

2024年10月21日



华鑫证券
CHINA FORTUNE SECURITIES

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

—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)	44.9	40.1	4.8
沪深300	22.6	10.9	11.8

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：英特尔发布新一代处理器，ByteDanceResearch 提出机器人模型》2024-10-14
- 《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
- 《计算机行业周报：苹果推出 iOS18，阿里云发布 Qwen2.5》2024-09-29

算力：算力租赁价格平稳，安卓首款 3nm AI 芯片天玑 9400 发布

安卓首款 3nm、PC 级 Arm V9 架构、第八代 NPU——天玑 9400 正式发布，比上一代生成式 AI 芯片天玑 9300，其 AI 能力更上一层楼，成为苏黎世 ETHZ 移动 SoC AI 性能榜单之首。具体提升如下：

- 1) 天玑 9400 首次实现了端侧 DiT 架构支持，无需联网就能在手机上体验 Sora 同款架构视频生成。还有业界首发的端侧 LoRA 训练，可以高效离线训练专属于自己的 LoRA 生图模型，同时保证隐私不泄露。
- 2) 大模型方面，天玑 9400 能够运行的模型窗口文本长度提升到了 32K，是天玑 9300 的 8 倍。同时还支持端侧运行多模态大模型，并以 50 Tokens 每秒的超高速速度超越了前 SOTA。
- 3) 综合 AI 性能方面，天玑 9400 以 6773 分的成绩领跑苏黎世 ETHZ 移动 SoC 榜单，是天玑 9300 的 1.4 倍。
- 4) 从整体上看，天玑 9400 采用台积电第二代 3nm 制程，相较上一代单核性能提升 35%，多核性能提升 28%，同性能功耗降低 40%，

AI 应用：零一万物正式对外发布新旗舰模型——Yi-Lightning

零一万物正式对外发布新旗舰模型——Yi-Lightning (闪电)。在大模型竞技场 (Chatbot Arena) 上，Yi-Lightning 性能直冲总榜单并列第 6，数学分榜并列第 3，代码等其它分榜也名列前茅。总成绩几乎与马斯克最新 xAI 大模型 Grok-2-08-13 持平，超越 GPT-4o-2024-05-13、GPT-4o-mini-2024-07-18、Claude 3.5 Sonnet 等顶流。在中文榜单上，Yi-Lightning 超越了 xAI 发布的 Grok-2-08-13、智谱发布的 GLM-4-Plus 等国内外优质模型，与 o1-mini 等模型并列排名世界第二。

Yi-Lightning 采用 MoE 混合专家模型架构。底层技术上，Yi-Lightning 在以下方面进行了提升。

优化混合注意力机制，只在模型的部分层次中将传统的全注意力 (Full Attention) 替换为滑动窗口注意力 (Sliding Window Attention)。由此以来，模型在保证处理长序列数据高性能表现的同时，还能大大降低推理成本。Yi-Lightning 还引入了跨层注意力 (Cross-Layer Attention, CLA)，允许模型在不同的层次之间共享键 (Key) 和值 (Value) 头，减少对存储需求。这使得 Yi-Lightning 能在不同层次之间更有效地共享信息。总的来说，KV cache 缩小了 2-4 倍，同时将计

算的复杂度从 $O(L^2)$ 降至 $O(L)$ 。”

动态 Top-P 路由机制。Yi-Lightning 可以根据任务的难度动态自动选择最合适的专家网络组合——训练过程中会激活所有专家网络，使模型能学习到所有专家知识；而推理阶段，根据任务的难度，模型会选择性激活更匹配的专家网络。

多阶段训练。团队将整个训练分为两块，一块做好以后就把它固定起来，然后在这个固定的模型上再做后段训练。训练前期，更注重数据多样性，使得 Yi-Lightning 尽可能学习不同的知识；训练后期更重内容更丰富、知识性更强的数据。同时团队还在不同阶段采用不同的 batch size 和 LR schedule 保证训练速度和稳定性。

谈起和国外头部大模型的差距，李开复博士表示这次 Yi-Lightning 的排名证明了国产大模型跟硅谷最顶尖模型的差距缩小到了五个月。去跟追上美国最顶尖的模型，缩短时间差非常困难，要付出很大的努力和有独特的打法。在国内不少公司都在努力，“模基共建”则是零一万物自己摸索的独特路径。

