

计算机

AI 大模型成果不断涌现，AGI 或将到来

➤ 人机交互体验优化，AGI 或将到来

北京时间5月14日，OpenAI发布了新一代多模态AI大模型GPT-4o，可接受文本、音频、图像和视频的任意组合作为输入，并生成文本、音频和图像的任意组合输出，响应时间与人类相似，人机交互效果更为自然。北京时间5月15日，谷歌召开年度开发者I/O大会，Gemini模型全面升级，发布文生图模型Imagen 3、视频生成模型Veo，此外全面赋能谷歌应用，包括AI搜索、Project Astra、Gemini Live等。

➤ 国产大模型能力对标GPT-4，成本快速下降

腾讯、阿里、字节跳动、商汤、幻方等近期发布了AI产品进展，国产大模型快速迭代升级，性能持续提升，创新成果不断涌现，如商汤日日新SenseNova 5.0综合性能全面对标GPT-4 Turbo，DeepSeek发布全球最强开源MoE模型，阿里云通义千问2.5 OpenCompass得分追平GPT-4 Turbo。商业化进程持续推进，如月之暗面推出Kimi+，MiniMax上架C端产品“海螺AI”。大模型成本快速下降，如智谱GLM-3 Turbo模型调用价格调整为0.001元/千tokens，豆包主力模型定价0.0008元/千tokens。

➤ 多模态能力提升、商业化推进、价格下降是行业发展趋势

AIGC的四个核心要素：数据、算力、模型和应用，正在逐步实现正反馈，呈现三大发展趋势。(1)多模态能力提升：大模型向原生多模态发展，带来更好的人机交互体验，拓展更广泛的应用场景；(2)商业化进程加速推进：各大AI厂商加速推进toC端产品、toB端服务的落地，利用AI能力重塑原有产品线；(3)大模型使用价格快速下降：随着模型、算力等的发展，大模型输入、输出价格快速下降，为商业化更好落地奠定了基础。

➤ 建议关注算力、端侧开发、办公软件、行业信息化等方面机遇

(1)算力基础设施建设机遇：大模型商用化进程不断推进，将提升算力需求，建议关注中科曙光、紫光股份、浪潮信息等；(2)端侧AI软件开发机遇：端侧设备将是人机交互的重要中介，带动软件开发机遇，建议关注中科达等；(3)生产力工具革新机遇：大模型有望重塑员工工作流程、企业组织架构等，建议关注金山办公、用友网络、泛微网络等；(4)行业信息化创新机遇：大模型在行业业务系统方面的应用场景不断扩展，建议关注恒生电子、宇信科技、中控技术、卫宁健康等。

风险提示：AI技术发展演进不及预期；商业化进程不及预期；法律政策监管风险；行业竞争加剧等。

投资建议： 强于大市（维持）
上次建议： 强于大市

相对大盘走势



作者

分析师：黄楷
执业证书编号：S0590522090001
邮箱：huangk@glsc.com.cn
分析师：陈安宇
执业证书编号：S0590523080004
邮箱：chenay@glsc.com.cn

相关报告

1、《计算机：2023年业绩恢复增长，2024Q1短期承压》2024.05.07
2、《计算机：金融IT业绩向好，受益信创、AI推进》2024.04.21

正文目录

1. 人机交互体验优化, AGI 或将到来	3
1.1 GPT-4o 迈向更自然的人机交互	3
1.2 谷歌全面升级基础大模型及应用	5
2. 国产大模型能力对标 GPT-4, 成本快速下降	6
2.1 商汤日日新: 综合性能全面对标 GPT-4 Turbo	6
2.2 幻方量化 Deepseek: 全球最强开源 MoE 模型	7
2.3 阿里云通义千问: OpenCompass 得分追平 GPT-4Turbo	7
2.4 腾讯混元: 拥抱开源, 产业实用	8
2.5 月之暗面 Kimi: AI Agent 进一步改善用户应用端体验	9
2.6 智谱 AI: 价格大幅下调, GLM 系列将迎来更新	9
2.7 Minimax: 上线产品对接 C 端用户, 大模型应用端持续拓展	10
2.8 字节豆包: 大模型费用进入“厘时代”, 月活达 2600 万	10
3. 投资建议: 建议关注算力、端侧开发、办公软件、行业信息化等方面机遇 ..	12
4. 风险提示	12

