

2026年03月25日



华鑫证券  
CHINA FORTUNE SECURITIES

# GPT-5.4mini 与 nano 轻量登场，SkyReels-V4 登顶全球 AI 视频大模型榜首

—计算机行业周报

## 推荐(维持)

## 投资要点

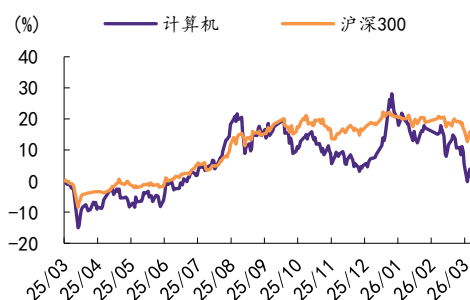
分析师：任春阳 S1050521110006

✉ rency@cfsc.com.cn

### 行业相对表现

表现	1M	3M	12M
计算机(申万)	-9.9	-2.8	3.9
沪深300	-4.2	-2.3	15.4

### 市场表现



资料来源：Wind，华鑫证券研究

### 相关研究

- 1、《(Netflix)：25Q4 业绩稳健兑现，内容与广告双轮驱动，平台化变现能力持续增强》2026-03-23
- 2、《微软 (Microsoft)：云与 AI 维持高景气，Azure 持续强势》2026-03-22
- 3、《计算机行业周报：OpenClaw 连续版本更新，英伟达联合谢赛宁团队发布 TMD 框架》2026-03-17

■ **算力：算力租赁价格平稳，SkyReels-V4 登顶全球 AI 视频大模型榜首**

2026年3月，昆仑万维旗下 SkyReels-V4AI 视频大模型，在业内权威第三方评测平台 ArtificialAnalysis 的文生视频全球榜单中登顶第一，超越 OpenAISora2、GoogleVeo3.1 等国际顶尖模型。该模型完成全模态强化学习体系升级，新增多帧参考等核心能力，大幅提升长视频生成稳定性。

■ **AI 应用：Discord 周访问量环比+2.09%，GPT-5.4mini 与 nano 轻量登场**

2026年3月18日，OpenAI 推出了两款全新轻量化模型——GPT-5.4mini 与 GPT-5.4nano。此次发布聚焦于真实生产环境中的实际需求，即在保证模型响应速度与输出质量的前提下，显著降低使用成本。两款模型继承了 GPT-5.4 的核心能力，但在运行效率和经济性方面做出了更为极致的优化。

■ **AI 融资动向：Deeptune 完成 4300 万美元 A 轮融资**

2026年3月，AI 智能体模拟训练环境平台 Deeptune 完成 4300 万美元 A 轮融资，由 a16z 领投，多家机构及 OpenAI 研究员等天使投资人参投，其 2023 年曾完成 310 万美元种子轮融资。该公司打造高保真强化学习模拟环境，还原各类职业工作场景，让 AI 通过交互式学习提升能力，成立半年便实现百万美元级 ARR。

■ **投资建议**

2026年3月，英伟达于美国加州圣何塞举办 GTC 大会，同时 OFC2026 光通信大会于洛杉矶同期召开，两大展会均释放 CPO 技术发展重要信号，英伟达相关布局及行业技术落地、共识进一步明确，核心内容如下：英伟达上调 2025-2027 年 AI 芯片营收预期至 1 万亿美元，同步了发布 VeraRubin 平台及 RubinUltra 机柜，并规划 2028 年量产 Feynman 架构芯片。CPO 领域，英伟达推出了 SN6800 等三款量产 CPO 交换机，Feynman 的路线图中也加上了 Nvlink8Cpo，黄仁勋明确表态“我们需要更多的 CPO 产能”，2028 年则有望作为 CPO 大规模落地关键节点。同期 OFC2026 大会上，英伟达进一步明确硅光 CPO 为下一代 AI 集群核心技术解，已基于 TSMCCOUPE 工艺打造光电协同封装，攻克微环调制器热敏感性难题并将探

索 DWDM 方案提升传输能力；同时该展会中 CPO 技术落地全面提速，博通、Marvell、Coherent 等企业分别发布 CPO 演进路线、光集成架构、三级集成定义及落地路径，英伟达联合 AMD 等六家头部企业成立 OCI-MSA 联盟，为 CPO 规模化落地制定统一开放标准。本次 GTC 和 OFC 大会的召开再次印证了 AI 算力板块的景气度，其中 CPO 的逐步商业化进程持续提速、发展路径愈发清晰。结合当前 AI 算力需求爆发、CPO 技术持续突破及产业链成熟度提升的现状，我们维持对 CPO 板块的看好。

中长期，建议关注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），AI 智能文字识别与商业大数据领域巨头的合合信息（688615.SH）、深耕工业 AI 与软件并长期服务高端装备等领域头部客户的能科科技（603859.SH）、专注于半导体等高端制造业的罗博特科（300757.SZ）。

