的反应速度。同时，OpenAI 与苹果需求深度匹配，有望定义新一代 AI 手机，引发新一轮硬件接入大模型的激烈竞争。

AI 融资动向：机器人公司 The Bot Company、无代码云分析服务商 Sigma Computing 获最高融资额

本期 AI 初创公司的融资中，Sigma Computing/The Bot Company 融资额前二，分别为 2 亿/1.5 亿美元。Sigma Computing 专注为分析人员提供便捷的分析工具。The Bot Company 则致力于利用技术解决家庭杂务的挑战，本次融资将用于开发和部署专门设计用于处理家庭杂物的机器人。

投资建议

ChatGPT 4o 进一步印证算力需求。尽管部分国产 AI 大模型在训练优化算法方面取得了显著效果，例如通过诸如模型剪枝、量化和知识蒸馏等技术手段，有效降低训练与推理过程中的算力消耗。但 ChatGPT 4o 的原生多模态大模型，无疑将推动 AI

模型的复杂度和精细化程度向更高水平发展，意味着目前仍处算力密集型 AI 时代。

人机交互效果逼近“贾维斯时刻”，大幅拓展应用场景。GPT-4o 的先进能力预示着 AI 可以提供更加自然、流畅且富有洞察力的交互体验，类似于贾维斯与钢铁侠的互动，真正开放使用后，将大幅打开实时视觉助手、虚拟陪伴人等领域的应用场景。

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	2024-05-19 股价	EPS			PE			投资评级
			2023	2024E	2025E	2023	2024E	2025E	
002230.SZ	科大讯飞	43.19	0.28	0.40	0.56	154.25	107.98	77.13	买入
002368.SZ	太极股份	22.37	0.79	1.01	1.28	28.32	22.15	17.48	买入
603918.SH	金桥信息	13.25	0.33	0.49	0.80	40.15	27.04	16.56	买入

资料来源：Wind，华鑫证券研究

正文目录

1、 算力动态：A100-80G 租赁价格下降，谷歌发布第六代 TPU 芯片	4
1.1、 数据跟踪：A100-80G 租赁价格环比下降	4
1.2、 产业动态：谷歌发布第六代 TPU 芯片，计算性能提升近 5 倍	4
2、 AI 应用动态：OPENAI 推出新一代旗舰 AI 模型 GPT-4o，多模态交互再创高峰	5
2.1、 流量跟踪：Kimi 访问量环比增速 46.73%	5
2.2、 产业动态：OpenAI 推出新一代旗舰 AI 模型 GPT-4o，引领人机交互新时代	6
3、 AI 融资动向：无代码云分析服务商 SIGMA COMPUTING、机器人公司 THE BOT COMPANY 获最高融资额 ..	8
4、 行情复盘	9
5、 投资建议	11
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：Google I/O 发布现场	5
图表 3：2024.5.8-2024.5.14 AI 相关网站流量	6
图表 4：启动语音功能，GPT-4o 进行意大利语-英语实时交替翻译	7
图表 5：启动摄像头，GPT-4o 实时给出过程及答案	7
图表 6：启用摄像头展现面部，由 GPT-4o 分析情绪	7
图表 7：输入购物要求，由 Kimi+ 进行商品推荐	8
图表 8：本周 AI 初创公司的融资动态	8
图表 9：本周指数日涨跌幅	10
图表 10：本周 AI 算力指数内部涨跌幅度排名	10
图表 11：本周 AI 应用指数内部涨跌幅度排名	10
图表 12：重点关注公司及盈利预测	11

1、算力动态：A100-80G 租赁价格下降，谷歌发布第六代 TPU 芯片

1.1、数据跟踪：A100-80G 租赁价格环比下降

恒源云 A100-80G 价格下降 10.00%，其余环比持平。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 10.00 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	-10.00%
	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

1.2、产业动态：谷歌发布第六代 TPU 芯片，计算性能提升近 5 倍

5 月 15 日，谷歌公司在召开的 I/O 2024 开发者大会上，发布了第六代数据中心 AI 芯片 Tensor 处理器单元 (TPU) Trillium，并将于今年推出交付。此款芯片旨在为最苛刻的生成式人工智能模型和工作负载提供支持，提供比现有 TPU 显著增强的计算、内存和网络功能。

