

2025 年 12 月 16 日



华鑫证券
CHINA FORTUNE SECURITIES

MistralAI 发布 Devstral2 系列，GPT-5.2 定义专家级智能

—计算机行业周报

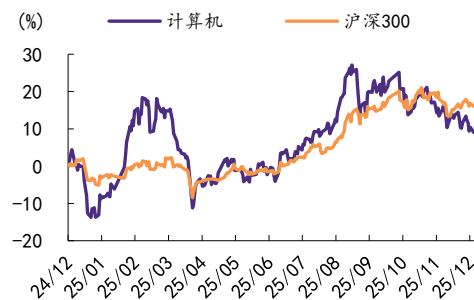
推荐(维持)

分析师：任春阳 S1050521110006
rancy@cfsc.com.cn

行业相对表现

表现	1M	3M	12M
计算机(申万)	-3.8	-10.8	6.1
沪深 300	-1.6	0.6	16.4

市场表现



相关研究

- 1、《计算机行业周报：DeepSeek 发布 V3.2 双模型，谷歌 DeepThink 引领推理革命》2025-12-08
- 2、《计算机行业周报：ClaudeOpus4.5 发布，DeepSeekMath-V2 重磅开源》2025-12-01
- 3、《计算机行业点评报告：小马智行（PONY.0）：商业化里程碑确立，跨城深耕与全球布局双驱》2025-11-29

投资要点

算力：算力租赁价格平稳，MistralAI 发布 Devstral2 系列

2025 年 12 月 10 日，MistralAI 发布了专为编程场景设计的开源模型家族 Devstral2，并同步推出命令行工具 MistralVibeCLI。这一技术迭代不仅将开源编程智能体的能力边界推向新高度，更通过企业级与本地化的双重解决方案设计，为开发者生态提供了兼具性能与实用性的 AI 编程工具链，成为大语言模型在代码生成领域落地的重要里程碑。Devstral2 系列采用密集 Transformer 架构，是 MistralAI 针对代码生成、多文件编辑与复杂代码库探索场景打造的专用模型，包含旗舰版（Devstral2，123B 参数）与轻量版（DevstralSmall2，24B 参数）两个版本，形成了覆盖企业级算力与消费级硬件的全场景适配体系。与 Devstral2 模型同步推出的 MistralVibeCLI，是 MistralAI 打造端到端编程自动化体系的关键组件。作为开源命令行工具，MistralVibeCLI 以 Devstral2 模型为推理核心，实现了自然语言与代码操作的直接映射。

AI 应用：Perplexity 周时长环比+10.58%，GPT-5.2 定义专家级智能

2025 年 12 月 12 日，OpenAI 正式发布了其新一代模型系列 GPT-5.2。此次发布正值该公司成立十周年之际，标志着其在通用人工智能发展道路上又迈出了坚实一步。新推出的模型家族包括即时版 GPT-5.2Instant、思考版 GPT-5.2Thinking 以及专业版 GPT-5.2Pro，旨在全方位满足从日常办公到深度专业任务的不同需求，并在多项基准测试中展现出卓越性能，特别是在处理高难度知识型工作方面表现突出。

AI 融资动向：Fal.ai 完成 1.4 亿美元融资，巩固 AI 内容生成设施地位

2025 年 12 月 10 日，多模态 AI 平台 Fal.ai 宣布完成 1.4 亿美元 D 轮融资，公司估值达到 45 亿美元。本轮由红杉资本 (Sequoia) 领投，KleinerPerkins 跟投，新进投资者 AlkeonCapital 和 NVIDIA 的风险投资部门 NVentures 参投，现有投资者包括 AndreessenHorowitz(a16z)、KindredVentures、Meritech、BessemerVenturePartners、NotableCapital、ShopifyVentures 和 SalesforceVentures。

继续参投支持。该轮融资巩固了 fal 作为 AI 生成内容基础设施层在开发者、企业和创意生态系统中的地位。Fal.ai 成立于 2021 年，由 BurkayGur 和 GorkemYurtseven 共同创立，专注于 AI 生成音频、视频和图像领域。其云平台提供超过 600 个优化用于生成图像、音频和视频的 AI 模型，其中包括 OpenAIGroupPBC 的 Sora2 和 GoogleLLC 上个月发布的 NanoBananaPro 图像生成器，以及大量开源算法。