图表目录

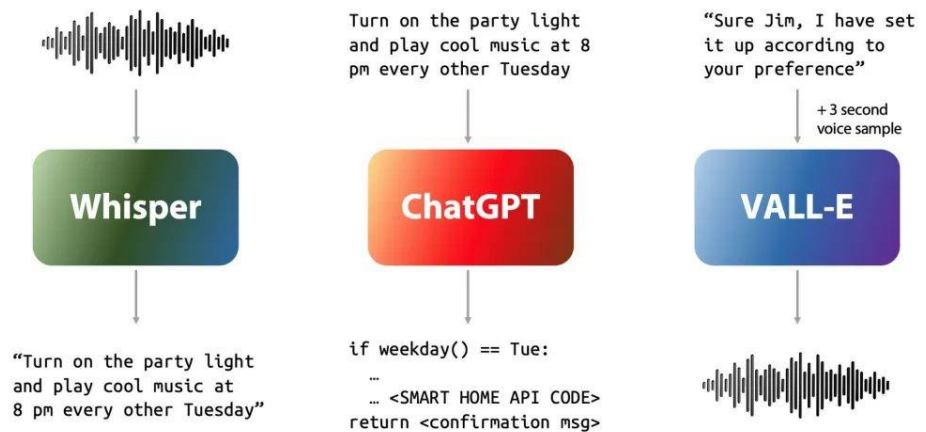
图表 1: 传统语音模式与 ChatGPT 对话的实现方式	3
图表 2: GPT-4o 在视觉感知基准上实现了最先进的性能	4
图表 3: 免费用户可体验 GPT-4o 的文本和图像功能	4
图表 4: 谷歌发布了一系列 AI 驱动的产品和功能	5
图表 5: 商汤日日新模型综合基准测试综合得分排名首位	6
图表 6: Deepseek-V2 媲美 GPT4 能力	7
图表 7: 阿里云通义千问模型矩阵	8
图表 8: 腾讯混元模型服务通过腾讯云对外开放	8
图表 9: Kimi+智能助手	9
图表 10: 智谱大模型标准版定价	10
图表 11: Minimax 发布海螺 AI 应用	10
图表 12: 豆包大模型产品矩阵	11

1. 人机交互体验优化，AGI 或将到来

1.1 GPT-4o 迈向更自然的人机交互

GPT-4o 实现原生多模态。在 GPT-4o 之前，用户使用语音模式与 ChatGPT 对话，平均延迟为 2.8 秒(GPT-3.5)和 5.4 秒(GPT-4)，其实现是通过三个独立模型组成的管道：一个简单模型将音频转录为文本，GPT-3.5 或 GPT-4 接收文本并输出文本，第三个简单模型将该文本转换回音频。这个过程意味着主要智能来源 GPT-4 丢失了大量信息——它无法直接观察音调、多个说话者或背景噪音，也无法输出笑声、歌唱或表达情感。北京时间 5 月 14 日凌晨，OpenAI 发布了新一代多模态人工智能 (AI) 大模型 GPT-4o，跨文本、视觉和音频端到端训练了一个新模型，这意味着所有输入和输出都由同一神经网络处理。

图表1：传统语音模式与 ChatGPT 对话的实现方式



@DrJimFan

资料来源：36Kr, X@Jim Fan 国联证券研究所

人机交互效果更为自然。GPT-4o 可接受文本、音频、图像和视频的任意组合作为输入，并生成文本、音频和图像的任意组合输出。它可以在短至 232 毫秒的时间内响应音频输入，平均为 320 毫秒，与人类的响应时间相似。根据基准测试，GPT-4o 在文本、推理和编码智能方面实现了 GPT-4 Turbo 级别的性能，同时，在多语言、音频和视觉功能上达到了新的水平。