图表 2: Google I/O 发布现场



资料来源: Google I/O, 华鑫证券研究

迄今为止性能最高、能效最高的 TPU。全新的 Trillium 能够以更快的速度训练新一代 AI 模型，同时减少延迟和降低成本。与上一代 TPU v5e 相比，Trillium TPU 的每芯片峰值计算性能提高了 4.7 倍，为了达到这种性能水平，谷歌扩大了矩阵乘法单元 (MXU) 的大小并提高了时钟速度；高带宽内存 (HBM) 容量和带宽提高了一倍，可以实现缩短大型模型的训练时间和服务延迟；芯片间互联 (ICI) 带宽也提高了一倍，使训练和推理作业能够扩展到数万个芯片；作为迄今为止谷歌最具可持续性的 TPU，能效提高了超过 67%。

配备了第三代 SparseCore。SparseCore 是专业的加速器，用于处理先进的基于人工智能的排名和推荐系统中的超大嵌入。借此，Trillium TPU 将能够训练下一代基础人工智能模型，并以更低的延迟和成本为这些模型提供服务。

Trillium 将为下一代 AI 模型提供动力。Google Cloud 的一部分客户已经计划使用 Trillium TPU。例如 Google DeepMind 将使用新芯片来训练和服务未来几代 Google 的 Gemini 模型；Deep Genomics 正在利用人工智能为药物发现的未来提供支持。此外，Trillium TPU 还将与 Google Cloud 的 AI Hypercomputer 项目集成，这是一种专为尖端人工智能工作负载而设计的突破性超级计算架构。AI 超级计算机将最先进的 TPU 和 GPU 与开源软件框架和灵活的消费模式集成在一起，以支持开发极其强大的新 AI 工作负载。

2、AI 应用动态：OpenAI 推出新一代旗舰 AI 模型 GPT-4o，多模态交互再创高峰

2.1、流量跟踪：Kimi 访问量环比增速 46.73%

本期 (2024.5.8-2024.5.14) AI 相关网站流量数据：访问量前三位分别为 Bing (314.9M)、Discord (252.8M) 和 Canva (143.6M)；访问量环比增速前三位分别为 Kimi (46.73%)、文心一言 (34.39%) 和 Character.AI (6.02%)；平均停留时长前三位分别为 Character.AI (0:14:34)、Canva (0:10:03) 和 DeepL (0:08:54)；平均停留时长环比增速前三位分别为 Kimi (2.56%)、Gemini (1.31%) 和 QuillBot (1.15%)。

图表 3: 2024. 5. 8-2024. 5. 14 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	114.3M	-65.14%	0:06:36	-10.41%
Bing	搜索	微软	314.9M	-3.40%	0:07:10	-0.46%
Discord	游戏社区	微软	252.8M	-1.17%	0:06:33	-0.25%
Canva	在线设计	Canva	143.6M	2.64%	0:10:03	0.50%
Github	代码托管	微软	104.7M	1.65%	0:07:26	-0.22%
Gemini	聊天机器人	谷歌	93.7M	1.11%	0:06:26	1.31%
Character.AI	聊天机器人	Character.AI	60.6M	6.02%	0:14:34	0.58%
NotionAI	文本/笔记	Notion	37.3M	1.19%	0:08:15	-0.40%
QuillBot	释义工具	QuillBot	15.2M	-0.26%	0:05:53	1.15%
Kimi	聊天机器人	Moonshot AI	5.2M	46.73%	0:03:20	2.56%
DeepL	翻译工具	DeepL	63.8M	4.66%	0:08:54	——
文心一言	聊天机器人	百度	4.3M	34.39%	0:01:33	——
Zoom	在线会议	Zoom	2.8M	-2.37%	0:01:28	-6.38%

资料来源: similarweb, 华鑫证券研究

2.2、产业动态: OpenAI 推出新一代旗舰 AI 模型 GPT-4o, 引领人机交互新时代

当地时间 5 月 13 日, OpenAI 正式发布 GPT-4o, 其原生多模态模型特性大幅提升整合方面的精细度, 可以在语音, 文本和视频分析之间快速切换。这类模型具有思维链和涌现能力, 生成结果不断接近人类思考的关键特征。同时, 多模态模型能够实现更好的常识推理性能, 允许模型在大量模态数据中自然学习能力。