### ■ 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	2026-03-25 股价	EPS			PE			投资评级
			2024	2025E	2026E	2024	2025E	2026E	
300757.SZ	罗博特科	365.23	0.41	-0.30	0.30	890.80	-1217.43	1217.43	买入
301196.SZ	唯科科技	110.16	1.76	2.53	3.34	62.59	43.54	32.98	买入
603859.SH	能科科技	37.60	0.78	0.96	1.18	48.21	39.17	31.86	买入
688615.SH	合合信息	174.88		3.24	4.22		53.98	41.44	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，SKYREELS-V4 登顶全球 AI 视频大模型榜首 .....	4
1.1、Tokens 跟踪.....	4
1.2、数据跟踪：算力租赁价格平稳 .....	5
1.3、产业动态：SkyReels-V4 登顶全球 AI 视频大模型榜首 .....	5
2、AI 应用动态：GITHUB 周访问量环比+2.67%，GPT-5.4MINI 与 NANO 轻量登场 .....	8
2.1、周流量跟踪：Github 周访问量环比+2.67%.....	8
2.2、产业动态：GPT-5.4mini 与 nano 轻量登场，AI 能力正以低成本方式加速普及.....	8
3、AI 融资动向：DEEPTUNE 完成 4300 万美元 A 轮融资.....	12
4、行情复盘 .....	14
5、投资建议 .....	16
6、风险提示 .....	16

## 图表目录

图表 1：TOKENS 规模 LEADERBOARD .....	4
图表 2：市场份额占据示意 .....	4
图表 3：上周算力租赁情况 .....	5
图表 4：ARTIFICIALANALYSIS 榜单示意图.....	5
图表 5：PROMPT：根据@图片-1 中的动漫情节生成的动画短片.....	6
表 6：MMDIT 结构示意图 .....	6
图表 7：2026.3.14-2026.3.20AI 相关网站流量.....	8
图表 8：PT-5.4MINI、GPT-5.4NANO 与 GPT-5.4 在多项评测中的结果对比.....	9
图表 9：PT-5.4MINI 与 GPT-5.4NANO 在多种工具调用测试中的表现.....	9
图表 10：PT-5.4MINI 与 GPT-5.4NANO 在 OSWORLD-VERIFIED 的表现.....	10
图表 11：PT-5.4MINI、GPT-5.4NANO 与 GPT-5.4 在定价方面的对比.....	10
图表 12：PT-5.4MINI 与 GPT-5.4NANO 在长上下文处理方面的表现.....	11
图表 13：上周 AI 初创公司融资动态 .....	12
图表 14：上周（2026.3.16-2026.3.20 日）指数日涨跌幅.....	14
图表 15：上周（2026.3.16-2026.3.20 日）AI 算力指数内部涨跌幅度排名 .....	14
图表 16：上周（2026.3.16-2026.3.20 日）AI 应用指数内部涨跌幅度排名 .....	15
图表 17：重点关注公司及盈利预测 .....	16

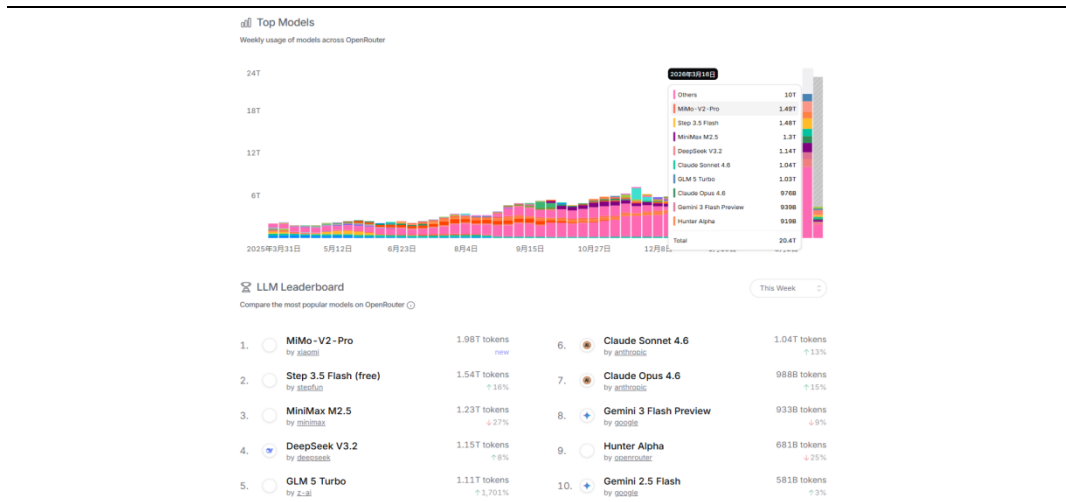
# 1、算力动态：算力租赁价格平稳，SkyReels-V4 登顶全球 AI 视频大模型榜首

## 1.1、Tokens 跟踪

根据 OpenRouter 公开数据，2026 年 3 月 16 日至 2026 年 3 月 22 日，周度 token 消耗量有所上升，调用量为 20.4T，环比上周 20.71%。在 tokens 规模 leaderboard 前五名中，xiaomi 旗下的 MiMo-V2-Pro 以 1.49Ttokens 位居榜首；stepfun 的 Step3.5Flash 的以 1.48T 位列第二；minimax 旗下的 MiniMax2.5 以 1.3Ttokens 位居第三；Deepseek 的 DeepseekV3.2 以 1.14Ttokens 位列第四；Anthropic 的 ClaudeSonnet 4.6 以 1.04Ttokens 位列第五。

