证券研究报告

2025 年 04 月 08 日

华鑫证券 CHINA FORTUNE SECURITIES

Llama 4 多版本参数亮眼，DeepSeek 公布推理时 Scaling 新论文

—计算机行业周报

推荐(维持)

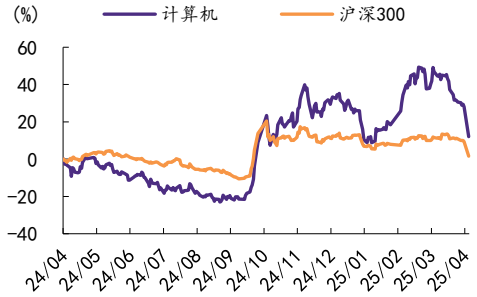
分析师：宝幼琛 S1050521110002

baoyc@cfsc.com.cn

行业相对表现

表现	1M	3M	12M
计算机(申万)	-23.8	0.7	12.5
沪深 300	-9.0	-5.3	1.5

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：阿里深夜开源 Qwen2.5-Omni，DeepSeek-V3 上线新版本》2025-04-05

2、《计算机行业点评报告：文远知行（WRD.O）：收入承压与商业化突破并行，自动驾驶长赛道静待拐点》2025-04-04

3、《计算机行业点评报告：禾赛科技（HSAI.O）：激光雷达龙头加速业绩兑现，多元布局打开成长空间》2025-04-04

投资要点

算力：Llama 4 多版本参数亮眼，2 万亿多模态巨兽重登王座

Meta 官宣开源首个原生多模态 Llama 4，首次采用 MoE 架构，支持 12 种语言，首批发布一共两款：第一款是 Llama 4 Scout，规模较小，其共有 1090 亿参数，17B 活跃参数，16 个专家，1000 万上下文；第二款是 Llama 4 Maverick，规模较大，其共有 4000 亿参数，17B 活跃参数，128 个专家，100 万上下文。

在大模型 LMSYS 排行榜上，Llama 4 Maverick 冲上第二（ELO 得分 1417），仅次于闭源 Gemini 2.5 Pro。Llama 4 Scout 最大亮点在于支持 1000 万上下文，相当于可以处理 20+小时的视频，仅在单个 H100 GPU（Int4 量化后）上就能跑。

在基准测试中，性能超越 Gemma 3、Gemini 2.0 Flash-Lite、Mistral 3.1。Llama 4 模型是 Llama 系列模型中首批采用混合专家（MoE）架构的模型。在 MoE 模型中，单独的 token 只会激活全部参数中的一小部分。与传统的稠密模型相比，MoE 架构在训练和推理时的计算效率更高，并且在相同的训练 FLOPs 预算下，能够生成更高质量的结果。

Llama 4 是一个原生多模态模型，采用了早期融合技术，能把文本和视觉 token 无缝整合到一个统一的模型框架里。早期融合是个大进步，因为它可以用海量的无标签文本、图片和视频数据一起来预训练模型。

Meta 还开发了一种叫做 MetaP 的新训练方法，能让他们更精确地设置关键的模型超参数，比如每层的学习率和初始化规模。这些精心挑选的超参数在不同的批大小、模型宽度、深度和训练 token 量上都能很好地适配。Llama 4 通过在 200 种语言上预训练实现了对开源微调的支持，其中超过 10 亿个 token 的语言有 100 多种，整体多语言 token 量比 Llama 3 多出 10 倍。

AI 应用：Gemini 搜索访问量环比+9.62%，DeepSeek 公布推理时 Scaling 新论文

近期，来自 DeepSeek、清华大学的研究人员探索了奖励模型（RM）的不同方法，发现逐点生成奖励模型（GRM）可以统一

行业研究

纯语言表示中单个、成对和多个响应的评分。基于这一初步成果，论文的作者提出了一种新学习方法，即自我原则批评调整（SPCT），以促进 GRM 中有效的推理时间可扩展行为。通过利用基于规则的在线 RL，SPCT 使 GRM 能够学习根据输入查询和响应自适应地提出原则和批评，从而在一般领域获得更好的结果奖励。