图表2: GPT-4o 在视觉感知基准上实现了最先进的性能

Eval Sets	GPT-4o	GPT-4T 2024-04-09	Gemini 1.0 Ultra	Gemini 1.5 Pro	Claude Opus
MMMU (%) (val)	69.1	63.1	59.4	58.5	59.4
MathVista (%) (testmini)	63.8	58.1	53.0	52.1	50.5
AI2D (%) (test)	94.2	89.4	79.5	80.3	88.1
ChartQA (%) (test)	85.7	78.1	80.8	81.3	80.8
DocVQA (%) (test)	92.8	87.2	90.9	86.5	89.3
ActivityNet (%) (test)	61.9	59.5	52.2	56.7	
EgoSchema (%) (test)	72.2	63.9	61.5	63.2	

资料来源: OpenAI 官网, 国联证券研究所

免费用户可体验 GPT-4o 的文本和图像功能。在 ChatGPT 网页端, 免费用户可以体验 GPT-4o 的文本和图像功能, 不过额度有限; 付费用户的额度是免费用户的 5 倍。与 GPT-4 Turbo 相比, GPT-4o 速度提高 2 倍, 价格降低一半, 速率限制提高 5 倍。官方计划在未来几周内 API 中向一小部分值得信赖的合作伙伴推出对 GPT-4o 新音频和视频功能的支持。此外 ChatGPT 推出桌面端应用, 目前已经有 Mac 原生应用可安装, Windows 客户端预计会在晚些时候推出。

图表3: 免费用户可体验 GPT-4o 的文本和图像功能

Free	Plus	Team
每月 \$0 美元	每月 \$20 美元	每人每月 \$25 美元*
您当前的套餐	升级至 Plus	升级至 Team 套餐
<ul style="list-style-type: none"> 帮助您进行写作、解决问题等 对 GPT-3.5 的访问权限 对 GPT-4o 的有限访问权限 对高级数据分析、文件上传、视觉、网页浏览和自定义 GPT 等功能的有限访问权限 	<ul style="list-style-type: none"> 抢先体验新功能 对 GPT-4、GPT-4o、GPT-3.5 的访问权限 Up to 5x more messages for GPT-4o 对高级数据分析、文件上传、视觉和网页浏览功能的访问权限 DALL-E 图片生成 创建和使用自定义 GPT 	除了 Plus 套餐中的所有功能外, 还包括: <ul style="list-style-type: none"> GPT-4 和 GPT-4o 的更高限额, 以及 DALL-E 图片生成、高级数据分析、网页浏览等工具 创建 GPT 并与您的工作空间共享 适用于工作空间管理的管理员控制台 默认情况下, 团队数据不会用于训练。了解更多
已有套餐? 请参见 结算帮助	Limits apply.	*按年结算价格, 最少 2 位用户

资料来源: ChatGPT 官网, 国联证券研究所

1.2 谷歌全面升级基础大模型及应用

Gemini 模型全面升级。北京时间 5 月 15 日，谷歌召开了年度开发者 I/O 大会，发布了一系列 AI 驱动的产品和功能。Gemini 模型全面升级：(1) 升级版 Gemini 1.5 Pro 上下文长度突破 200 万 token，超过目前所有商用大模型；(2) 轻量级模型 Gemini 1.5 Flash 发布，相较于旧有模型，实现了低延迟响应并降低使用成本，100 万 tokens 只要 0.35 美元；(3) Gemini Nano 将具备多模态能力。多模态模型方面：(1) 谷歌发布文生图模型 Imagen 3 模型，能够更准确理解图像的文本提示词，并且生成的图片更具创造性和细节；(2) 发布视频生成模型 Veo，可以根据文本提示创建大约一分钟长的 1080p 视频剪辑，可捕捉不同的视觉和电影风格，包括风景和延时镜头，并对已生成的镜头进行编辑和调整。