■ AI 融资动向：Abridge/Terray Therapeutics 分别获 2.5 亿美元/1.2 亿美元融资

本期 AI 初创公司的融资中，Abridge/Terray Therapeutics 融资额位居前二，分别获 2.5 亿美元/1.2 亿美元融资。Abridge 的平台可通过移动应用程序访问，医生在每次预约前激活该应用，软件即可记录对话、转录并自动生成医疗记录。虽然这一过程并非完全自动化，医生仍需检查草稿的准确性并进行必要编辑，但 Abridge 提供了一项名为链接证据的功能，将每个 AI 生成的句子与相应的笔录摘录链接起来，从而简化了错误发现的过程。Terray 一直致力于通过规模实验和生成式 AI，提高小分子药物发现和开发的速度和成功率。依托其专有实验平台，Terray 构建了全球最大的化学数据集之一。目前 Terray 已定量测量了超过 50 亿次靶标-配体相互作用，该数据集仍在每年翻倍增长。Terray 的数据优势赋予其潜在 best-in-class 的 AI 能力，可识别并优化解决复杂问题的新型小分子。Terray 将这一专长应用于专注于免疫疾病的内部管线，同时在与 BMS 和 Calico 的合作中，针对更多具有挑战性的靶点开展研发。

■ 投资建议

一方面，国产 ai 大模型有望迎来新机遇，10 月 17 日，微软宣布将于 10 月 21 日停止在中国大陆的个人 Azure OpenAI 服务。OpenAI 停止服务后，多家国内大型企业，如阿里巴巴、百度与汇量科技迅速推出了针对这一市场变动的应对措施，力求吸引迁移流失客户，昆仑万维与讯飞等公司表明 OpenAI 用户的迁移需求在短时间内显著增加。另一方面，台积电 Q3 财报全面超预期提振市场信心，在截至 9 月 30 日的季度中，台积电营收录得 7596.9 亿新台币（235 亿美元），较上年同期增长了 39%（以美元计增长了 36%），超过了市场预期的 7421.66 亿元新台币，智能手机和人工智能对 3nm 和 5nm 技术的强劲需求支撑了其业务。我们继续看好 AI 软硬件机会，近期优先关注铜缆高速连接板块。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	2024-10-21 股价	EPS			PE			投资评级
			2023	2024E	2025E	2023	2024E	2025E	
002230.SZ	科大讯飞	45.83	0.28	0.40	0.56	163.68	114.58	81.84	买入
002368.SZ	太极股份	23.29	0.79	1.01	1.28	29.48	23.06	18.20	买入
688256.SH	寒武纪-U	429.71	-2.04	-1.21	-0.50	-210.64	-355.13	-859.42	买入
688668.SH	鼎通科技	43.93	0.67	1.04	1.41	65.57	42.24	31.16	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， 安卓首款 3NM AI 芯片天玑 9400 发布	5
1.1、 数据跟踪：算力租赁价格平稳	5
1.2、 产业动态：安卓首款 3nm AI 芯片天玑 9400 发布， AI 能力更上一层楼	5
2、 AI 应用动态：KIMI 搜索访问量环比+32.36% ， 零一万物正式发布新旗舰模型 YI-LIGHTNING.....	7
2.1、 流量跟踪： Kimi 搜索访问量环比+32.36%.....	7
2.2、 产业动态：国产大模型竞技场首超 GPT-4o， 零一万物正式发布新旗舰模型 Yi-Lightning.....	8
3、 AI 融资动向：ABRIDGE/TERRAY THERAPEUTICS 分别获 2.5 亿美元/1.2 亿美元融资.....	12
4、 行情复盘.....	13
5、 投资建议.....	15
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： 天玑 9400 芯片	6
图表 3： 2024. 10. 14-2024. 10. 18 AI 相关网站流量.....	7
图表 4： Yi-Lightning 排名榜单.....	8
图表 5： Yi-Lightning 中文能力.....	9
图表 6： Yi-Lightning 数学能力.....	9
图表 7： Yi-Lightning 代码能力.....	10
图表 8： Hard Prompts 分榜.....	11
图表 9： Longer Query 分榜.....	11
图表 10： 本周 AI 初创公司的融资动态	12
图表 11： 本周指数日涨跌幅	13
图表 12： 本周 AI 算力指数内部涨跌幅度排名	13
图表 13： 本周 AI 应用指数内部涨跌幅度排名	14
图表 14： 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，安卓首款 3nm AI 芯片天玑 9400 发布

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 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	8.50/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	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：安卓首款 3nm AI 芯片天玑 9400 发布，AI 能力更上一层楼

安卓首款 3nm、PC 级 Arm V9 架构、第八代 NPU——天玑 9400 正式发布，比上一代生成式 AI 芯片天玑 9300，其 AI 能力更上一层楼，成为苏黎世 ETHZ 移动 SoC AI 性能榜单之首。具体提升如下：

1) 天玑 9400 首次实现了端侧 DiT 架构支持，无需联网就能在手机上体验 Sora 同款架构视频生成。还有业界首发的端侧 LoRA 训练，可以高效离线训练专属于自己的 LoRA 生图模型，同时保证隐私不泄露。