基于此技术，DeepSeek 提出了 DeepSeek-GRM-27B，它基于 Gemma-2-27B 用 SPCT 进行后训练。对于推理时间扩展，它通过多次采样来扩展计算使用量。通过并行采样，DeepSeek-GRM 可以生成不同的原则集和相应的批评，然后投票选出最终的奖励。通过更大规模的采样，DeepSeek-GRM 可以更准确地判断具有更高多样性的原则，并以更细的粒度输出奖励，从而解决挑战。

除了投票以获得更好的扩展性能外，DeepSeek 还训练了一个元 RM。从实验结果上看，SPCT 显著提高了 GRM 的质量和可扩展性，在多个综合 RM 基准测试中优于现有方法和模型，且没有严重的领域偏差。作者还将 DeepSeek-GRM-27B 的推理时间扩展性能与多达 671B 个参数的较大模型进行了比较，发现它在模型大小上可以获得比训练时间扩展更好的性能。虽然当前方法在效率和特定任务方面面临挑战，但凭借 SPCT 之外的努力，DeepSeek 相信，具有增强可扩展性和效率的 GRM 可以作为通用奖励系统的多功能接口，推动 LLM 后训练和推理的前沿发展。

AI 融资动向：星海图“小步快跑式”融资，今年估值已翻倍

4 月 3 日，星海图宣布接连完成 A2、A3 轮系列融资，领投方为凯辉基金，总融资额超 3 亿元人民币。这意味着 2025 年以来星海图已累计融资近 1 亿美元。

星海图本次 A2、A3 轮系列融资由凯辉基金领投，联想创投、海尔资本等产业资本参投，老股东 IDG 资本、高瓴创投、百度风投、同歌创投等追投，其中部分老股东多轮满额、超额持续加注。

星海图 A1 轮融资于今年 2 月完成，总融资额近 3 亿元，由蚂蚁集团独家领投，高瓴创投、IDG 资本、北京机器人产业基金、百度风投、同歌创投等老股东追加投资。由此可见，星海图于 2025 年展开的 A 轮系列累计融资总额已达约 1 亿美元。

星海图介绍，投资人最关注的是公司全栈要素齐备且实力较强的特点。具身智能产品的成功不只靠模型，而是底层零部件、整机设计及制造、场景理解能力等的系统性能力。公司创始团队具有业内领先的模型技术实力和产业落地经验，硬件能力也在过去一年里快速补齐。星海图目前已成为国内极少数同时具备端到端 AI 算法能力、全链路正向研发制造能力

以及实际商业化验证能力的具身智能公司之一。星海图若估值达到 50 亿元，将成为业内第二梯队的“排头兵”。

投资建议

4 月 8 日消息，美国时间周一，白宫发布指令，要求联邦各机构任命首席人工智能官，并制定扩大政府人工智能应用的战略。备忘录还指示各机构在六个月内“制定人工智能战略，识别并消除负责任使用该技术的障碍，并实现全机构范围内的提升应用成熟度。我们仍然坚定认为，AI 应用有望在今年诞生部份现象级应用。建议关注临床 AI 产品成功落地验证的嘉和美康（688246.SH）、以 AI 为核心的龙头厂商科大讯飞（002230.SZ）、芯片技术有望创新突破的寒武纪（688256.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）、已与 Rokid 等多家知名 AI 眼镜厂商建立紧密合作的亿道信息（001314.SZ）、加快扩张算力业务的精密零部件龙头迈信林（688685.SH）、持续加码高速铜缆的泓淋电力（301439.SZ）、新能源业务高增并供货科尔摩根等全球电机巨头的唯科科技（301196.SZ）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。

重点关注公司及盈利预测

公司代码	名称	2025-04-08 股价	EPS			PE			投资评级
			2023	2024E	2025E	2023	2024E	2025E	
001314.SZ	亿道信息	42.01	0.91	0.92	1.03	46.16	45.66	40.79	买入
002230.SZ	科大讯飞	43.39	0.28	0.40	0.56	154.96	108.48	77.48	买入
301196.SZ	唯科科技	41.59	1.35	1.66	2.02	30.81	25.05	20.59	买入
301439.SZ	泓淋电力	12.27	0.55	0.57	0.66	22.31	21.53	18.59	买入
688246.SH	嘉和美康	27.79	0.31	0.56	0.77	89.65	49.63	36.09	买入
688256.SH	寒武纪-U	565.16	-2.04	-1.21	-0.50	-277.04	-467.07	-1130.32	买入
688668.SH	鼎通科技	35.40	0.67	1.04	1.41	52.84	34.04	25.11	买入
688685.SH	迈信林	43.88	0.14	0.41	1.34	313.43	107.02	32.75	买入