图表4：谷歌发布了一系列 AI 驱动的产品和功能

分类	名称	特点	发布状态
AI 模型	Gemini 1.5 Pro	性能显著提升，拥有 200 万令牌上下文窗口，适合广泛任务的最佳模型。	升级
	Gemini 1.5 Flash	更轻量化、速度更快且成本更低，适用于大规模高频任务。	新发布
	Gemini Nano	多模态理解能力，支持文本和图像输入，适用于设备内应用。	升级
	Gemma 2.0	具有新架构，突破性的性能和效率，适用于负责任的 AI 创新。	新发布
	PaliGemma	第一个视觉语言模型，基于 PaLI-3，扩展了 Gemma 家族。	新发布
多模态生成模型	Veo	最先进的视频生成模型，生成高质量 1080p 视频，理解自然语言和视觉语义。	新发布
	Imagen 3	最高质量的文本到图像生成模型，细节丰富，逼真度高。	升级
AI 延展产品	Google Messages	集成 Gemini，实现更自然的聊天体验。	新增功能
	Gemini Live	使用最先进的语音技术，实现更加自然和直观的对话体验。	新发布
	Search Labs	提供生成性 AI 功能的实验平台，支持复杂问题解决和计划功能。	升级
	Gems	允许用户创建定制版本的 Gemini，满足特定需求和个性化响应。	新发布
	Project Astra	未来 AI 助理的愿景，目标是开发通用 AI 代理，能够理解和响应复杂环境。	新发布
	Music AI Sandbox	生成音乐的 AI 工具套件，支持音乐创作和实验。	升级
硬件	Trillium TPU	第六代 TPU，性能和能源效率显著提升，支持训练和服务最强大的 AI 模型。	新发布

资料来源：36Kr，国联证券研究所

Gemini 全面赋能谷歌应用。(1) AI 搜索：功能包括搜索摘要、多步推理功能、在搜索中做计划、以及用视频提问等；(2) Project Astra：通过手机和 AR 眼镜的联合使用，AI 能够在目之所及的范围内完成和用户的实时交互。(3) Gemini Live：用户可以通过手机和 Gemini 直接进行语音交互，Gemini Live 还能与谷歌日历、任务、备忘录等应用无缝连接。

2. 国产大模型能力对标 GPT-4，成本快速下降

国内互联网头部企业、人工智能创新企业、垂直领域厂商等重视 AI 产业发展趋势，持续进行研发投入，腾讯、阿里、字节跳动、商汤、幻方等近期发布了 AI 产品进展。国产大模型快速迭代升级，性能持续提升，创新成果不断涌现，商业化进程持续推进，成本快速下降。

2.1 商汤日日新：综合性能全面对标 GPT-4 Turbo

4 月 23 日，商汤科技带来全新升级的“日日新 SenseNova 5.0”大模型，具备更强的知识、数学、推理及代码能力，综合性能全面对标 GPT-4 Turbo，并在主流客观评测上达到或超越 GPT-4 Turbo。

商汤日日新提升能力主要方面：采用混合专家架构 (MoE)，激活少量参数就能完成推理。且推理时上下文窗口达到 200K 左右。基于超过 10TB tokens 训练、覆盖数千亿量级的逻辑型合成思维链数据。商汤 AI 大装置 SenseCore 算力设施与算法设计的联合调优。

自然语言能力上，创意写作、推理以及总结能力均大幅提升。文生图能力有较大提升。多模态和数据分析能力上，商汤多模态大模型的图文感知能力达到全球领先水平，在权威综合基准测试 MMBench 中综合得分排名首位；支持高清长图的解析和理解以及文生图交互式生成，还可以实现复杂的跨文档知识抽取及总结问答展示。

图表5：商汤日日新模型综合基准测试综合得分排名首位

	MMBench Test	MMBench-CN Test	MME	MMStar	MMMU Val	MathVista MiniTest	Hallusion Bench Avg.	AIZD Test	OCRBench	SEEDBench_IMG	MMVet
InternVL-Chat-v1.5 Shanghai AI Laboratory & SenseTime & Tsinghua	82.3	80.7	2189.6	57.1	46.8	52.7	47.4	80.6	720	75.2	55.4
Step-1V StepFun	80.7	79.9	2206.4	50	49.9	43	48.4	79.2	625	70.3	63.3
InternLM-XComposer2-VL Shanghai AI Lab	80.7	79.4	2220.4	56.2	41.4	57.9	41	81.2	532	74.9	46.7
Qwen-VL-Max Alibaba Group	77.6	75.7	2281.7	36.5	52	41.4	41.2	75.7	640	72.7	61.8
GPT-4v OpenAI	77	74.4	1771.5	47.7	53.8	47.8	46.5	75.5	516	71.6	56.8
GeminiProVision Google	73.6	74.3	2148.9	38.6	48.9	45.8	45.2	70.2	680	70.7	59.2