投资建议

2025 年 12 月 12 日，博通 CEO 陈福阳在 2025 财年第四季度财报电话会议中披露，Anthropic 已累计向博通订购谷歌 TPUIronwood 机架，订单总额达 210 亿美元，折合人民币约 1481.8 亿元。财报同时显示，博通该季度实现营收 180.2 亿美元，同比增幅达 28.2%；其中 AI 芯片业务表现尤为亮眼，销售额同比激增 74%，累计未完成订单规模攀升至 730 亿美元，同时成功拓展第五家 XPU 定制芯片客户。当前谷歌 TPU 由谷歌主导架构设计，博通则承担核心工程实现工作，且该代 TPU 产品能效比较前代实现六倍提升。目前，谷歌正持续推进 TPU 产品的商业化布局，逐步向云服务客户及 Meta 等科技企业开放供应。在 AI 算力领域电力瓶颈愈发突出的行业背景下，谷歌与博通联手打造的 TPU 产品，有望成为破解行业算力能效困境的重要替代方案。此次博通财报的超预期表现，结合其与谷歌在 TPU 芯片领域的商业化落地推进及供应链扩张，为整个 AI 算力板块注入了强劲增长动能。从行业基本面的看，博通 AI 芯片销售额的高速增长及未来 18 个月 730 亿美元的未完成订单储备，充分印证了 AI 算力基础设施领域需求的刚性增长态势，彰显了行业的高景气度。谷歌与博通的 TPU 合作凭借双方技术协同优势，有望加速全球算力基础设施的迭代升级，进一步巩固算力作为 AI 产业核心底座的地位。叠加行业需求确定性与持续技术迭代，算力板块的业绩增长具备坚实支撑，我们持续看好 AI 算力板块。

中长期，建议关注加快扩张算力业务的精密零部件龙头迈信林（688685.SH）、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）。

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	2025-12-16 股价	EPS			PE			投资评级
			2024	2025E	2026E	2024	2025E	2026E	
301196.SZ	唯科科技	76.15	1.76	2.53	3.34	43.27	30.10	22.80	买入
603859.SH	能科科技	38.97	0.78	0.96	1.18	49.96	40.59	33.03	买入
688615.SH	合合信息	213.60	4.01	3.37	4.11	53.27	63.38	51.97	买入
688685.SH	迈信林	50.63	0.31	1.64	2.26	163.32	30.87	22.40	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，DEEPSEEK 发布 V3.2 双模型 5