2) 大模型方面，天玑 9400 能够运行的模型窗口文本长度提升到了 32K，是天玑 9300 的 8 倍。同时还支持端侧运行多模态大模型，并以 50 Tokens 每秒的超高速度超越了前 SOTA。

3) 综合 AI 性能方面，天玑 9400 以 6773 分的成绩领跑苏黎世 ETHZ 移动 SoC 榜单，是天玑 9300 的 1.4 倍。

4) 从整体上看，天玑 9400 采用台积电第二代 3nm 制程，相较上一代单核性能提升 35%，多核性能提升 28%，同性能功耗降低 40%，

图表 2：天玑 9400 芯片



资料来源：量子位、华鑫证券研究

2、AI 应用动态：Kimi 搜索访问量环比 +32.36%，零一万物正式发布新旗舰模型 Yi-Lightning

2.1、流量跟踪：Kimi 搜索访问量环比+32.36%

本期（2024.10.14-2024.10.18）AI 相关网站流量数据：访问量前三位分别为 ChatGPT（816.0M）、Bing（408.4M）和 Canva（174.8M）；访问量环比增速前三位分别为 Kimi（32.36%）、文心一言（24.14%）和 QuillBot（5.22%）；平均停留时长前三位分别为 Discord（00:12:26）、Character.AI（00:12:11）和 DeepL（00:08:59）；平均停留时长环比增速前三位分别为 Kimi（2.20%）、文心一言（1.11%）和 Github（0.50%）。

图表 3：2024.10.14-2024.10.18 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	816.0	2.60%	6:11	0.00%
Bing	搜索	微软	408.4	0.44%	6:54	-0.24%
Discord	游戏社区	微软	126.2	-1.02%	12:26	0.27%
Canva	在线设计	Canva	174.8	0.75%	8:01	-0.82%
Github	代码托管	微软	118.2	2.16%	6:39	0.50%
Gemini	聊天机器人	谷歌	66.76	0.66%	4:55	-1.34%
Character.AI	聊天机器人	Character.AI	41.81	1.98%	12:11	-1.88%
NotionAI	文本/笔记	Notion	35.95	-2.44%	8:53	0.00%
QuillBot	释义工具	QuillBot	18.54	5.22%	4:07	-0.40%
Kimi	聊天机器人	Moonshot AI	7.334	32.36%	3:06	2.20%
DeepL	翻译工具	DeepL	48.76	0.85%	8:59	0.00%
文心一言	聊天机器人	百度	5.713	24.14%	3:02	1.11%
Perplexity	AI 搜索	Perplexity	18.80	4.97%	6:09	0.00%

资料来源：similarweb, 华鑫证券研究

2.2、产业动态：国产大模型竞技场首超 GPT-4o，零一万物正式发布新旗舰模型 Yi-Lightning

零一万物正式对外发布新旗舰模型——Yi-Lightning（闪电）。在大模型竞技场（Chatbot Arena）上，Yi-Lightning 性能直冲总榜单并列第 6，数学分榜并列第 3，代码等其它分榜也名列前茅。总成绩几乎与马斯克最新 xAI 大模型 Grok-2-08-13 持平，超越 GPT-4o-2024-05-13、GPT-4o-mini-2024-07-18、Claude 3.5 Sonnet 等顶流。在中文子榜单上，Yi-Lightning 超越了 xAI 发布的 Grok-2-08-13、智谱发布的 GLM-4-Plus 等国内外优质模型，与 o1-mini 等模型并列排名世界第二。

图表 4：Yi-Lightning 排名榜单

零一万物 Yi-Lightning: 进击全球前3大模型企业, 中国第一 零一万物
中国大模型发展里程碑: 首度公开超越 GPT4o, 中美AI全球竞技至今最佳成绩