资料来源：Wind，华鑫证券研究

正文目录

图表目录

1、 算力动态：算力租赁价格平稳，Llama 4 多版本参数亮眼

1.1、 数据跟踪：算力租赁价格平稳

本周算力租赁价格保持平稳。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7.50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	7.50/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：Llama 4 多版本参数亮眼，2万亿多模态巨兽重登王座

Meta官宣开源首个原生多模态Llama 4，首次采用MoE架构，支持12种语言，首批发布一共两款：第一款是Llama 4 Scout，规模较小，其共有1090亿参数，17B活跃参数，16个专家，1000万上下文；第二款是Llama 4 Maverick，规模较大，其共有4000亿参数，17B活跃参数，128个专家，100万上下文。

在大模型LMSYS排行榜上，Llama 4 Maverick冲上第二（ELO得分1417），仅次于闭源Gemini 2.5 Pro。Llama 4 Scout最大亮点在于支持1000万上下文，相当于可以处理20+小时的视频，仅在单个H100 GPU（Int4量化后）上就能跑。

在基准测试中，性能超越Gemma 3、Gemini 2.0 Flash-Lite、Mistral 3.1。Llama 4模型是Llama系列模型中首批采用混合专家（MoE）架构的模型。在MoE模型中，单独的token只会激活全部参数中的一小部分。与传统的稠密模型相比，MoE架构在训练和推理时的计算效率更高，并且在相同的训练FLOPs预算下，能够生成更高质量的结果。

Llama 4是一个原生多模态模型，采用了早期融合技术，能把文本和视觉token无缝整合到一个统一的模型框架里。早期融合是个大进步，因为它可以用海量的无标签文本、图片和视频数据一起来预训练模型。

Meta还开发了一种叫做MetaP的新训练方法，能让他们更精确地设置关键的模型超参

数，比如每层的学习率和初始化规模。这些精心挑选的超参数在不同的批大小、模型宽度、深度和训练 token 量上都能很好地适配。Llama 4 通过在 200 种语言上预训练实现了对开源微调的支持，其中超过 10 亿个 token 的语言有 100 多种，整体多语言 token 量比 Llama 3 多出 10 倍。

图表 2：Llama 4 Maverick instruction-tuned benchmarks & Llama 4 Scout instruction-tuned benchmarks

Llama 4 Maverick instruction-tuned benchmarks

Category / Benchmark	Llama 4 Maverick	Gemini 2.0 Flash	DeepSeek v3.1	GPT-4o
Inference Cost — Cost per 1M input & output tokens (3:1 blended)	$0.19-$0.49[5]	$0.17	$0.48	$4.38
Image Reasoning — MMMU	73.4	71.7	No multimodal support	69.1
MathVista	73.7	73.1		63.8
Image Understanding — ChartQA	90.0	88.3		85.7
DocVQA (test)	94.4	—		92.8
Coding — LiveCodeBench (10/01/2024-02/01/2025)	43.4	34.5	45.8/49.2[3]	32.3[3]
Reasoning & Knowledge — MMLU Pro	80.5	77.6	81.2	—
GPQA Diamond	69.8	60.1	68.4	53.6
Multilingual — Multilingual MMLU	84.6	—	—	81.5
Long Context — MTOB (half book) eng → kgv/kgv → eng	54.0/46.4	48.4/39.8[4]	Context window is 128K	Context window is 128K
MTOB (full book) eng → kgv/kgv → eng	50.8/46.7	45.5/39.6[4]		

1. For Llama model results, we report 0 shot evaluation with temperature = 0 and no majority voting or parallel test time compute. For high-variance benchmarks (GPQA Diamond, LiveCodeBench), we average over multiple generations to reduce uncertainty.
2. For non-Llama models, we source the highest available self-reported eval results, unless otherwise specified. We only include evals from models that have reproducible evals (via API or open weights), and we only include non-thinking models. Cost estimates are sourced from Artificial Analysis for non-Llama models.
3. DeepSeek v3.1's date range is unknown (49.2), so we provide our internal result (45.8) on the defined date range. Results for GPT-4o are sourced from the LCB leaderboard.
4. Specialized long context evals are not traditionally reported for generalist models, so we share internal runs to showcase Llama's frontier performance.
5. $0.19/Mtok (3:1 blended) is our cost estimate for Llama 4 Maverick assuming distributed inference. On a single host, we project the model can be served at $0.30-$0.49/Mtok (3:1 blended).