 1.1、Tokens 跟踪..... 5

 1.2、数据跟踪：算力租赁价格平稳 6

 1.3、产业动态：MistralAI 发布 Devstral2 系列，推动大语言模型代码生成落地 6

2、AI 应用动态：PERPLEXITY 周时长环比+10.58%，GPT-5.2 定义专家级智能 10

 2.1、周流量跟踪：Perplexity 周时长环比+10.58%..... 10

 2.2、产业动态：模型实力再突破，GPT-5.2 定义专家级智能 10

3、AI 融资动向：FAL.AI 完成 1.4 亿美元融资，巩固 AI 内容生成设施地位 14

4、行情复盘 16

5、投资建议 18

6、风险提示 18

图表目录

图表 1：TOKENS 规模 LEADERBOARD 示意..... 5

图表 2：市场份额占据示意 5

图表 3：上周算力租赁情况 6

图表 4：SWE-BENCHVERIFIED 基准测试结果示意..... 7

图表 5：不同规模模型在 SWE-BENCH 验证常规性能表现对比 7

图表 6：MISTRALAI 在 GITHUB 上的开源项目 MISTRAL-VIBE 代码页面..... 8

图表 7：修改版 MIT 许可证编码 8

图表 7：2025.12.05-2025.12.11AI 相关网站流量..... 10

图表 8：GPT-5.2 在多项基准测试中的表现..... 11

图表 9：GPT-5.2THINKING 在 GDPVAL 上的表现..... 11

图表 10：GPT-5.2 的性能及价格..... 12

图表 11：GPT-5.2 在 SWE-BENCHPRO 上的表现..... 12

图表 12：GPT-5.2THINKING 与 GPT-5.1THINKING 在错误回答比例方面的对比..... 13

图表 13：上周 AI 初创公司融资动态 14

图表 14：上周（12.8-12.12 日）指数日涨跌幅..... 16

图表 15：上周（12.8-12.12 日）AI 算力指数内部涨跌幅度排名 16

图表 16：上周（12.8-12.12 日）AI 应用指数内部涨跌幅度排名 17

图表 17：重点关注公司及盈利预测 18

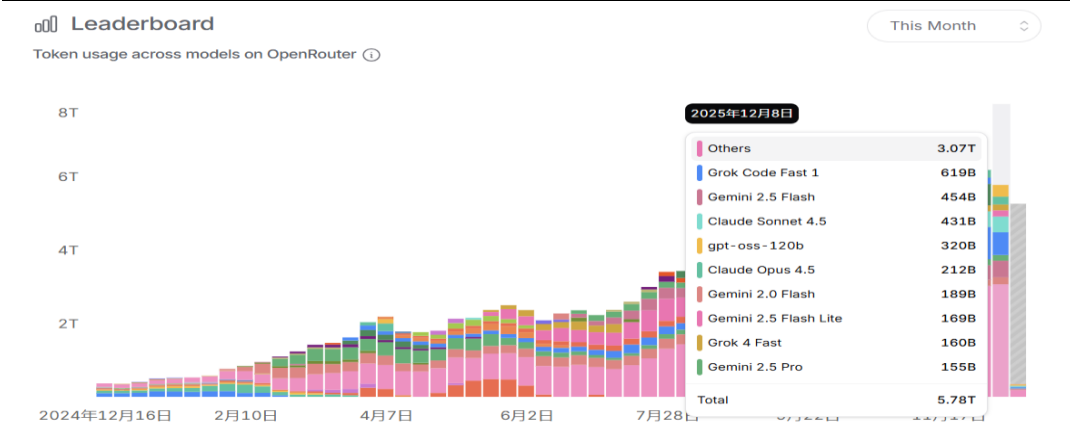
1、算力动态：算力租赁价格平稳，DeepSeek 发布 V3.2 双模型

1.1、Tokens 跟踪

根据 OpenRouter 公开数据，2025 年 12 月 8 日至 12 月 14 日，周度 token 消耗量有所下降，调用量为 5.78T，环比上周-6.6%。在 tokens 规模 leaderboard 前五名中，xAI 旗下 GrokCodeFast1 以 619Btokens 位居榜首；Google 的 Gemini2.5Flash 以 454B 位列第二；Anthropic 的 ClaudeSonnet4.5 以 431Btokens 排第三；OpenAI 的 GPT-OSS-120B 和 Anthropic 的 ClaudeOpus4.5 分别以 320B、212Btokens 位列第四、第五。

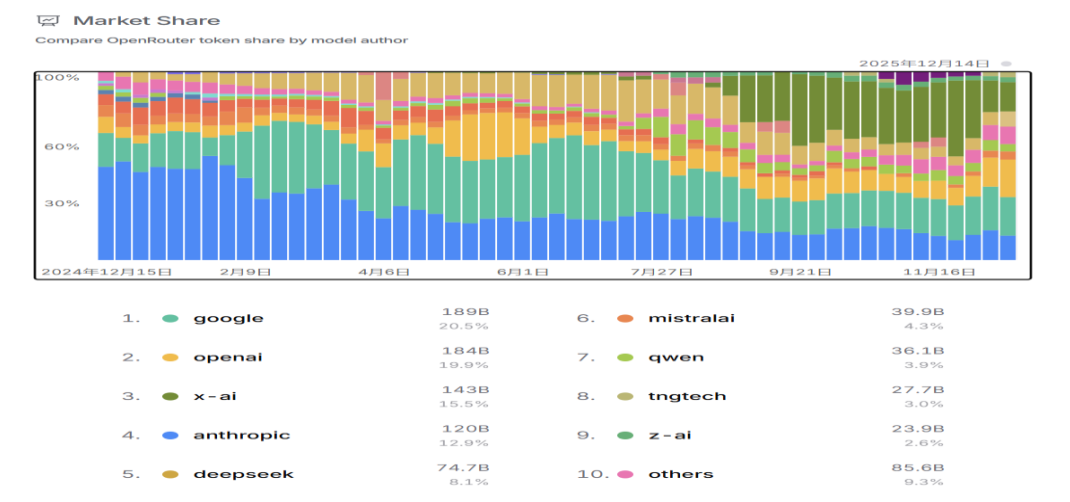
从市场份额维度来看，Google 以 189Btokens 占据 20.5%的份额，稳居首位；OpenAI 以 184Btokens 占比 19.9%，位列第二；xAI、Anthropic、Deepseek 则分别以 143B、120B、74.7Btokens，对应占据 15.5%、12.9%、8.1%的市场份额。

图表 1：tokens 规模 leaderboard 示意



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；显卡配置为 A800-80G 中，恒源云 16 核+256G 价格为 7.50 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	3.99%
	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	-	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