Rank# (UB)	Rank (StyleCtrl)	Model	Arena Score	95% CI	Votes	Organization	License	Knowledge Cutoff
1	1	ChatGPT-4o-latest (2024-09-03)	1339	+4/-4	28488	OpenAI	Proprietary	2023/10
1	1	o1-preview	1335	+4/-5	17562	OpenAI	Proprietary	2023/10
3	3	o1-mini	1313	+4/-4	17919	OpenAI	Proprietary	2023/10
3	3	Gemini-1.5-Pro-002	1305	+5/-4	11430	Google	Proprietary	Unknown
4	3	Gemini-1.5-Pro-Exp-0027	1299	+4/-3	12437	Google	Proprietary	2023/11
6	8	Grok-2-08-13	1291	+3/-3	35661	xAI	Proprietary	2024/3
6	9	Yi-Lightning	1287	+5/-3	13262	01 AI	Proprietary	Unknown
7	5	GPT-4o-2024-05-13	1285	+3/-2	99251	OpenAI	Proprietary	2023/10
9	15	GLM-4-Plus	1274	+5/-5	13674	Zhipu AI	Proprietary	Unknown
9	17	GPT-4o-mini-2024-07-18	1274	+4/-3	38831	OpenAI	Proprietary	2023/10
9	13	Gemini-1.5-Flash-Exp-0027	1269	+3/-4	25555	Google	Proprietary	2023/11
9	20	Gemini-1.5-Flash-002	1269	+8/-5	8957	Google	Proprietary	Unknown
9	5	Claude 3.5 Sonnet	1268	+3/-3	75957	Anthropic	Proprietary	2024/4
9	26	Grok-2-Mini-08-13	1267	+3/-5	30597	xAI	Proprietary	2024/3
9	7	Meta-Llama-3.1-405B-Instruct-F16	1266	+5/-4	14696	Meta	Llama 3.1 Community	2023/12

LMSYS 盲测竞技场总榜 2024年10月16日

- 全球大模型企业排行第 3 -- 仅次于 OpenAI, Google, 打平 xAI Grok
- 全球大模型总榜排行第 6
- 中国大模型首度公开超越 GPT4o
- 中美 AI 全球竞技至今最佳成绩

(2024年10月16日)

资料来源：APPSO，华鑫证券研究

Yi-Lightning 采用 MoE 混合专家模型架构。底层技术上，Yi-Lightning 在以下方面进行了提升：

1) 优化混合注意力机制，只在模型的部分层次中将传统的全注意力（Full Attention）替换为滑动窗口注意力（Sliding Window Attention）。由此以来，模型在保证处理长序列数据高性能表现的同时，还能大大降低推理成本。Yi-Lightning 还引入了跨层注意力（Cross-Layer Attention, CLA），允许模型在不同的层次之间共享键（Key）和值（Value）头，减少对存储需求。这使得 Yi-Lightning 能在不同层次之间更有效地共享信息。总的来说，KV cache 缩小了 2-4 倍，同时将计算的复杂度从 $O(L^2)$ 降至 $O(L)$ 。”

2) 动态 Top-P 路由机制。也就是说，Yi-Lightning 可以根据任务的难度动态自动选择最合适的专家网络组合——训练过程中会激活所有专家网络，使模型能学习到所有专家知识；而推理阶段，根据任务的难度，模型会选择性地激活更匹配的专家网络。

3) 多阶段训练。团队将整个训练分为两块，一块做好以后就把它固定起来，然后在这个固定的模型上再做后段训练。训练前期，更注重数据多样性，使得 Yi-Lightning 尽可能

学习不同的知识；训练后期更重内容更丰富、知识性更强的数据。同时团队还在不同阶段采用不同的 batch size 和 LR schedule 保证训练速度和稳定性。李开复博士还表示，零一万物开发模型讲究“模基共建”，也就是共建模型和基础架构。模型的训练、服务、推理设计，与底层的 Allinfra 架构和模型结构必须高度适配。这样做的目的，不仅是让模型更好，而且让它在推理的时候能够更便宜。

细看大模型竞技场分类榜上的“赛况”，Yi-Lightning 各项能力都排在前头。

1) 在中文能力上，Yi-Lightning 和 GLM-4-Plus 两个国产大模型都名列前茅。Yi-Lightning 跃居并列第二，和 o1-mini 相差无几。

图表 5: Yi-Lightning 中文能力

Category		Apply filter		Chinese Prompts					
Chinese		<input type="checkbox"/> Style Control <input type="checkbox"/> Show Deprecated		#models: 129 (83%) #votes: 207,638 (10%)					
Rank* (UB)	Delta	Model	Arena Score	95% CI	Votes	Organization	License	Knowledge Cutoff	
1	0	ChatGPT-4o-latest (2024-09-03)	1423	+14/-11	2260	OpenAI	Proprietary	2023/10	
1	2	Gemini-1.5-Pro-002	1400	+20/-20	1006	Google	Proprietary	Unknown	
2	-1	o1-preview	1390	+16/-14	1464	OpenAI	Proprietary	2023/10	
2	2	Gemini-1.5-Pro-Exp-0827	1385	+13/-13	2377	Google	Proprietary	2023/11	
2	1	o1-mini	1375	+18/-17	1432	OpenAI	Proprietary	2023/10	
2	4	Yi-Lightning	1373	+21/-17	1240	01 AI	Proprietary	Unknown	
5	4	Gemini-1.5-Flash-Exp-0827	1353	+11/-14	2070	Google	Proprietary	2023/11	
5	1	Grok-2-08-13	1353	+12/-13	2778	xAI	Proprietary	2024/3	
5	4	GLM-4-Plus	1345	+16/-21	1157	Zhipu AI	Proprietary	Unknown	
5	4	Gemini-1.5-Flash-002	1341	+21/-17	794	Google	Proprietary	Unknown	
7	12	Deepseek-v2.5	1337	+16/-20	1135	DeepSeek	DeepSeek	Unknown	
7	0	GPT-4o-2024-05-13	1336	+8/-7	9961	OpenAI	Proprietary	2023/10	
7	3	Gemini Advanced App (2024-05-14)	1332	+10/-7	6074	Google	Proprietary	Online	
7	8	Qwen2.5-72b-Instruct	1328	+14/-16	1119	Alibaba	Qwen	2024/9	