Llama 4 Scout instruction-tuned benchmarks

Category / Benchmark	Llama 4 Scout	Llama 3.3 70B	Llama 3.1 405B	Gemma 3 27B	Mistral 3.1 24B	Gemini 2.0 Flash-Lite
Image Reasoning — MMMU	69.4	No multimodal support	No multimodal support	64.9	62.8	68.0
MathVista	70.7			67.6	68.9	57.6
Image Understanding — ChartQA	88.8			76.3	86.2	73.0
DocVQA (test)	94.4			90.4	94.1	91.2
Coding — LiveCodeBench (10/01/2024-02/01/2025)	32.8	33.3	27.7	29.7	—	28.9
Reasoning & Knowledge — MMLU Pro	74.3	68.9	73.4	67.5	66.8	71.6
GPQA Diamond	57.2	50.5	49.0	42.4	46.0	51.5
Long Context — MTOB (half book) eng → kgv/kgv → eng	42.2/36.6	Context window is 128K	Context window is 128K	Context window is 128K	Context window is 128K	42.3/35.1[3]
MTOB (full book) eng → kgv/kgv → eng	39.7/36.3					35.1/30.0[3]

1. For Llama model results, we report 0 shot evaluation with temperature = 0 and no majority voting or parallel test time compute. For high-variance benchmarks (GPQA Diamond, LiveCodeBench), we average over multiple generations to reduce uncertainty.
2. For non-Llama models, we source the highest available self-reported eval results unless otherwise specified. We only include evals from models that have reproducible evals (via API or open weights), and we only include non-thinking models.
3. Specialized long context evals are not traditionally reported for generalist models, so we share internal runs to showcase Llama's frontier performance.

资料来源：新智元，华鑫证券研究

2、AI 应用动态：Gemini 搜索访问量环比 +9.62%，DeepSeek 公布推理时 Scaling 新论文

2.1、流量跟踪：Gemini 搜索访问量环比 +9.62%

本期（2025.3.31-2025.4.4）AI 相关网站流量数据：访问量前三位分别为 ChatGPT（1102.0M）、Bing（332.4M）和 Canva（181.7M），访问量环比增速第一为 Gemini（9.62%）；平均停留时长前三位分别为 Character.AI（00:17:04）、Discord（00:11:46）和 NotionAI（00:9:05）;平均停留时长环比增速第一为文心一言（4.13%）。

图表 3：2025.3.31-2025.4.4 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	1102.0	8.46%	6:50	0.74%
Bing	搜索	微软	332.4	-0.63%	6:33	-2.00%
Discord	游戏社区	微软	131.8	1.38%	11:46	-0.14%
Canva	在线设计	Canva	181.7	0.55%	7:09	-0.69%
Github	代码托管	微软	116.3	-0.94%	6:32	0.00%
Gemini	聊天机器人	谷歌	88.58	9.62%	4:59	0.67%
Character.AI	聊天机器人	Character.AI	43.65	-1.11%	17:04	-0.29%
NotionAI	文本/笔记	Notion	36.28	1.34%	9:05	-0.37%
QuillBot	释义工具	QuillBot	14.38	-1.37%	4:09	2.05%
Kimi	聊天机器人	Moonshot AI	8.05	-8.96%	3:54	-0.85%
Deepl	翻译工具	DeepL	45.6	-0.39%	8:34	0.19%
文心一言	聊天机器人	百度	2.604	-1.21%	3:47	4.13%
Perplexity	AI 搜索	Perplexity	27.43	-2.18%	6:24	-0.52%

资料来源：similarweb,华鑫证券研究

2.2、产业动态：DeepSeek 公布推理时 Scaling 新论文

近期，来自 DeepSeek、清华大学的研究人员探索了奖励模型（RM）的不同方法，发现逐点生成奖励模型（GRM）可以统一纯语言表示中单个、成对和多个响应的评分。基于这一初步成果，论文的作者提出了一种新学习方法，即自我原则批评调整（SPCT），以促进 GRM 中有效的推理时间可扩展行为。通过利用基于规则的在线 RL，SPCT 使 GRM 能够学习根据输入查询和响应自适应地提出原则和批评，从而在一般领域获得更好的结果奖励。