1.3、产业动态：MistralAI 发布 Devstral2 系列，推动大语言模型代码生成落地

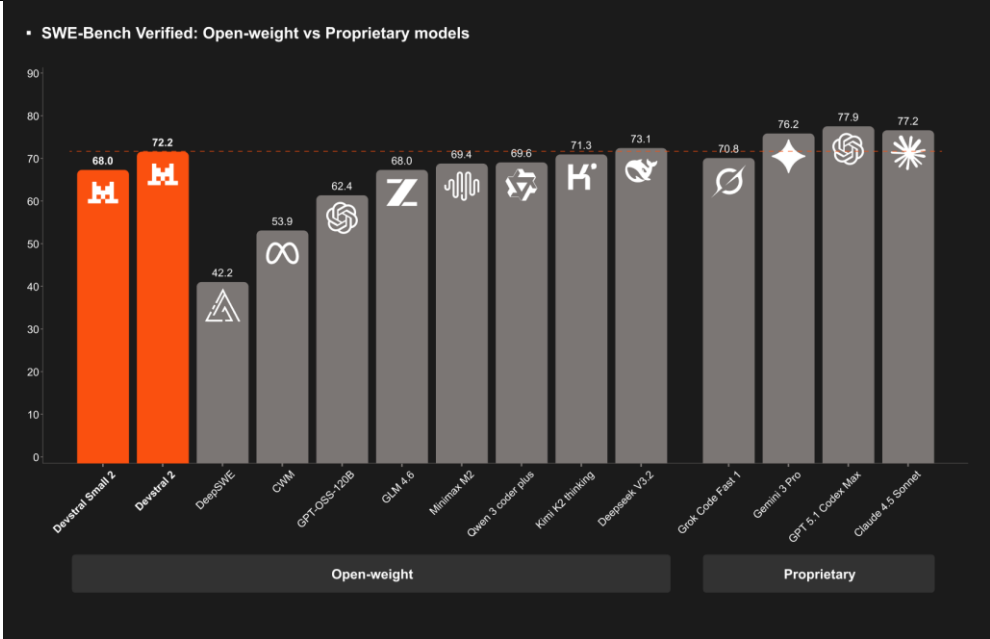
2025 年 12 月 10 日，MistralAI 发布了专为编程场景设计的开源模型家族 Devstral2，并同步推出命令行工具 MistralVibeCLI。这一技术迭代不仅将开源编程智能体的能力边界推向新高度，更通过企业级与本地化的双重解决方案设计，为开发者生态提供了兼具性能与实用性的 AI 编程工具链，成为大语言模型在代码生成领域落地的重要里程碑。

Devstral2 系列采用密集 Transformer 架构，是 MistralAI 针对代码生成、多文件编辑与复杂代码库探索场景打造的专用模型，包含旗舰版（Devstral2，123B 参数）与轻量版（DevstralSmall2，24B 参数）两个版本，形成了覆盖企业级算力与消费级硬件的全场景适配体系。该系列最核心的技术特征是 256K 超长上下文窗口，这一设计使其能够处理大规模多文件代码库的关联分析与逻辑生成任务，突破了传统代码大模型在上下文长度上的限制，解决了复杂项目中跨文件依赖、全局逻辑梳理的技术痛点。

从功能定位来看，123B 参数的 Devstral2 作为企业级主力模型，擅长工具调用、多文件编辑与复杂代码库探索，可支撑大型软件开发团队的自动化编程需求；24B 参数的 DevstralSmall2 则具备轻量高效的特性，支持在消费级硬件上本地运行，且兼容图像输入能力，适配隐私敏感场景或离线开发环境的需求。在性能评测中，Devstral2 在 SWE-benchVerified 基准测试中取得 72.2 的高分，超越多数开源竞品；同时其成本效率显著优于同参数规模的闭源模型。

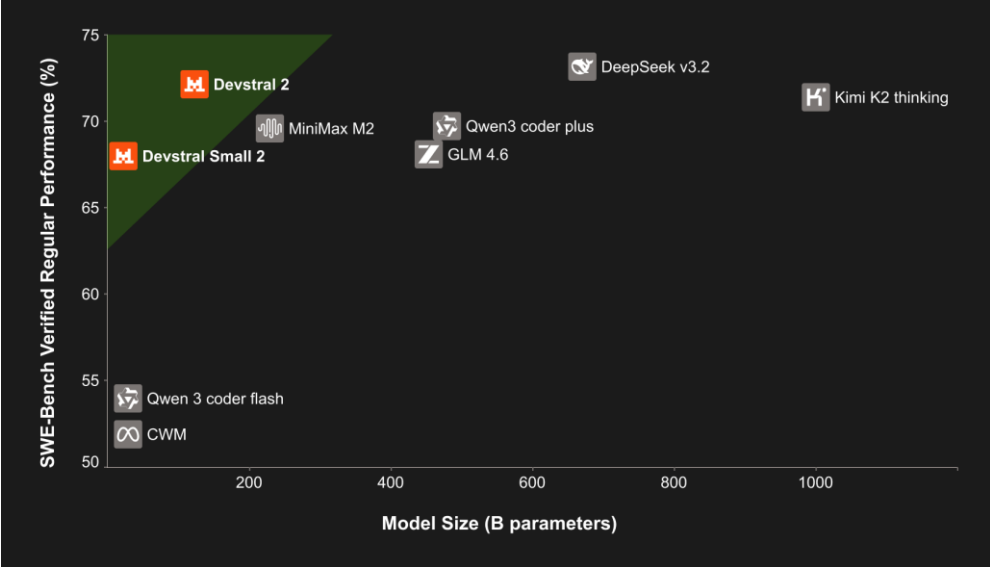
Mistral 官方指出，Devstral2（123B）和 DevstralSmall2（24B）分别比 DeepSeekV3.2 小 5 倍和 28 倍，比 KimiK2 小 8 倍和 41 倍。这证明了紧凑型模型可以匹敌甚至超越更大型竞争对手的性能。