资料来源：商汤科技公众号，国联证券研究所

2.2 幻方量化 Deepseek: 全球最强开源 MoE 模型

5月6日,幻方量化正式开源第二代 MoE 模型:DeepSeek-V2。该模型中文综合能力 (AlignBench) 开源模型中最强:与 GPT-4-Turbo, 文心 4.0 等闭源模型在评测中处于同一梯队。英文综合能力 (MT-Bench) 处于第一梯队:英文综合能力 (MT-Bench) 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。知识、数学、推理、编程等榜单结果位居前列。开源模型支持 128K 上下文,对话官网/API 支持 32K 上下文。该模型使用 236B 参数量,32K 上下文 (对话/API),费用上采取 1 元/百万输入 Tokens, 2 元/百万输出 Tokens, 同时兼容 OpenAI API 接口。

图表6: Deepseek-V2 媲美 GPT4 能力

Model	是否	中文综合	英文综合	知识	基础算数	数学解题	逻辑推理	编程
	开源	AlignBench	MT-Bench	MMLU	GSM8K	MATH	BBH	HumanEval
DeepSeek-V2	✓	7.91	8.97	77.8	92.2	53.9	79.7	81.1
GPT-4-Turbo-1106	✗	8.01	9.32	84.6	93.0	64.1	-	82.2
GPT-4-0613	✗	7.53	8.96	86.4	92.0	52.9	83.1	84.1
GPT-3.5	✗	6.08	8.21	70.0	57.1	34.1	66.6	48.1
Gemini 1.5 Pro	✗	7.33	8.93	81.9	91.7	58.5	84.0	71.9
Claude 3 Opus	✗	7.62	9.00	86.8	95.0	61.0	86.8	84.9
Claude 3 Sonnet	✗	6.70	8.47	79.0	92.3	40.5	82.9	73.0
Claude 3 Haiku	✗	6.42	8.39	75.2	88.9	40.9	73.7	75.9
abab-6.5 (MiniMax)	✗	7.97	8.82	79.5	91.7	51.4	82.0	78.0
abab-6.5s (MiniMax)	✗	7.34	8.69	74.6	87.3	42.0	76.8	68.3
ERNIE-4.0 (文心一言)	✗	7.89	7.69	-	91.3	52.2	-	72.0
GLM-4 (智谱清言)	✗	7.88	8.60	81.5	87.6	47.9	82.3	72.0
Moonshot-v1 (月之暗面)	✗	7.22	8.59	-	89.5	44.2	-	82.9
Baichuan 3 (百川)	✗	-	8.70	81.7	88.2	49.2	84.5	70.1
Qwen1.5 72B (通义千问)	✓	7.19	8.61	76.2	81.9	40.6	65.9	68.9
LLaMA 3 70B	✓	7.42	8.95	80.3	93.2	48.5	80.1	76.2
Mixtral 8x22B	✓	6.49	8.66	77.8	87.9	49.8	78.4	75.0