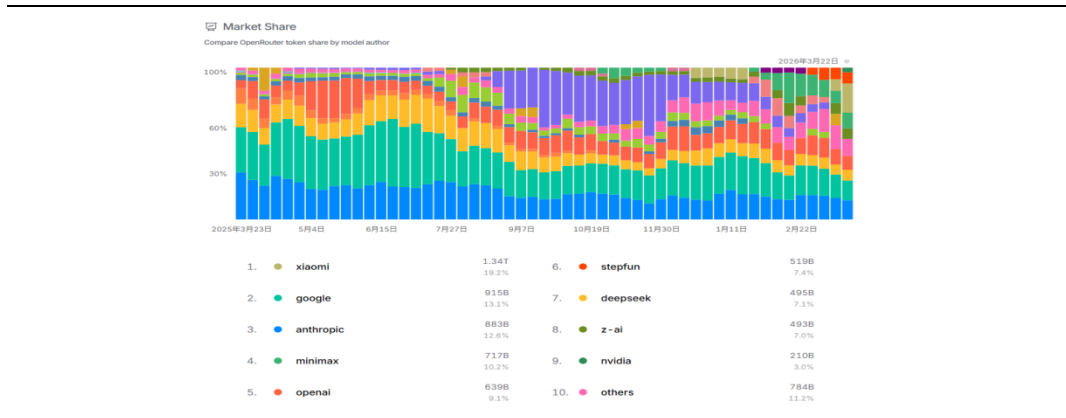
从市场份额维度来看，xiaomi 以 1.34Ttokens 占据 19.2% 的份额，稳居首位；google 以 915B 占据 13.1%，位列第二；Anthropic、Minimax、OpenAI 则分别以 883B、717B、639Btokens，对应占据 12.6%、10.2%、9.1% 的市场份额。

图表 1：tokens 规模 leaderboard



资料来源：OpenRouter，华鑫证券研究

图表 2：市场份额占据示意



资料来源：OpenRouter，华鑫证券研究

## 1.2、数据跟踪：算力租赁价格平稳

上周算力租赁价格平稳。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；A100-80G 价格为 5.66 元，A800-80G 价格为 6.80 元/时。

图表 3：上周算力租赁情况

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	5.66/元	0.00%
A800-80G	16	256	系统盘: 20G 数据盘: 50GB	恒源云	6.80/元	0.00%

资料来源：腾讯云，阿里云，恒源云，华鑫证券研究

## 1.3、产业动态：SkyReels-V4 登顶全球 AI 视频大模型榜首

2026 年 3 月，昆仑万维旗下 SkyworkAI 研发的 SkyReels-V4AI 视频大模型，在权威第三方评测平台 ArtificialAnalysis 的文生视频全球排行榜中登顶全球第一，超越 OpenAISora2、GoogleVeo3.1 等一众国际顶尖模型，成为当前全球 AI 视频生成能力最强的大模型。此前在 2026 年 2 月，该模型的 Preview 版本已在同平台榜单中拿下全球第二的成绩，实现了不到一个月内的排名跃升。

图表 4：ArtificialAnalysis 榜单示意图

Current models	All models	No Audio	With Audio	All	Open weights	Global Leaderboard	Personal Leaderboard
Creator	Model	ELO	95% CI	Samples	Released	API Pricing	
1   Skywork AI	SkyReels V4	1,129	-14/14	2,324	Mar 2026	\$8.40 /min	
2   KlingAI	Kling 3.0 1080p (Pro)	1,097	-10/10	4,589	Feb 2026	\$20.16 /min	
3   KlingAI	Kling 3.0 720p (Standard)	1,090	-10/10	4,470	Feb 2026	\$15.12 /min	
4   KlingAI	Kling 3.0 Omni 1080p (Pro)	1,090	-10/10	4,325	Feb 2026	\$16.80 /min	
5   Google	Veo 3.1 Fast	1,087	-11/11	3,838	Jan 2026	\$9.00 /min	
6   KlingAI	Kling 3.0 Omni 720p (Standard)	1,084	-10/10	4,166	Feb 2026	\$13.44 /min	
7   Vidu	Vidu Q3 Pro	1,083	-9/9	5,131	Jan 2026	\$9.60 /min	
8   Google	Veo 3.1	1,081	-10/10	4,745	Jan 2026	\$24.00 /min	
9   xAI	grok-imagine-video	1,081	-9/9	5,040	Jan 2026	\$4.20 /min	
10   OpenAI	Sora 2 (December)	1,079	-12/12	3,377	Dec 2025	\$6.00 /min	

资料来源：机器之心，华鑫证券研究

相较于上一代版本，SkyReels-V4 通过全模态强化学习大幅提升了模型的语义理解和整体逻辑能力，新增多帧参考与网格参考两项高阶任务，重点补强了角色一致性和长叙事视频的生成稳定性。该模型可稳定完成多镜头复杂场景的视频生成，支持首帧参考的图生视频、多图片参考的多镜头短剧生成，可实现多语言台词的口型精准同步，同时支持动作迁移、视频编辑等功能，覆盖局部添加、区域删除、去台词、去水印等常见应用场景。其新增的多帧或网格图参考能力，可基于用户给定的关键帧精准推演逻辑严密、动作连贯的中间画面，适配短剧工业化生产的核心需求。