基于此技术，DeepSeek 提出了 DeepSeek-GRM-27B，它基于 Gemma-2-27B 用 SPCT 进行后训练。对于推理时间扩展，它通过多次采样来扩展计算使用量。通过并行采样，DeepSeek-GRM 可以生成不同的原则集和相应的批评，然后投票选出最终的奖励。通过更大规模的采样，DeepSeek-GRM 可以更准确地判断具有更高多样性的原则，并以更细的粒度输

出奖励，从而解决挑战。

除了投票以获得更好的扩展性能外，DeepSeek 还训练了一个元 RM。从实验结果上看，SPCT 显著提高了 GRM 的质量和可扩展性，在多个综合 RM 基准测试中优于现有方法和模型，且没有严重的领域偏差。作者还将 DeepSeek-GRM-27B 的推理时间扩展性能与多达 671B 个参数的较大模型进行了比较，发现它在模型大小上可以获得比训练时间扩展更好的性能。虽然当前方法在效率和特定任务方面面临挑战，但凭借 SPCT 之外的努力，DeepSeek 相信，具有增强可扩展性和效率的 GRM 可以作为通用奖励系统的多功能接口，推动 LLM 后训练和推理的前沿发展。

如下图所示，SPCT 包含两个阶段：1. 拒绝式微调（rejective fine-tuning），作为冷启动阶段；2. 基于规则的在线强化学习（rule-based online RL），通过不断优化生成的准则和评论，进一步增强泛化型奖励生成能力。此外，SPCT 还能促使奖励模型在推理阶段展现出良好的扩展能力。

图表 4：SPCT 的两个阶段

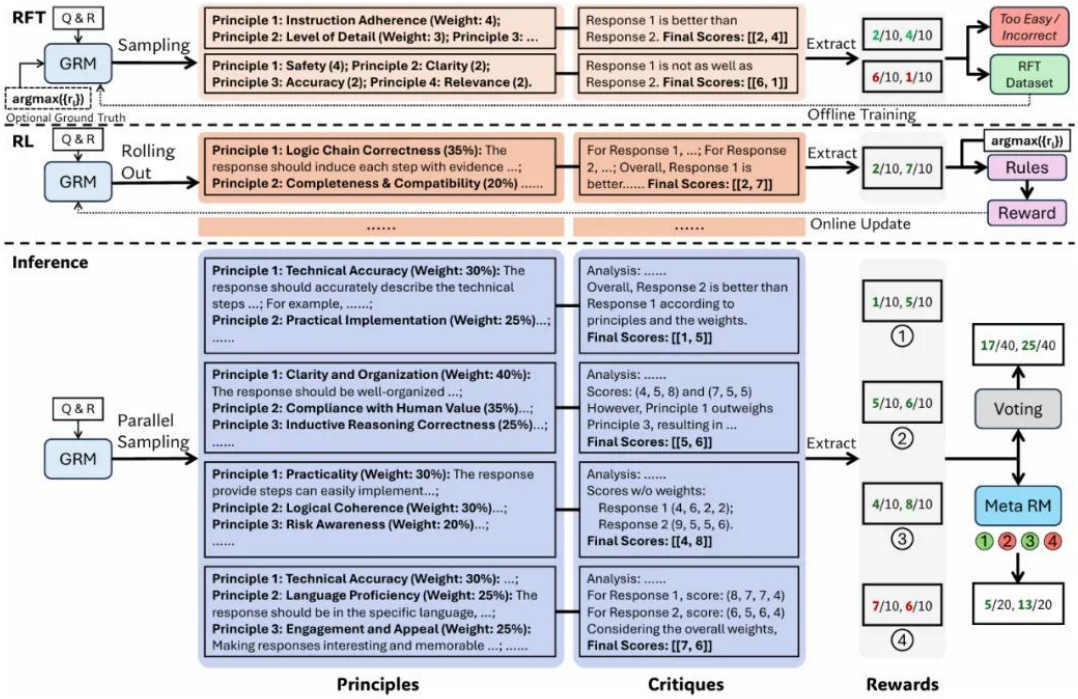


Figure 3: Illustration of SPCT, including rejective fine-tuning, rule-based RL, and corresponding scalable behaviors during inference. The inference-time scaling is achieved via naive voting or meta RM guided voting with principles generated at scale, resulting in finer-grained outcome rewards within a expanded value space.