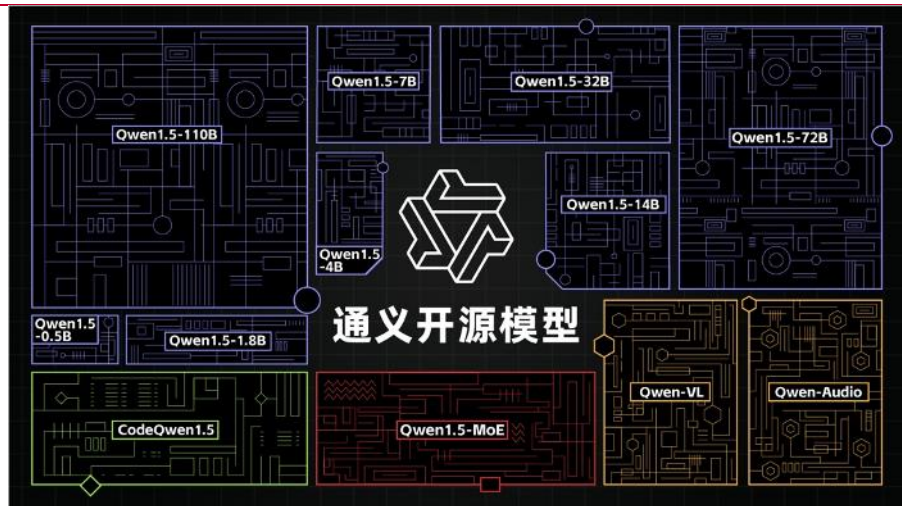
资料来源: DeepSeek 公众号, 国联证券研究所

2.3 阿里云通义千问: OpenCompass 得分追平 GPT-4Turbo

5月9日,在北京阿里云 AI 智领者峰会上,阿里云正式发布通义千问 2.5, 并开源通义千问 1100 亿参数模型。2.5 版模型的理解能力、逻辑推理、指令遵循、代码能力分别提升 9%、16%、19%、10%, 中文能力更是持续领先业界。

通义发布 1100 亿参数开源模型 Qwen1.5-110B, 该模型在 MMLU、TheoremQA、GPQA 等基准测评中超越了 Meta 的 Llama-3-70B 模型; 在 HuggingFace 推出的开源大模型排行榜 Open LLM Leaderboard 上, Qwen1.5-110B 冲上榜首, 再度证明通义开源系列业界最强的竞争力。

图表7：阿里云通义千问模型矩阵



资料来源：阿里云公众号，国联证券研究所

2.4 腾讯混元：拥抱开源，产业实用

5月14日，腾讯宣布旗下的混元文生图大模型全面升级并对外开源。混元文生图大模型是首个采用中文原生开发，同时支持中英文输入和理解的 DiT 架构模型，参数规模达到 15 亿。模型采用了与 sora 模型相同的 DiT 架构，使其不仅适用于文本生成图像，也能作为视频等多模态视觉内容生成的基础。

5月17日，腾讯云生成式 AI 产业应用峰会在北京召开，公布大模型研发、应用产品的系列进展。腾讯混元大模型能力持续升级，多个版本模型 hunyuan-pro、hunyuan-standard、hunyuan-lite 通过腾讯云对外开放，满足企业客户、开发者在不同场景下的模型需求，落地最优性价比模型方案。腾讯云大模型知识引擎、图像创作引擎、视频创作引擎三大工具发布，打造大模型时代原生工具链，通过 PaaS 服务简化数据接入、模型精调、应用开发流程，助力企业更高效、简单地用大模型开发 AI 原生应用，快速接入生产场景。

图表8：腾讯混元模型服务通过腾讯云对外开放

腾讯混元大模型目前提供3个版本：可根据不同业务需求按需选择。



资料来源：51CTO，国联证券研究所

2.5 月之暗面 Kimi：AI Agent 进一步改善用户应用端体验

5月6日，月之暗面旗下kimi智能助手迎来大更新推出Kimi+。Kimi+首批上线23个智能体，分别是官方推荐、办公提效、辅助写作、社交娱乐、生活实用5大领域。在输入框里打“@”，就可以召唤出用户所需要的工具人Kimi+。Kimi与应用端的联动进一步推动了AI应用的发展，从而给客户带来更好的应用体验。

图表9：Kimi+智能助手



资料来源：Kimi+官网，国联证券研究所

2.6 智谱 AI：价格大幅下调，GLM 系列将迎来更新

GLM-3 Turbo 模型调用价格由 5 元/百万 tokens 调整为 1 元/百万 tokens，最新价格相当于 0.001 元/千 tokens。公司即将推出 GLM-3 Turbo Batch 批处理 API（异步任务处理），适用于无需实时响应场景，包括效果评估、数据批处理等任务，价格为 1 元/200 万 token。公司对于新注册的用户，赠送额度从 500 万 tokens 提升至 2500 万 tokens（包含 2000 万入门级额度和 500 万企业级额度）。