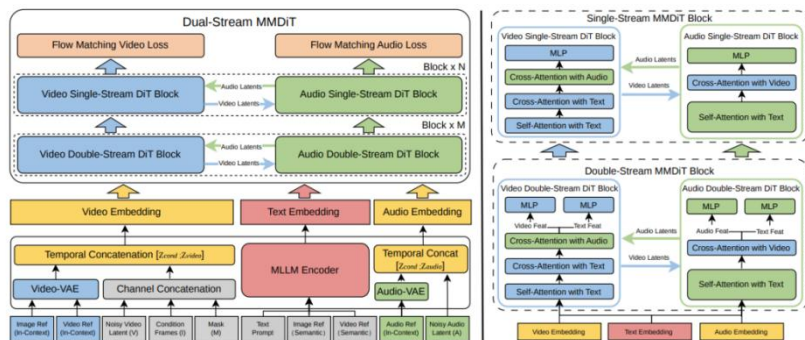
图表 5: Prompt: 根据@图片-1 中的动漫情节生成的动画短片



资料来源：机器之心，华鑫证券研究

技术层面，SkyReels-V4 实现了两大核心突破，为性能表现提供了底层支撑。其一为全模态强化学习体系的全面升级，针对传统扩散模型重局部像素生成、轻整体语义逻辑与叙事连贯性的行业痛点，模型搭建了完整的强化学习体系，一方面构建覆盖文生视频、图生视频、视频编辑、音视频对齐全场景的全模态语义 Reward 模型，为生成提供全局精准的实时反馈，另一方面采用阶梯式课程强化学习路径，从分辨率与时长、任务复杂度、数据难度三个维度循序渐进，让模型由简入繁掌握复杂能力，最终实现 1080p、15 秒商用长序列生成，以及多任务大一统的能力框架，同时该体系也为模型带来了极强的跨任务泛化性。其二是新增关键帧参考与九宫格参考两大高阶参考任务，全面提升视频生成的稳定性与灵活性，直接解决了短剧生成中角色走形、场景跳跃的行业痛点。

表 6: MMDiT 结构示意图



资料来源：机器之心，华鑫证券研究

在架构设计上，SkyReels-V4 采用了创新的 MMDiT 结构，该结构包含两个并行分支，分别负责视频合成与音频生成，模型在每个 Transformer 块中都嵌入了双向跨注意力机制，确保音画同步效果，同时模型共享基于多模态大语言模型的强力文本编码器，可理解包含文本、参考图、参考视频片段在内的复杂组合指令。研发团队还通过统一的通道连接公式，将图像转视频、视频扩展及编辑任务建模为不同配置下的视频补全问题，让模型可在同一界面下灵活处理多种生成 workflow；模型引入带偏移的 3DRoPE，解决了不同时域分辨率的音视频对齐问题，可精准捕捉人物身份特征或复杂运动轨迹。针对高分辨率长视频生成的计算挑战，模型采用基础模型生成低分辨率完整序列和高分辨率关键帧、再由专门模块进行细节重塑的方案，其帧插值模块引入的视频稀疏注意力机制，在维持硬件运行效率的同时，将注意力计算成本降低了约 3 倍。实验结果显示，SkyReels-V4 在多项公开榜单中取得了当前最优成绩，在面向音视频综合素质的人类评估中，其指令遵循能力、运动质量以及多镜头叙事连贯性，均超越了部分知名闭源商业系统。

目前，SkyReels-V4 的能力已覆盖短剧、影视、广告、数字人等主流内容形态，统一的生成接口可实现同一套基础设施服务不同生产场景，无需为单独需求搭建专属技术栈。昆仑万维已将 SkyReels-V4 直接嵌入自身的内容生产体系，自 2024 年 12 月宣布布局海外短剧以来，旗下付费平台 DramaWave 与免费平台 FreeReels 已携手站稳海外短剧第一梯队，MAU 突破 8000 万，月流水超 4000 万美金，其中 DramaWave 平台漫剧模块上线数月累计上架近千部 AI 剧，凭借千万级月活与数百万美金的月收入，占据海外 AI 剧市场龙头地位。

## 2、AI 应用动态：Discord 周访问量环比 +2.09%，GPT-5.4mini 与 nano 轻量登场

### 2.1、周流量跟踪：Discord 周访问量环比+2.09%

本期（2026.3.14-2026.3.20）AI 相关网站流量数据：访问量前三位分别为 ChatGPT（1272.0M）、Bing（809.5M）和 Gemini（575.5M），访问量环比增速第一为 Discord（2.09%）；平均停留时长前三位分别为 Character.AI（00:17:12）、Discord（00:10:51）和 Kimi（00:08:41）；平均停留时长环比增速第一为 Canva（0.57%）。