资料来源：量子位，华鑫证券研究

2) 数学能力，Yi-Lightning 和 Gemini-1.5-Pro-002 并列第 3，仅次于 o1-preview、o1-mini。

图表 6: Yi-Lightning 数学能力

Category: **Math** | Apply filter: Style Control Show Deprecated | **Math** #models: 147 (94%) #votes: 275,859 (13%)

Rank* (UB)	Delta	Model	Arena Score	95% CI	Votes	Organization	License	Knowledge Cutoff
1	0	o1-preview	1347	+13/-12	2784	OpenAI	Proprietary	2023/10
1	2	o1-mini	1342	+9/-11	2964	OpenAI	Proprietary	2023/10
3	0	Gemini-1.5-Pro-002	1283	+11/-13	1723	Google	Proprietary	Unknown
3	3	Yi-Lightning	1278	+13/-12	2119	01 AI	Proprietary	Unknown
3	-2	ChatGPT-4o-latest (2024-09-03)	1277	+7/-9	4375	OpenAI	Proprietary	2023/10
3	6	Claude 3.5 Sonnet	1273	+7/-7	10549	Anthropic	Proprietary	2024/4
3	1	Gemini-1.5-Pro-Exp-0827	1273	+8/-9	5397	Google	Proprietary	2023/11
5	4	Meta-Llama-3.1-405b-Instruct-bf16	1255	+12/-11	2179	Meta	Llama 3.1 Community	2023/12
8	-1	GPT-4o-2024-05-13	1256	+7/-8	13307	OpenAI	Proprietary	2023/10
8	-2	Grok-2-08-13	1253	+11/-9	5072	xAI	Proprietary	2024/3
8	3	GPT-4o-2024-08-06	1253	+7/-10	4913	OpenAI	Proprietary	2023/10
8	7	Qwen2.5-72b-Instruct	1248	+12/-13	1966	Alibaba	Qwen	2024/9
8	2	Meta-Llama-3.1-405b-Instruct-fp8	1247	+8/-9	5612	Meta	Llama 3.1 Community	2023/12

资料来源：量子位，华鑫证券研究

3) 代码能力 Yi-Lightning 排名并列第 4。

图表 7: Yi-Lightning 代码能力

Category: **Coding** | Apply filter: Style Control Show Deprecated | **Coding: whether conversation contains code snippets** #models: 150 (96%) #votes: 400,720 (19%)

Rank* (UB)	Delta	Model	Arena Score	95% CI	Votes	Organization	License	Knowledge Cutoff
1	2	o1-mini	1373	+11/-11	3821	OpenAI	Proprietary	2023/10
1	0	o1-preview	1354	+11/-11	3431	OpenAI	Proprietary	2023/10
2	-1	ChatGPT-4o-latest (2024-09-03)	1338	+7/-9	5650	OpenAI	Proprietary	2023/10
4	2	Yi-Lightning	1300	+12/-14	2541	01 AI	Proprietary	Unknown
4	5	Claude 3.5 Sonnet	1295	+7/-6	14760	Anthropic	Proprietary	2024/4
4	3	GPT-4o-2024-05-13	1293	+6/-6	20711	OpenAI	Proprietary	2023/10
4	0	Gemini-1.5-Pro-Exp-0827	1287	+8/-9	6283	Google	Proprietary	2023/11
4	-1	Gemini-1.5-Pro-002	1287	+11/-14	2384	Google	Proprietary	Unknown
4	2	Grok-2-08-13	1287	+8/-5	7143	xAI	Proprietary	2024/3
4	5	Meta-Llama-3.1-405b-Instruct-bf16	1285	+8/-11	2818	Meta	Llama 3.1 Community	2023/12
4	5	GLM-4-Plus	1284	+9/-9	2836	Zhipu AI	Proprietary	Unknown
4	5	GPT-4o-mini-2024-07-18	1284	+7/-7	7436	OpenAI	Proprietary	2023/10
4	15	Deepseek-v2.5	1283	+9/-11	2809	DeepSeek	Proprietary	Unknown
4	11	Qwen2.5-72b-Instruct	1282	+9/-11	2810	Alibaba	Qwen	2024/9