企业版 GLM-4/GLM-4V 实现了视觉语言特征的深度融合，支持视觉问答、图像字幕、视觉定位、复杂目标检测等各类图像理解任务。在上海人工智能实验室发布的大模型开源开放评测体系司南（OpenCompass 2.0）中，GLM-4 综合排名第二，仅次于 GPT-4 Turbo。

图表10：智谱大模型标准版定价

模型	说明	上下文长度	单价
GLM-4	提供了更强大的问答和文本生成能力。适合于复杂的对话交互和深度内容创作设计的场景。	128K	0.1元 / 千tokens
GLM-4V	实现了视觉语言特征的深度融合，支持视觉问答、图像字幕、视觉定位、复杂目标检测等各类图像理解任务。	2K	0.1元 / 千tokens 
GLM-3-Turbo	适用于对知识量、推理能力、创造力要求较高的场景，比如广告文案、小说写作、知识类写作、代码生成等。	128K	0.001元 / 千tokens

资料来源：智谱 AI 开放平台，国联证券研究所

2.7 Minimax：上线产品对接 C 端用户，大模型应用端持续拓展

5月15日，大模型公司MiniMax宣布上架C端产品“海螺AI”，背后接入了MiniMax自研的多模态大模型，包括万亿参数 MoE 大语言模型 abab6.5，语音大模型和图像大模型。该产品支持输入 200ktokens 的上下文长度，1秒内处理近 3 万字的文本。可以进行书籍、长篇报告、学术论文等长篇内容的阅读、分析和文本写作。

图表11：Minimax 发布海螺 AI 应用



资料来源：海螺 AI 官网，国联证券研究所

2.8 字节豆包：大模型费用进入“厘时代”，月活达 2600 万

5月15日，在 2024 春季火山引擎 Force 原动力大会上，字节跳动正式发布了字节跳动豆包大模型家族。目前，豆包大模型日均处理 1200 亿 tokens 文本、生成 3000

万张图片，双端月活用户量达到 2600 万。豆包模型已用于豆包 App、扣子、河马爱学、飞书智能伙伴、抖音电商、剪映、番茄小说等字节跳动旗下产品及业务，并通过火山方舟向智能终端、汽车、金融、消费等行业的众多客户提供服务。基于豆包大模型同名打造的豆包 APP，在苹果 APP Store 和各大安卓应用市场，豆包 APP 的下载量在 AIGC 类应用中排名第一。豆包主力模型在企业市场的定价只有 0.0008 元/千 tokens，0.8 厘就能处理 1500 多个汉字，比行业便宜 99.3%。

图表12：豆包大模型产品矩阵



资料来源：澎湃网，国联证券研究所

3. 投资建议：建议关注算力、端侧开发、办公软件、行业信息化等方面机遇

AIGC 的四个核心要素：数据、算力、模型和应用，正在逐步实现正反馈。海内外厂商积极投入 AI 大模型研发，发布新产品及应用，呈现三大发展趋势。(1) **多模态能力提升**：大模型正在由多个模型简单协作向原生多模态发展，带来更好的人机交互体验，拓展更广泛的应用场景；(2) **商业化进程加速推进**：各大 AI 厂商在提升基础大模型性能的同时，加速推进 toC 端产品、toB 端服务的落地，利用 AI 能力重塑原有产品线；(3) **大模型使用价格快速下降**：随着模型、算力等的发展，大模型输入、输出价格快速下降，为商业化更好落地奠定了基础。