图表 7：2026.3.14-2026.3.20AI 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	1272.0	-2.90%	5:53	-1.12%
Bing	搜索	微软	809.5	-1.32%	7:59	-0.21%
Gemini	聊天机器人	谷歌	575.5	-0.19%	7:17	-1.13%
Canva	在线设计	Canva	216.3	-4.38%	5:54	0.57%
Github	代码托管	微软	143.3	0.56%	6:39	0.50%
Discord	游戏社区	微软	136.8	2.09%	10:51	0.00%
Character.AI	聊天机器人	Character.AI	43.82	-1.97%	17:12	-0.48%
NotionAI	文本/笔记	Notion	40.83	-1.40%	7:58	-0.21%
Perplexity	AI 搜索	Perplexity	37.21	-2.59%	4:43	-
DeepL	翻译工具	DeepL	28.04	-1.68%	2:24	-
QuillBot	释义工具	QuillBot	10.97	-3.18%	2:46	0.00%
Kimi	聊天机器人	MoonshotAI	10.59	-9.72%	8:41	-0.95%
文心一言	聊天机器人	百度	0.66	1.92%	2:46	-

资料来源：similarweb, 华鑫证券研究

### 2.2、产业动态：GPT-5.4mini 与 nano 轻量登场，AI 能力正以低成本方式加速普及

2026年3月18日，OpenAI 推出了两款全新轻量化模型——GPT-5.4mini 与 GPT-5.4nano。此次发布聚焦于真实生产环境中的实际需求，即在保证模型响应速度与输出质量的前提下，显著降低使用成本。两款模型继承了 GPT-5.4 的核心能力，但在运行效率和经济性方面做出了更为极致的优化。

从关键性能指标来看，GPT-5.4mini 在多项评测中展现出与旗舰版 GPT-5.4 极为接近的能力水平。在 SWE-BenchPro 基准测试中，mini 版本取得了 54.4% 的成绩，与满血版的 57.7% 仅相差 3.3 个百分点。相比之下，上一代 mini 版本在该项测试中的得分为 45.7%，新一代模型在编码任务上的提升幅度接近 9%。在 Terminal-Bench2.0 测试中，GPT-5.4mini 更是以 60.0% 对 38.2% 的显著优势超越了前代产品。即使是定位更为轻量的 nano 版本，在 SWE-BenchPro 上也获得了 52.4% 的分数，高于上一代 mini 版本的表现，表明轻量化模型在解决真实工程问题上的能力正在迅速逼近旗舰模型，为开发者在代码修改、前端生成等场景中提供了更具效率的选择。

图表 8：PT-5.4mini、GPT-5.4nano 与 GPT-5.4 在多项评测中的结果对比

	OpenAI			Anthropic	Google
	GPT-5.4	GPT-5.4 mini	GPT-5.4 nano	Claude Haiku 4.5	Gemini 3 Flash
SWE-Bench Pro (Public) Software engineering	57.7%	54.4%	52.4%	—	—
Terminal-Bench 2.0 Terminal-based coding	75.1%	60.0%	46.3%	41.0%	47.60%
OSWorld-Verified Computer use	75.0%	72.1%	39.0%	50.7%	—
MCP Atlas Tool calling	67.2%	57.7%	56.1%	34.6%	57.4%
GPQA Diamond Expert scientific reasoning (no tools)	93.0%	88.0%	82.8%	73.0%	90.4%

All OpenAI models run with maximum available reasoning effort.  
Non-OpenAI evals are publicly reported by respective companies or leaderboards managed by the eval providers.

资料来源：新智元，华鑫证券研究

在推理与工具调用方面，GPT-5.4mini 同样表现出色。在博士级科学推理基准 GPQADiamond 中，其得分达到 88%，与旗舰版的差距仅为 5%。在工具调用能力的评估中，GPT-5.4mini 在 Toolathlon 基准上取得 42.9% 的得分，远超 GPT-5mini 的 26.9%。在电信行业专用基准 τ2-bench 上，mini 版本的得分高达 93.4%，几乎与旗舰版的 98.9% 持平。这些结果说明，GPT-5.4mini 不仅具备较强的推理能力，更能在生产环境中独立完成复杂任务链。

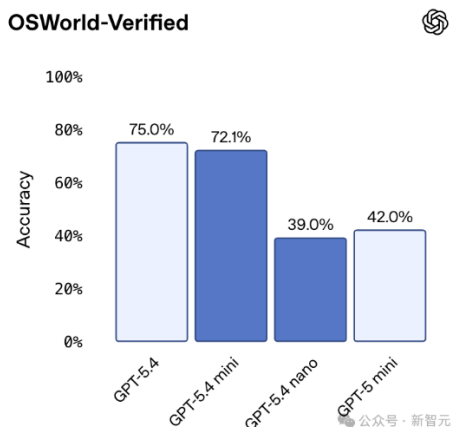
图表 9：PT-5.4mini 与 GPT-5.4nano 在多种工具调用测试中的表现

Tool-calling	GPT-5.4 (xhigh)	GPT-5.4 mini (xhigh)	GPT-5.4 nano (xhigh)	GPT-5 mini (high)
MCP Atlas	67.2%	57.7%	56.1%	47.6%
Toolathlon	54.6%	42.9%	35.5%	26.9%
τ2-bench (telecom)	98.9%	93.4%	92.5%	74.1%