资料来源：量子位，华鑫证券研究

4) 另外在 Hard Prompts 和 Longer Query 分榜, Yi-Lightning 也都排在第 4 位。

图表 8: Hard Prompts 分榜

Category		Apply filter		Hard Prompts (Overall): details in blog post					
Hard Prompts (Overall)		<input type="checkbox"/> Style Control	<input type="checkbox"/> Show Deprecated	#models: 155 (99%)		#votes: 565,664 (27%)			
Rank* (UB)	Delta	Model	Arena Score	95% CI	Votes	Organization	License	Knowledge Cutoff	
1	0	o1-preview	1357	+10/-10	4923	OpenAI	Proprietary	2023/10	
1	2	o1-mini	1347	+9/-9	5237	OpenAI	Proprietary	2023/10	
2	-1	ChatGPT-4o-latest...(2024-09-03)	1334	+7/-8	8087	OpenAI	Proprietary	2023/10	
4	2	Yi-Lightning	1298	+8/-9	3533	01 AI	Proprietary	Unknown	
4	-1	Gemini-1.5-Pro-002	1298	+9/-10	3259	Google	Proprietary	Unknown	
4	0	Gemini-1.5-Pro-Exp-0827	1288	+7/-8	9146	Google	Proprietary	2023/11	
5	2	GPT-4o-2024-05-13	1284	+5/-6	28836	OpenAI	Proprietary	2023/10	
6	0	Grok-2-08-13	1279	+6/-7	9811	xAI	Proprietary	2024/3	
6	3	Claude 3.5 Sonnet	1279	+4/-7	21350	Anthropic	Proprietary	2024/4	
6	13	Deepseek-v2.5	1273	+11/-9	4004	DeepSeek	DeepSeek	Unknown	
6	3	GLM-4-Plus	1273	+8/-9	3794	Zhipu AI	Proprietary	Unknown	
6	3	Meta-llama-3.1-405b-Instruct-bf16	1272	+8/-11	4044	Meta	Llama 3.1 Community	2023/12	
6	4	Yi-Lightning-lite	1271	+11/-10	3557	01 AI	Proprietary	Unknown	
6	5	Qwen-Max-0919	1271	+10/-10	3248	Alibaba	Qwen	Unknown	

资料来源: 量子位, 华鑫证券研究

图表 9: Longer Query 分榜

Rank* (UB)	Delta	Model	Arena Score	95% CI	Votes	Organization	License	Knowledge Cutoff
1	0	ChatGPT-4o-latest...(2024-09-03)	1365	+11/-9	3971	OpenAI	Proprietary	2023/10
1	0	o1-preview	1355	+11/-12	2554	OpenAI	Proprietary	2023/10
2	1	o1-mini	1335	+15/-13	2616	OpenAI	Proprietary	2023/10
3	1	Gemini-1.5-Pro-Exp-0827	1322	+10/-13	3661	Google	Proprietary	2023/11
3	0	Gemini-1.5-Pro-002	1317	+14/-13	1789	Google	Proprietary	Unknown
4	5	Gemini-1.5-Flash-Exp-0827	1310	+12/-12	3201	Google	Proprietary	2023/11
4	2	Yi-Lightning	1303	+12/-14	1849	01 AI	Proprietary	Unknown
4	3	GPT-4o-2024-05-13	1302	+8/-7	12818	OpenAI	Proprietary	2023/10
4	5	Gemini-1.5-Flash-002	1301	+20/-16	1439	Google	Proprietary	Unknown
4	7	Qwen-Max-0919	1297	+15/-12	1800	Alibaba	Qwen	Unknown
5	4	GPT-4o-mini-2024-07-18	1295	+10/-14	4798	OpenAI	Proprietary	2023/10
5	4	GLM-4-Plus	1292	+13/-11	2104	Zhipu AI	Proprietary	Unknown
6	11	Gemini-1.5-Pro-001	1295	+8/-8	11286	Google	Proprietary	2023/11
6	0	Grok-2-08-13	1293	+6/-11	4645	xAI	Proprietary	2024/3
6	4	Yi-Lightning-lite	1288	+15/-11	1867	01 AI	Proprietary	Unknown

资料来源: 量子位, 华鑫证券研究

谈起和国外头部大模型的差距，李开复博士表示这次 Yi-Lightning 的排名证明了国产大模型跟硅谷最顶尖模型的差距缩小到了五个月。去跟上美国最顶尖的模型，缩短这个时间差非常困难，要付出很大的努力和有独特的打法。在国内不少公司都在努力，“模基共建”则是零一万物自己摸索的独特路径。