目前用户可免费使用 GPTs, browsing, 图像理解, 代码解释器等此前付费功能。GPT-4o 在 GPT-4 的基础上改进了文本, 视觉以及音频方面的能力, 实现从模型到硬件 Infra 到推理工程软件的全面优化。OpenAI 通过跨文本, 视觉和音频端到端训练新模型, 使所有输入和输出由同一神经网络处理。其次, 一月前英伟达交付 H200 与 OpenAI, 其推理速度为 H100 的 2 倍, 进一步降低延迟, 几乎达到人类的反应速度。发布会展现的 GPT-4o 部分具体功能如下:

1) **无延迟对话**: 区别于 GPT-4 模型先将用户语音转为文字, 生成回复文本, 再转回语音的长时间过程, GPT-4o 模型最短可以在 232 毫秒内相应对话。

图表 4: 启动语音功能, GPT-4o 进行意大利语-英语实时交替翻译



资料来源: ChatGPT-4o 发布会, 华鑫证券研究

2) 视频交互: 用户可以打开摄像头, 由 GPT-4o 给出提示引导用户完成每一步解答, 实时识别书写结果, 帮助用户解出 x 的值。

图表 5: 启动摄像头, GPT-4o 实时给出过程及答案



资料来源: ChatGPT-4o 发布会, 华鑫证券研究

3) 识别人类面部情绪: 用户可以启用摄像头向 GPT-4o 展现自己的面部, 由 GPT-4o 描述外貌并识别情绪。目前 GPT-4o 可以读取人类情绪, 但存在一定误差与困难, 或需尝试多次得出正确答案。

图表 6: 启用摄像头展现面部, 由 GPT-4o 分析情绪

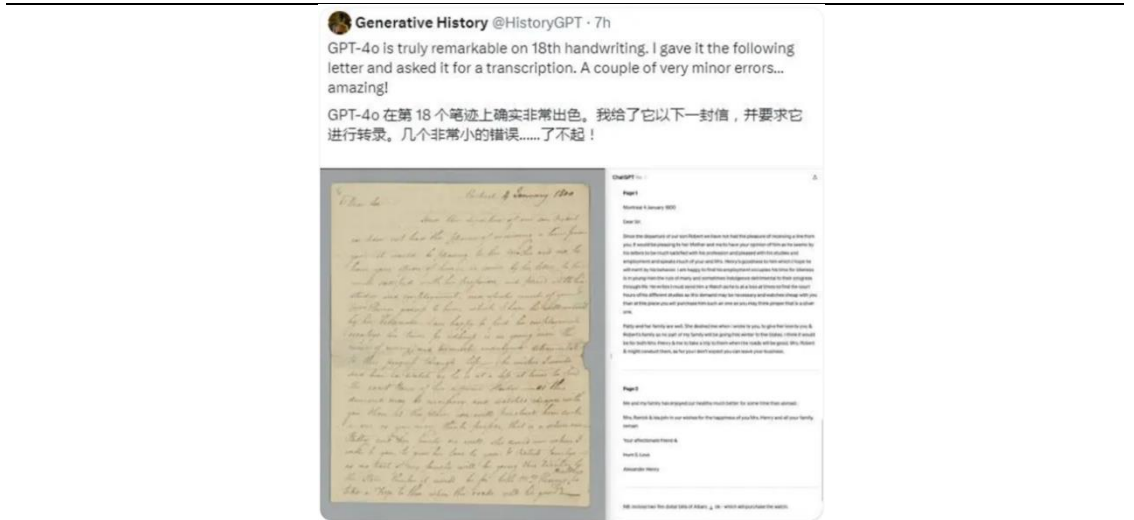


资料来源: OpenAI, 华鑫证券研究

4) 文本转录: 上传需要转录的图片, 由 GPT-4o 识别理解图片内容, 将手写体转录为

电子文本。

图表 7: 输入购物要求, 由 Kimi+ 进行商品推荐



资料来源: 华尔街见闻, 华鑫证券研究

GPT-4o 引领人机交互新时代, OpenAI 与苹果合作可期。近年来苹果为提高用户体验, 在设备中集成越来越多的 AI 技术, 可能与 OpenAI 达成合作。对端侧应用权限, 系统级权限的需求与苹果对 GPT-4o 此类成熟大模型的需求深度匹配, 或在 iOS 18 中达成紧密集成。用户无需下载 ChatGPT 应用软件, AI 助手就能实现无缝衔接的跨应用服务, 有望定义新一代 AI 手机, 引发新一轮硬件接入大模型的激烈竞争。