资料来源：机器之心，华鑫证券研究

为了进一步提高 DeepSeek-GRM 在使用更多推理计算生成通用奖励方面的性能，研究者探索了基于采样的策略，以实现有效的推理时可扩展性。DeepSeek-GRM 的投票过程需要多次采样，由于随机性或模型的局限性，少数生成的原则和评论可能存在偏差或质量不高。因此，研究者训练了一个元 RM 来指导投票过程。引导投票非常简单：元 RM 对 k 个采样奖励输出元奖励，最终结果由 k_meta ≤ k 个元奖励的奖励投票决定，从而过滤掉低质量样本。

图表 5：不同方法和模型在奖励模型基准测试上的整体结果

Model	Reward Bench	PPE Preference	PPE Correctness	RMB	Overall
Reported Results of Public Models					
Skywork-Reward-Gemma-2-27B	94.1	56.6	56.6	60.2	66.9
DeepSeek-V2.5-0905	81.5	62.8	58.5	65.7	67.1
Gemini-1.5-Pro	86.8	66.1	59.8	56.5	67.3
ArmoRM-8B-v0.1	90.4	60.6	61.2	64.6	69.2
InternLM2-20B-Reward	90.2	61.0	63.0	62.9	69.3
LLaMA-3.1-70b-Instruct	84.1	65.3	59.2	68.9	69.4
Claude-3.5-sonnet	84.2	65.3	58.8	70.6	69.7
Nemotron-4-340B-Reward	92.0	59.3	60.8	69.9	70.5
GPT-4o	86.7	67.1	57.6	73.8	71.3
Reproduced Results of Baseline Methods					
LLM-as-a-Judge	83.4	64.2	58.8	64.8	67.8
DeepSeek-BTRM-27B	81.7	**68.3**	**66.7**	57.9	68.6
CLoud-Gemma-2-27B	82.0	67.1	62.4	63.4	68.7
DeepSeek-PairRM-27B	87.1	65.8	64.8	58.2	69.0
Results of Our Method					
DeepSeek-GRM-27B-RFT (Ours)	84.5	64.1	59.6	67.0	68.8
DeepSeek-GRM-27B (Ours)	86.0	64.7	59.8	69.0	69.9
Results of Inference-Time Scaling (Voting@32)					
DeepSeek-GRM-27B (Ours)	88.5	65.3	60.4	69.0	71.0
DeepSeek-GRM-27B (MetaRM) (Ours)	**90.4**	67.2	63.2	**70.3**	**72.8**

Table 2: Overall results of different methods and models on RM benchmarks. Underlined numbers indicate the best performance, **bold numbers** indicate the best performance among baseline and our methods, and *italicized font* denotes scalar or semi-scalar RMs. For meta RM guided voting (MetaRM), $k_{\text{meta}} = \frac{1}{2}k$.

资料来源：机器之心，华鑫证券研究

3、AI 融资动向：星海图“小步快跑式”融资，今年估值已翻倍

4 月 3 日，星海图宣布接连完成 A2、A3 轮系列融资，领投方为凯辉基金，总融资额超 3 亿元人民币。这意味着 2025 年以来星海图已累计融资近 1 亿美元。

星海图本次 A2、A3 轮系列融资由凯辉基金领投，联想创投、海尔资本等产业资本参投，老股东 IDG 资本、高瓴创投、百度风投、同歌创投等追投，其中部分老股东多轮满额、超额持续加注。

星海图 A1 轮融资于今年 2 月完成，总融资额近 3 亿元，由蚂蚁集团独家领投，高瓴创投、IDG 资本、北京机器人产业基金、百度风投、同歌创投等老股东追加投资。

由此可见，星海图于 2025 年展开的 A 轮系列累计融资总额已达约 1 亿美元。

星海图介绍，投资人最关注的是公司全栈要素齐备且实力较强的特点。具身智能产品的成功不只靠模型，而是底层零部件、整机设计及制造、场景理解能力等的系统性能力。公司创始团队具有业内领先的模型技术实力和产业落地经验，硬件能力也在过去一年里快速补齐。星海图目前已成为国内极少数同时具备端到端 AI 算法能力、全链路正向研发制造能力以及实际商业化验证能力的具身智能公司之一。