资料来源：新智元，华鑫证券研究

值得关注的是，GPT-5.4mini 在计算机使用能力方面也取得了突破性进展。在衡量模型对 UI 界面的视觉理解与操作能力的 OSWorld-Verified 基准中，mini 版本取得了 72.1% 的得分，而旗舰版 GPT-5.4 为 75.0%，两者差距不足三个百分点。相比之下，GPT-5mini 在该项测试中的得分仅为 42.0%。不过，nano 版本在该项测试中得分为 39.0%，略低于上一代 mini 版本，显示出计算机使用任务对模型能力有较高门槛，并非所有轻量模型均能胜任。

图表 10: PT-5.4mini 与 GPT-5.4nano 在 OSWorld-Verified 的表现



资料来源: 新智元, 华鑫证券研究

在模型架构与使用策略方面, OpenAI 提出了明确的“大模型决策, 小模型执行”协作思路。旗舰模型 GPT-5.4 负责整体规划、协调与最终决策, 而 GPT-5.4mini 则作为子智能体, 承担需要快速完成的具体执行任务, 如代码库搜索、文件审查、支持文档处理等。在 Codex 平台中, 这一架构已实现落地, mini 在执行任务时仅消耗旗舰模型 30% 的资源配额, 在成本与效率之间取得更优平衡。

在定价与可用性方面, GPT-5.4mini 的输入价格为每百万 token0.75 美元, 输出价格为每百万 token4.50 美元, 支持 400k 的上下文窗口, 具备文本与图像输入、工具使用、函数调用、网络搜索、文件搜索及计算机使用等完整功能。GPT-5.4nano 则仅在 API 中提供, 输入价格为每百万 token0.20 美元, 输出价格为每百万 token1.25 美元, 其输入价格约为 mini 的四分之一, 输出价格不足 mini 的三分之一, 在高频低复杂度的分类、数据提取等任务中具备极高的性价比。在 ChatGPT 端, GPT-5.4mini 已面向免费用户和 Go 用户开放, 当付费用户的旗舰版额度用尽时, mini 将作为降级备选方案自动启用。

图表 11: PT-5.4mini、GPT-5.4nano 与 GPT-5.4 在定价方面的对比

	GPT-5.4	GPT-5.4 mini	GPT-5.4 nano
Reasoning	★★★★★	★★★★	★★★
Speed	★★★★	★★★★	★★★★
Input	📄 📄 📄 📄	📄 📄 📄 📄	📄 📄 📄 📄
Output	📄 📄 📄 📄	📄 📄 📄 📄	📄 📄 📄 📄
Reasoning tokens	●	●	●
PRICING	PER 1M TOKENS	PRICING	PER 1M TOKENS
Input	\$2.50	Input	\$0.75
Cached Input	\$0.25	Cached Input	\$0.08
Output	\$15.00	Output	\$4.50

资料来源: 新智元, 华鑫证券研究

当然，GPT-5.4mini 在长上下文处理方面仍存在明显短板。在 MRCRv2 测试中，针对 64K 至 128K 窗口下的八针搜索任务，旗舰版得分 86.0%，而 mini 仅为 47.7%；在 128K 至 256K 窗口下，两者差距进一步扩大至 79.3%对 33.6%。在 Graphwalks 系列测试中，父节点追踪任务上旗舰版得分 89.8%，mini 为 71.5%。这说明在需要对超长文本进行精确信息检索和逻辑追踪的场景中，mini 的能力上限仍显著低于旗舰版，旗舰模型在长文档分析、长对话记忆保持等任务中依然不可替代。

图表 12: PT-5.4mini 与 GPT-5.4nano 在长上下文处理方面的表现

Long context				
	GPT-5.4 (xhigh)	GPT-5.4 mini (xhigh)	GPT-5.4 nano (xhigh)	GPT-5 mini (high')
OpenAI MRCR v2 8-needle 64K-128K	86.0%	47.7%	44.2%	35.1%
OpenAI MRCR v2 8-needle 128K-256K	79.3%	33.6%	33.1%	19.4%
Graphwalks BFS 0K-128K	93.1%	76.3%	73.4%	73.4%
Graphwalks parents 0- 128K (accuracy)	89.8%	71.5%	50.8%	64.3%

资料来源：新智元，华鑫证券研究

### 3、AI 融资动向：Deeptune 完成 4300 万美元 A 轮融资

2026 年 3 月，AI 智能体模拟强化学习环境平台 Deeptune 宣布完成 4300 万美元 A 轮融资，本轮融资由 AndreessenHorowitz (a16z) 领投，776、AbstractVentures、InspiredCapital 参投，OpenAI 研究员 NoamBrown 等行业资深人士也以天使投资人身份入局。Deeptune 成立于 2022 年 8 月，此前曾在 2023 年 3 月完成 310 万美元种子轮融资，本轮新筹资金将全部用于 AI 智能体训练“训练场”的搭建与迭代。目前公司团队约 20 人，核心成员来自 Anthropic、ScaleAI 等头部企业，成立后仅六个月便实现百万美元级年度经常性收入 (ARR)，计划在后续六个月将该数字提升十倍。