AI 大模型产业的发展有望带动四方面投资机遇。(1) **算力基础设施建设机遇**：大模型商用化进程的不断推进，将为算力基础设施建设提出新的要求，国产 GPU 生态体系正在加速发展，建议关注中科曙光、紫光股份、浪潮信息等领军企业；(2) **端侧 AI 软件开发机遇**：云端协同的混合 AI 是未来的方向，端侧设备将是人机交互的重要中介，将带动端侧 AI 应用的软件开发机遇，建议关注中科创达等相关公司；(3) **生产力工具革新机遇**：大模型有望为生产力工具带来降低专业门槛、减少重复劳动等变化，员工工作流程、企业组织架构等有望重塑，建议关注金山办公、用友网络、泛微网络等；(4) **行业信息化创新机遇**：受益于人机交互能力提升，大模型有望率先在政务、金融等领域窗口服务或培训场景落地，多模态能力提升也将扩展大模型在行业业务系统方面的应用场景，行业信息化厂商将是连接基础大模型厂商和行业客户的重要环节，建议关注恒生电子、宇信科技、中控技术、卫宁健康等行业信息化头部企业。

4. 风险提示

AI 技术发展演进不及预期；商业化进程不及预期；法律政策监管风险；行业竞争加剧等。

分析师声明

本报告署名分析师在此声明：我们具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，本报告所表述的所有观点均准确地反映了我们对标的证券和发行人的个人看法。我们所得报酬的任何部分不曾与，不与，也将不会与本报告中的具体投资建议或观点有直接或间接联系。

评级说明

投资建议的评级标准		评级	说明
报告中投资建议所涉及的评级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：以报告发布日后的6到12个月内的公司股价（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。其中：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；香港市场以摩根士丹利中国指数为基准；美国市场以纳斯达克综合指数或标普500指数为基准；韩国市场以柯斯达克指数或韩国综合股价指数为基准。	股票评级	买入	相对同期相关证券市场代表指数涨幅20%以上
		增持	相对同期相关证券市场代表指数涨幅介于5%~20%之间
		持有	相对同期相关证券市场代表指数涨幅介于-10%~5%之间
		卖出	相对同期相关证券市场代表指数跌幅10%以上
	行业评级	强于大市	相对同期相关证券市场代表指数涨幅10%以上
		中性	相对同期相关证券市场代表指数涨幅介于-10%~10%之间
		弱于大市	相对同期相关证券市场代表指数跌幅10%以上

一般声明

除非另有规定，本报告中的所有材料版权均属国联证券股份有限公司（已获中国证监会许可的证券投资咨询业务资格）及其附属机构（以下统称“国联证券”）。未经国联证券事先书面授权，不得以任何方式修改、发送或者复制本报告及其所包含的材料、内容。所有本报告中使用的商标、服务标识及标记均为国联证券的商标、服务标识及标记。

本报告是机密的，仅供我们的客户使用，国联证券不因收件人收到本报告而视其为国联证券的客户。本报告中的信息均来源于我们认为可靠的已公开资料，但国联证券对这些信息的准确性及完整性不作任何保证。本报告中的信息、意见等均仅供客户参考，不构成所述证券买卖的出价或征价邀请或要约。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。客户应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专家的意见。对依据或者使用本报告所造成的一切后果，国联证券及/或其关联人员均不承担任何法律责任。

本报告所载的意见、评估及预测仅为本报告出具日的观点和判断。该等意见、评估及预测无需通知即可随时更改。过往的表现亦不应作为日后表现的预示和担保。在不同时期，国联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。

国联证券的销售人员、交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。国联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。国联证券的资产管理部门、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的意见或建议不一致的投资决策。

特别声明

在法律许可的情况下，国联证券可能会持有本报告中提及公司所发行的证券并进行交易，也可能为这些公司提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。因此，投资者应当考虑到国联证券及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突，投资者请勿将本报告视为投资或其他决定的唯一参考依据。

版权声明

未经国联证券事先书面许可，任何机构或个人不得以任何形式翻版、复制、转载、刊登和引用。否则由此造成的一切不良后果及法律责任有私自翻版、复制、转载、刊登和引用者承担。

联系我们

北京：北京市东城区安定门外大街208号中粮置地广场A塔4楼
无锡：江苏省无锡市金融一街8号国联金融大厦12楼
 电话：0510-85187583

上海：上海浦东新区世纪大道1198号世纪汇一座37楼
深圳：广东省深圳市福田区益田路4068号卓越时代广场1期13楼