记者了解到，星海图若估值达到 50 亿元，将成为业内第二梯队的“排头兵”。

图表 6：本周 AI 初创公司融资动态

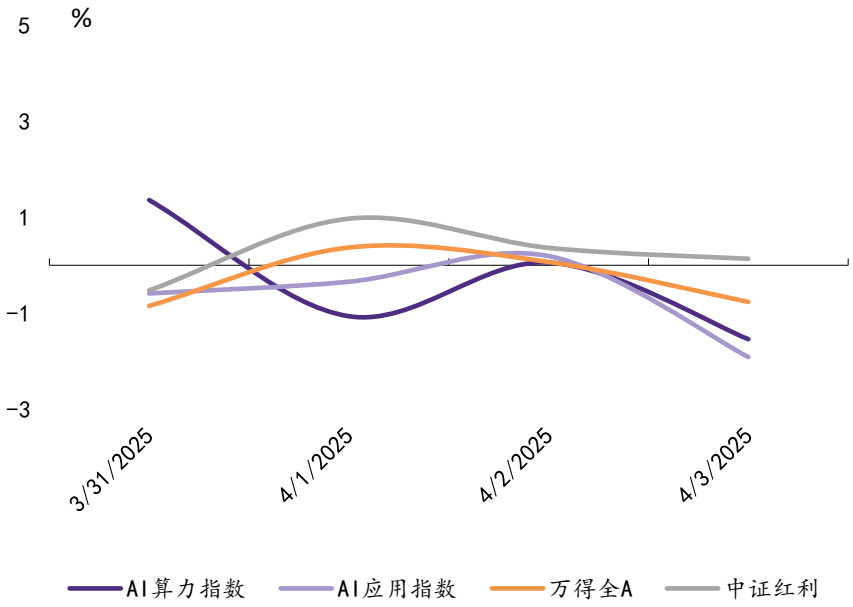
应用	应用类型	领投方	融资轮	融资额	目前累计融资额	目前估值
星海图	Embodied AI	凯辉基金，联想创投、海尔资本，IDG 资本	A3	超 3 亿元		超 50 亿元
原力灵机	AI 工业	君联资本、九坤创投、启明创投	天使轮	2 亿元		
它石智航	AI 大模型	蓝驰创投、启明创投，线性资本、恒旭资本、洪泰基金、联想创投	天使轮	8.7 亿		

资料来源：上海证券报，财经杂志，AI 科技评论，华鑫证券研究

4、 行情复盘

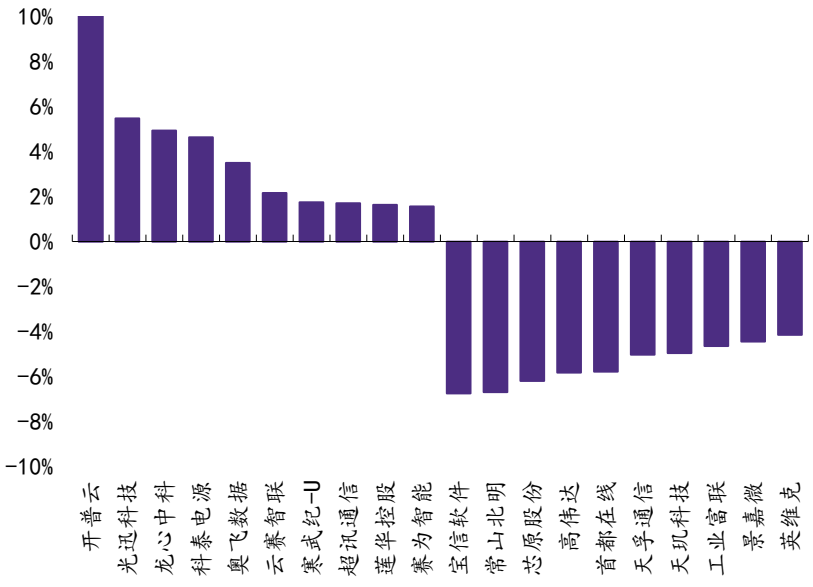
本周，AI 算力指数/AI 应用指数/万得全 A/中证红利日涨幅最大值分别为 1.36%/0.19%/0.37%/0.97%， AI 算力指数/AI 应用指数/万得全 A/中证红利日跌幅最大值分别为-1.54%/-1.91%/-0.85%/-0.52%。 AI 算力指数内部，开普云以+10.52%录得本周最大涨幅，宝信软件以-6.75%录得本周最大跌幅。 AI 应用指数内部，海信视像以+6.99%得本周最大涨幅，国光电器以-15.03%录得本周最大跌幅。