Deeptune 的核心技术创新在于打造了高保真的 AI 智能体强化学习模拟环境，为 AI 智能体提供可交互的模拟器，破解行业高质量训练数据枯竭的核心痛点。不同于传统 AI 依赖静态标注数据的训练模式，Deeptune 的模拟环境可高度还原会计师、客户支持代表、DevOps 工程师等职业的日常工作流程，让 AI 智能体通过交互式操作学习多步骤复杂工作任务，在动态环境中通过行动、反馈完成能力迭代，而非仅依赖书本化的教程数据。

目前 Deeptune 已为全球领先的 AI 实验室搭建了数百个专属训练环境，其技术已助力 AI 智能体实现从简单问答到真实软件多步骤 workflow 执行的能力跃升，覆盖视频编辑、金融模型搭建等多元场景。此次融资不仅印证了资本市场对 AI 训练新范式的高度认可，更标志着 AI 训练正从静态网络数据抓取，转向合成交互式环境的强化学习模式。随着高质量公共训练数据资源逐渐见顶，Deeptune 的技术路线为 AI 智能体的企业级落地提供了核心支撑，也为 AI 训练基础设施赛道开辟了全新的增长空间。

图表 13：上周 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
Deeptune	AI 训练基础设施	AndreessenHorowitz (a16z)	A 轮	4300 万美元	4610 万美元	未披露
Replit	AI 技术服务	Georgian	D 轮	4 亿美元	4 亿美元	90 亿美元

---

Science	脑机接口	Lightspeed Venture Partners 等	C 轮	2.3 亿美元	4.9 亿美元	15 亿美元
---------	------	-------------------------------	-----	---------	---------	--------

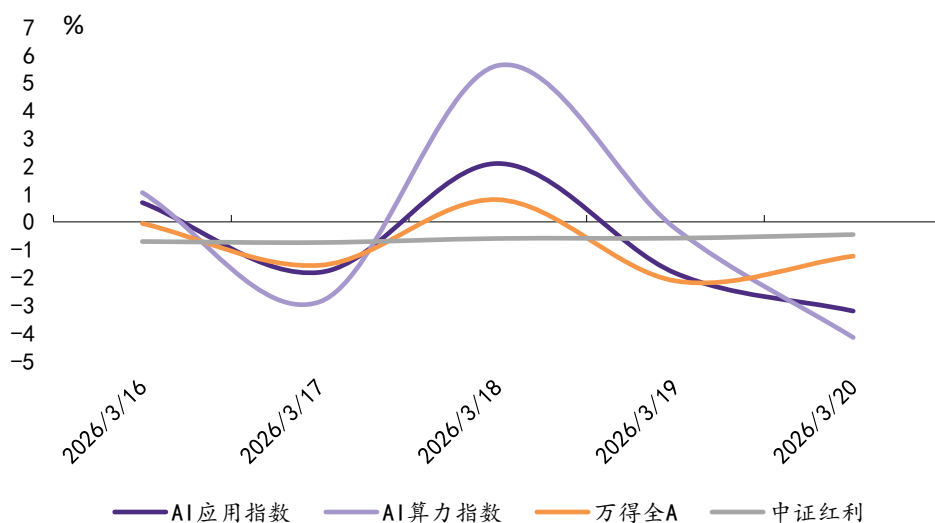
---

资料来源: wind, Saasverse, 华鑫证券研究

## 4、行情复盘

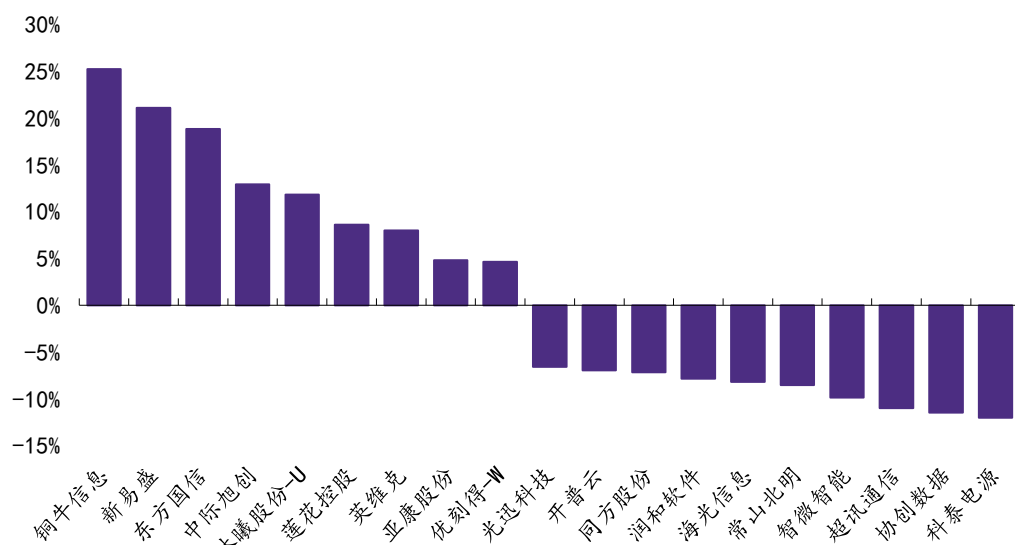
上周（2026.3.16-2026.3.20日），AI应用指数/AI算力指数/万得全A/中证红利日涨幅最大值分别为2.1%/5.62%/0.8%/-0.46%，AI应用指数/AI算力指数/万得全A/中证红利日跌幅最大值分别为-3.21%/-4.16%/-2.12%/-0.74%。AI算力指数内部，铜牛信息以25.23%录得上周最大涨幅，科泰电源以-12.01%录得上周最大跌幅。AI应用指数内部，东方国信以18.84%录得上周最大涨幅，每日互动以-15.47%录得上周最大跌幅。