3、AI 融资动向：Abridge/Terray Therapeutics 分别获 2.5 亿美元/1.2 亿美元融资

本期 AI 初创公司的融资中，Abridge/Terray Therapeutics 融资额位居前二，分别获 2.5 亿美元/1.2 亿美元融资。Abridge 的平台可通过移动应用程序访问，医生在每次预约前激活该应用，软件即可记录对话、转录并自动生成医疗记录。虽然这一过程并非完全自动化，医生仍需检查草稿的准确性并进行必要编辑，但 Abridge 提供了一项名为链接证据的功能，将每个 AI 生成的句子与相应的笔录摘录链接起来，从而简化了错误发现的过程。Terray 一直致力于通过规模实验和生成式 AI，提高小分子药物发现和开发的速度和成功率。依托其专有实验平台，Terray 构建了全球最大的化学数据集之一。目前 Terray 已定量测量了超过 50 亿次靶标-配体相互作用，该数据集仍在每年翻倍增长。Terray 的数据优势赋予其潜在 best-in-class 的 AI 能力，可识别并优化解决复杂问题的新型小分子。Terray 将这一专长应用于专注于免疫疾病的内部管线，同时在与 BMS 和 Calico 的合作中，针对更多具有挑战性的靶点开展研发。

图表 10：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
Gladia	AI 转录音频	XAnge	—	1600 万美元	—	—
Abridge	人工智能软件	Elad Gil、IVP	—	2.5 亿美元	—	25 亿美元
玻色量子	量子计算机技术	启赋资本	A 轮	数亿元	—	—

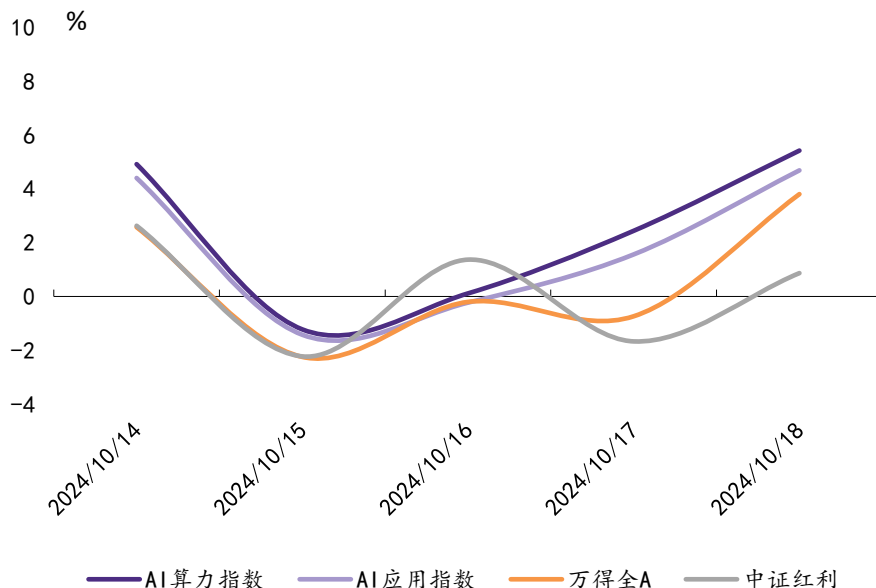
Terray Therapeutics	生成式 AI	英伟达风投部门	B 轮	1.2 亿美元	——	——
星动纪元	AI 大模型	清流资本、元璟资本、阿里巴巴	Pre-A 轮	3 亿元	——	——