图表 7：本周指数日涨跌幅



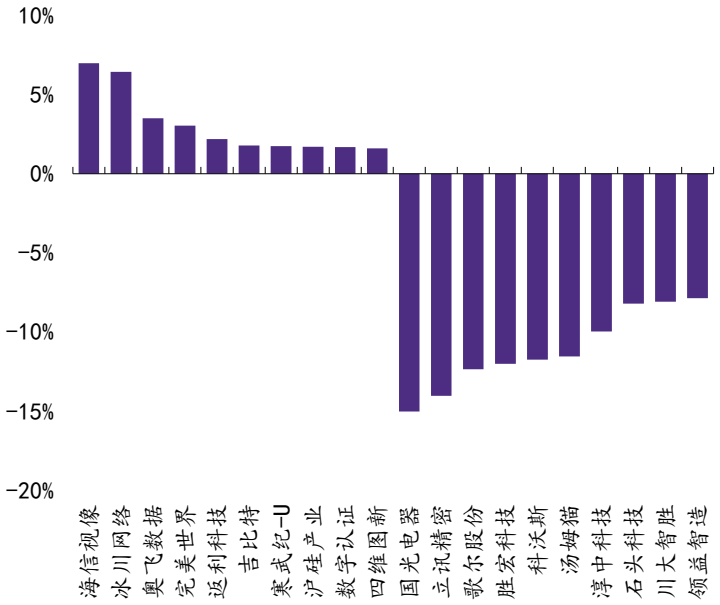
资料来源：wind，华鑫证券研究

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

4 月 8 日消息，美国时间周一，白宫发布指令，要求联邦各机构任命首席人工智能官，并制定扩大政府人工智能应用的战略。备忘录还指示各机构在六个月内“制定人工智能战略，识别并消除负责任使用该技术的障碍，并实现全机构范围内的提升应用成熟度。我们仍然坚定认为，AI 应用有望在今年诞生部份现象级应用。建议关注临床 AI 产品成功落地验证的嘉和美康（688246.SH）、以 AI 为核心的龙头厂商科大讯飞（002230.SZ）、芯片技术有望创新突破的寒武纪（688256.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）、已与 Rokid 等多家知名 AI 眼镜厂商建立紧密合作的亿道信息（001314.SZ）、加快扩张算力业务的精密零部件龙头迈信林（688685.SH）、持续加码高速铜缆的泓淋电力（301439.SZ）、新能源业务高增并供货科尔摩根等全球电机巨头的唯科科技（301196.SZ）等。

图表 10：重点关注公司及盈利预测

公司代码	名称	2025-04-03	EPS			PE			投资评级
		股价	2023	2024E	2025E	2023	2024E	2025E	2023
001314.SZ	亿道信息	47.97	0.91	0.92	1.03	001314.SZ	亿道信息	47.97	0.91
002230.SZ	科大讯飞	48.25	0.28	0.40	0.56	002230.SZ	科大讯飞	48.25	0.28
301196.SZ	唯科科技	53.95	1.35	1.66	2.02	301196.SZ	唯科科技	53.95	1.35
301439.SZ	泓淋电力	15.83	0.55	0.57	0.66	301439.SZ	泓淋电力	15.83	0.55
688246.SH	嘉和美康	33.38	0.31	0.56	0.77	688246.SH	嘉和美康	33.38	0.31
688256.SH	寒武纪-U	629.35	-2.04	-1.21	-0.50	688256.SH	寒武纪-U	629.35	-2.04
688668.SH	鼎通科技	45.19	0.67	1.04	1.41	688668.SH	鼎通科技	45.19	0.67
688685.SH	迈信林	53.50	0.14	0.41	1.34	688685.SH	迈信林	53.50	0.14

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&AI&互联网组介绍

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。

谢孟津：伦敦政治经济学院硕士，2023 年加入华鑫证券。

费强：曼彻斯特大学硕士，2023 年加入华鑫证券研究所。

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。