3、AI 融资动向: 无代码云分析服务商 Sigma Computing、机器人公司 The Bot Company 获最高融资额

本期 AI 初创公司的融资中, Sigma Computing/The Bot Company 融资额前二, 分别为 2 亿/1.5 亿美元。Sigma Computing 是无代码云分析服务商, 致力于为分析人员提供便捷的分析工具。本轮融资将主要用于产品研发和市场拓展。The Bot Company 是一家新兴的机器人初创公司, 公司致力于利用技术解决家庭杂务的挑战, 并在竞争激烈的市场中寻求突破。本次融资将用于开发和部署专门设计用于处理家庭杂物的机器人。

图表 8: 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
Niqo Robotics	AI 农业机器人	Bidra Innovation Ventures	B 轮	1300 万美元	2100 万美元	——

SmarterDx	临床人工智能解决方案	BVP Flare Capital Partners Floodgate	B 轮	5000 万 美元	7100 万 美元	——
The Bot Company	家务机器人	Nat Friedman Daniel Gross Nabeel Hyatt Patrick Collison John Collison Quiet Capital	种子 轮	1.5 亿美元	——	——
Nabla Bio	AI 药物开发 AI 蛋白质设计	Radical Ventures	A 轮	2600 万 美元	——	——
高仙机器人	商业清洁机器人	——	D 轮	5000 万 美元	——	——
WEKA	AI 原生数据 平台	NVIDIA Generation Investment Management Atreides Management	E 轮	1.4 亿美 元	——	16 亿美元
Sigma Computing	商业智能云 分析	Altimeter Capital Avenir Growth Capital	D 轮	2 亿美元	——	——
PolyAI	AI 呼叫中心 客服助手	英伟达 Hedosophia	——	5000 万 美元	——	近 5 亿美 元

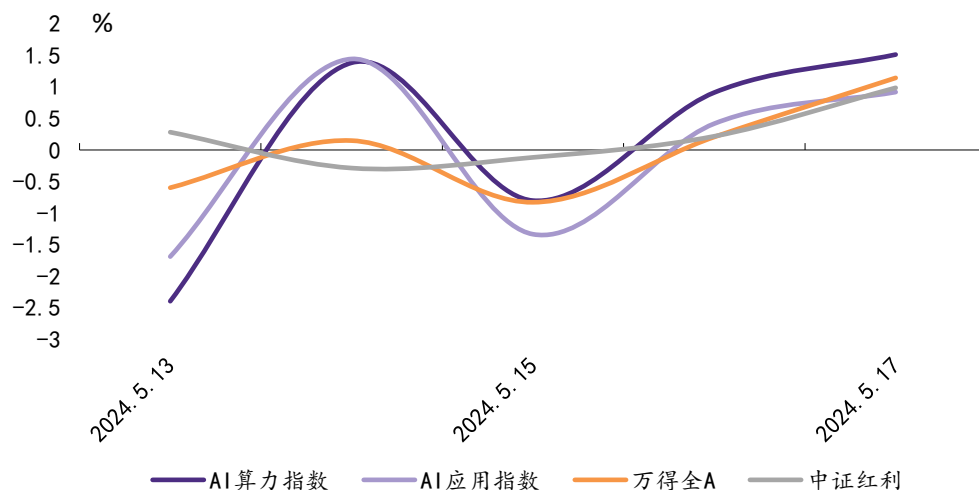
资料来源：亿欧数据，AI 工具集 (ai-bot.cn)，businessline，PR Newswire，BusinessWire，高仙机器人官网，动脉网，华鑫证券研究

4、行情复盘

本周，AI 算力指数/AI 应用指数/万得全 A/中证红利的单日涨幅最大值分别为 +1.51%/+1.44%/+1.14%/+0.99%，单日跌幅最大值分别为 -2.40%/-1.69%/-0.83%/-0.29%。AI