资料来源：36kr，医药魔法，MSN，融中财经，华鑫证券研究

4、行情复盘

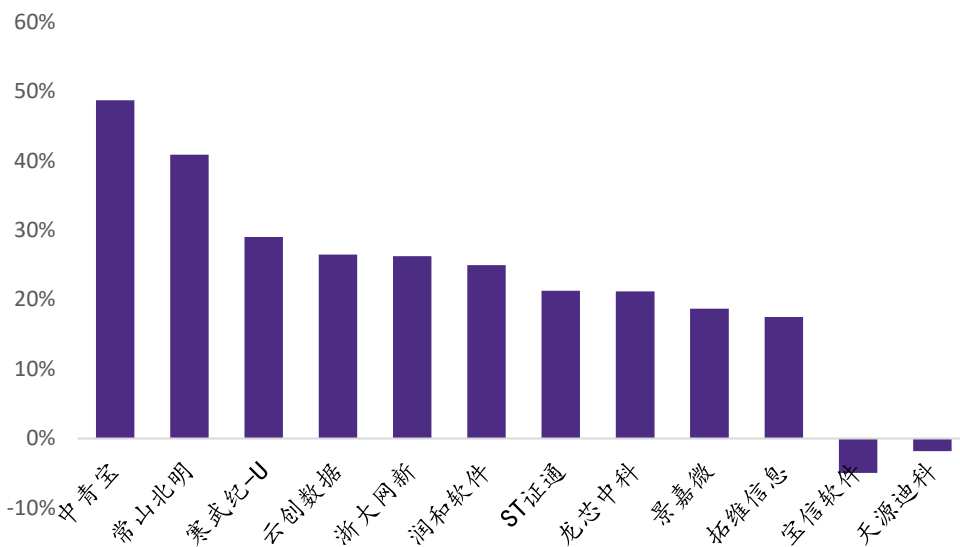
本周，AI 算力指数/AI 应用指数/万得全 A/中证红利日涨幅最大值分别为 5.42%/4.69%/3.80%/2.63%，日跌幅最大值分别为-1.22%/-1.42%/-2.25%/-2.23%。AI 算力指数内部，中青宝以+48.70%录得本周最大涨幅，宝信软件以-4.96%录得本周最大跌幅。AI 应用指数内部，光弘科技以+34.00%录得本周最大涨幅，传音控股以-5.50%录得本周最大跌幅。

图表 11：本周指数日涨跌幅



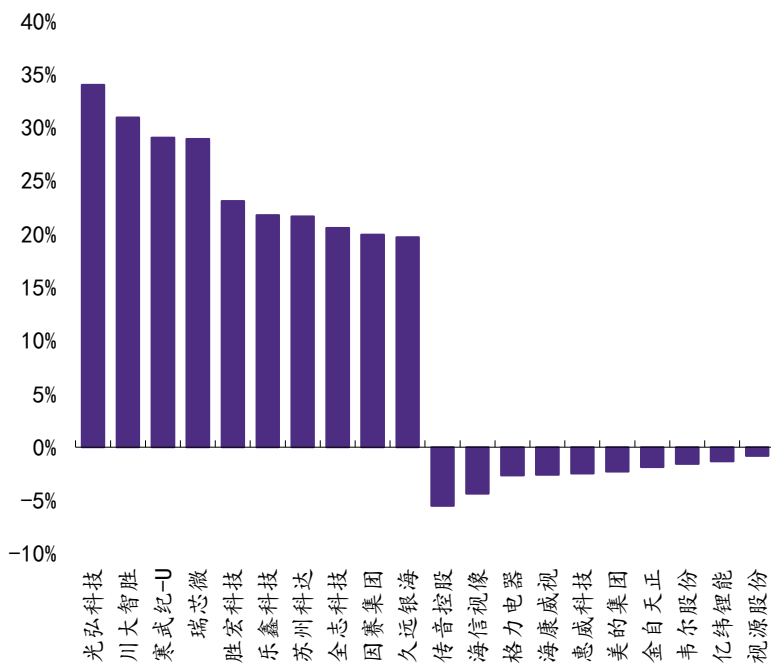
资料来源：wind，华鑫证券研究

图表 12：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 13: 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

一方面，国产 ai 大模型有望迎来新机遇，10 月 17 日，微软宣布将于 10 月 21 日停止在中国大陆的个人 Azure OpenAI 服务。OpenAI 停止服务后，多家国内大型企业，如阿里巴巴、百度与汇量科技迅速推出了针对这一市场变动的应对措施，力求吸引迁移流失客户，昆仑万维与讯飞等公司表明 OpenAI 用户的迁移需求在短时间内显著增加。另一方面，台积电 Q3 财报全面超预期提振市场信心，在截至 9 月 30 日的季度中，台积电营收录得 7596.9 亿新台币（235 亿美元），较上年同期增长了 39%（以美元计增长了 36%），超过了市场预期的 7421.66 亿元新台币，智能手机和人工智能对 3nm 和 5nm 技术的强劲需求支撑了其业务。

我们继续看好 AI 软硬件机会，近期优先关注铜缆高速连接板块。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 14：重点关注公司及盈利预测

公司代码	名称	2024-10-21 股价	EPS			PE			投资评级
			2023	2024E	2025E	2023	2024E	2025E	
002230.SZ	科大讯飞	45.83	0.28	0.40	0.56	163.68	114.58	81.84	买入
002368.SZ	太极股份	23.29	0.79	1.01	1.28	29.48	23.06	18.20	买入
688256.SH	寒武纪-U	429.71	-2.04	-1.21	-0.50	-210.64	-355.13	-859.42	买入
688668.SH	鼎通科技	43.93	0.67	1.04	1.41	65.57	42.24	31.16	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。

■ 免责条款

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。