图表 14：上周（2026.3.16-2026.3.20日）指数日涨跌幅



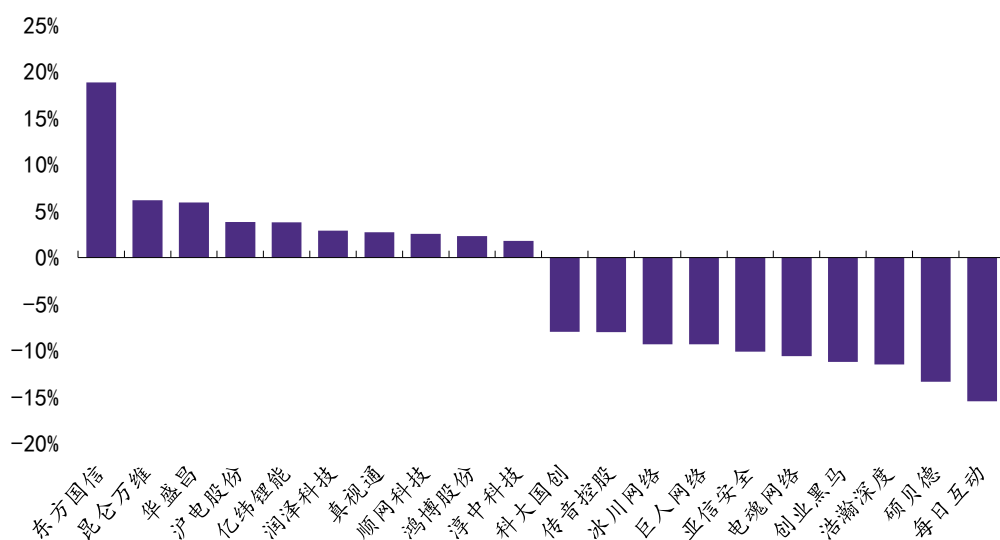
资料来源：wind, 华鑫证券研究

图表 15：上周（2026.3.16-2026.3.20日）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 16: 上周 (2026. 3. 16-2026. 3. 20 日) 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

2026年3月，英伟达于美国加州圣何塞举办GTC大会，同时OFC2026光通信大会于洛杉矶同期召开，两大展会均释放CPO技术发展重要信号，英伟达相关布局及行业技术落地、共识进一步明晰，核心内容如下：英伟达上调2025-2027年AI芯片营收预期至1万亿美元，同步了发布VeraRubin平台及RubinUltra机柜，并规划2028年量产Feynman架构芯片。CPO领域，英伟达推出了SN6800等三款量产CPO交换机，Feynman的路线图中也加上了Nvlink8Cpo，黄仁勋明确表态“我们需要更多的CPO产能”，2028年则有望作为CPO大规模落地关键节点。同期OFC2026大会上，英伟达进一步明确硅光CPO为下一代AI集群核心技术解，已基于TSMCCOUPE工艺打造光电协同封装，攻克微环调制器热敏感性难题并将探索DWDM方案提升传输能力；同时该展会中CPO技术落地全面提速，博通、Marvell、Coherent等企业分别发布CPO演进路线、光集成架构、三级集成定义及落地路径，英伟达联合AMD等六家头部企业成立OCI-MSA联盟，为CPO规模化落地制定统一开放标准。本次GTC和OFC大会的召开再次印证了AI算力板块的景气度，其中CPO的逐步商业化进程持续提速、发展路径愈发清晰。结合当前AI算力需求爆发、CPO技术持续突破及产业链成熟度提升的现状，我们维持对CPO板块的看好。

中长期，建议关注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），AI智能文字识别与商业大数据领域巨头的合合信息（688615.SH）、深耕工业AI与软件并长期服务高端装备等领域头部客户的能科科技（603859.SH）、专注于半导体等高端制造业的罗博特科（300757.SZ）。

图表 17：重点关注公司及盈利预测

公司代码	名称	2026-03-25 股价	EPS			PE			投资评级
			2024	2025E	2026E	2024	2025E	2026E	
300757.SZ	罗博特科	365.23	0.41	-0.30	0.30	890.80	-1217.43	1217.43	买入
301196.SZ	唯科科技	110.16	1.76	2.53	3.34	62.59	43.54	32.98	买入
603859.SH	能科科技	37.60	0.78	0.96	1.18	48.21	39.17	31.86	买入
688615.SH	合合信息	174.88		3.24	4.22		53.98	41.44	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。

## ■ 中小盘&主题&北交所组介绍

**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。

**倪汇康：**金融学士，2025年8月加盟华鑫证券研究所。

## ■ 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。