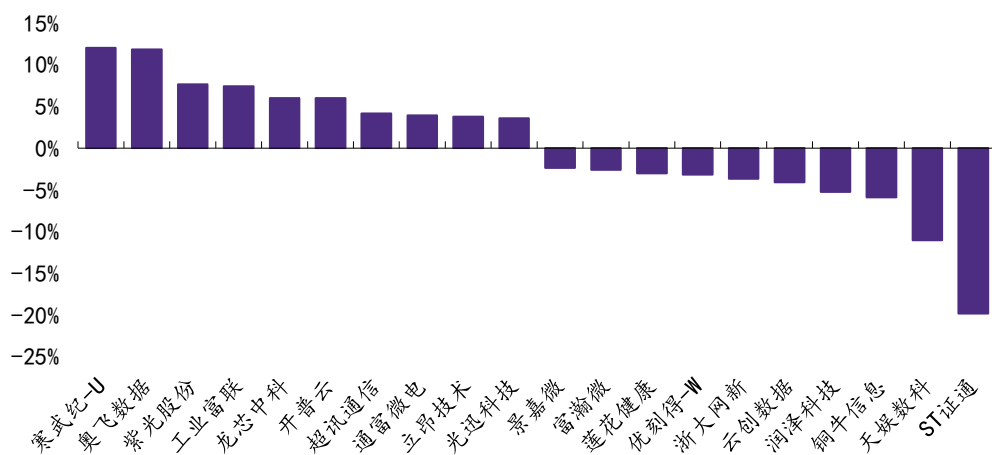
算力指数内部，寒武纪-U 以+12.02%录得本周最大涨幅，ST 证通以-19.81%录得本周最大跌幅。AI 应用指数内部，惠威科技以+15.03%录得本周最大涨幅，ST 联络以-15.54%录得本周最大跌幅。

图表 9：本周指数日涨跌幅



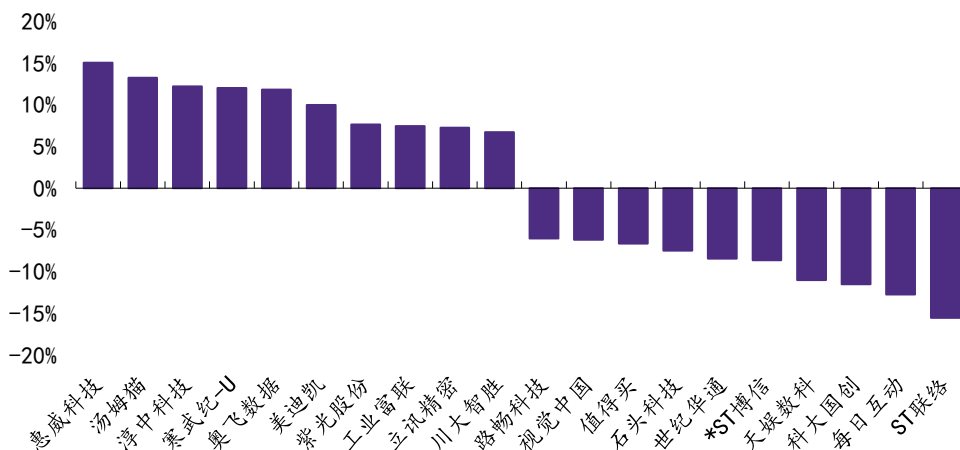
资料来源：wind, 华鑫证券研究

图表 10：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 11：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

ChatGPT 4o 进一步印证算力需求。尽管部分国产 AI 大模型在训练优化算法方面取得了显著效果, 例如通过诸如模型剪枝、量化和知识蒸馏等技术手段, 有效降低训练与推理过程中的算力消耗。但 ChatGPT 4o 的原生多模态大模型, 无疑将推动 AI 模型的复杂度和精细化程度向更高水平发展, 意味着目前仍处算力密集型 AI 时代。

人机交互效果逼近“贾维斯时刻”, 大幅拓展应用场景。GPT-4o 的先进能力预示着 AI 可以提供更加自然、流畅且富有洞察力的交互体验, 类似于贾维斯与钢铁侠的互动, 真正开放使用后, 将大幅打开实时视觉助手、虚拟陪伴人等领域的应用场景。

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 12: 重点关注公司及盈利预测

公司代码	名称	2024-05-19 股价	EPS			PE			投资评级
			2023	2024E	2025E	2023	2024E	2025E	
002230.SZ	科大讯飞	43.19	0.28	0.40	0.56	154.25	107.98	77.13	买入
002368.SZ	太极股份	22.37	0.79	1.01	1.28	28.32	22.15	17.48	买入
603918.SH	金桥信息	13.25	0.33	0.49	0.80	40.15	27.04	16.56	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。