图表 4：SWE-benchVerified 基准测试结果示意



资料来源：机器之心，华鑫证券研究

图表 5：不同规模模型在 SWE-Bench 验证常规性能表现对比

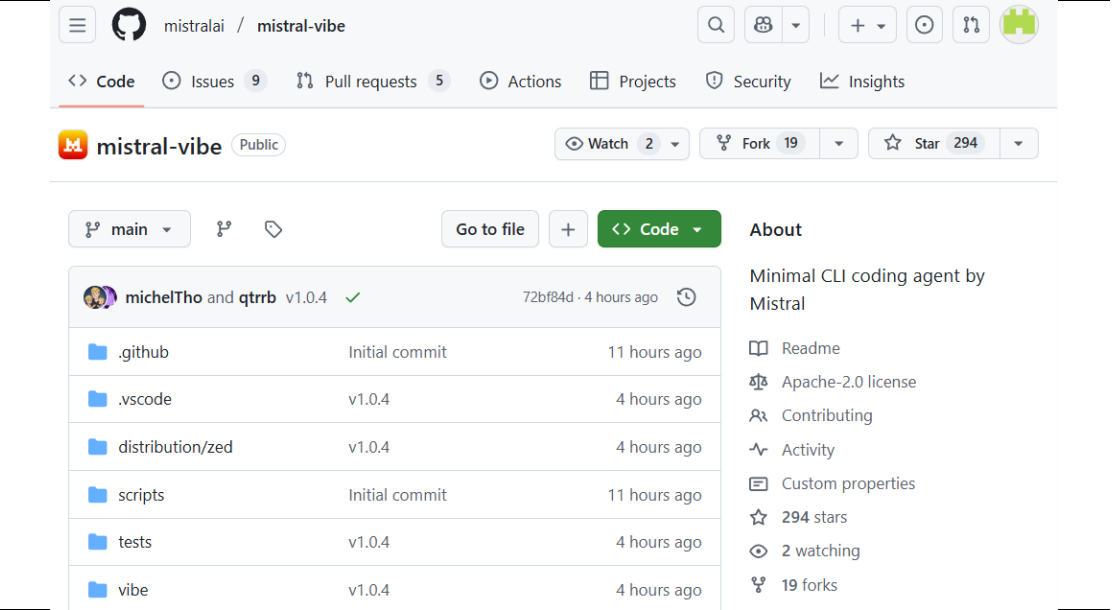


资料来源：机器之心，华鑫证券研究

与 Devstral2 模型同步推出的 MistralVibeCLI，是 MistralAI 打造端到端编程自动化体系的关键组件。作为开源命令行工具，MistralVibeCLI 以 Devstral2 模型为推理核心，实现了自然语言与代码操作的直接映射，其核心功能可归纳为三方面：

其一，自动化项目工程管理。工具能够自动扫描代码仓库的文件结构与 Git 状态，为模型提供准确的上下文，减少开发者在非业务逻辑开发上的时间投入。其二，交互式命令与自定义配置。开发者可通过自然语言执行 Shell 命令、调试代码、配置开发环境，且支持对工具的执行逻辑进行自定义配置，适配不同开发场景的操作习惯。其三，深度上下文理解与安全增强。与传统代码工具仅能处理单文件上下文不同，MistralVibeCLI 依托 Devstral2 的超长上下文能力，可深度理解整个代码库的逻辑关联；同时工具内置输入验证、自动补全与代码审计功能，降低了自动化编程过程中的语法错误与安全风险。

图表 6: MistralAI 在 GitHub 上的开源项目 mistral-vibe 代码页面

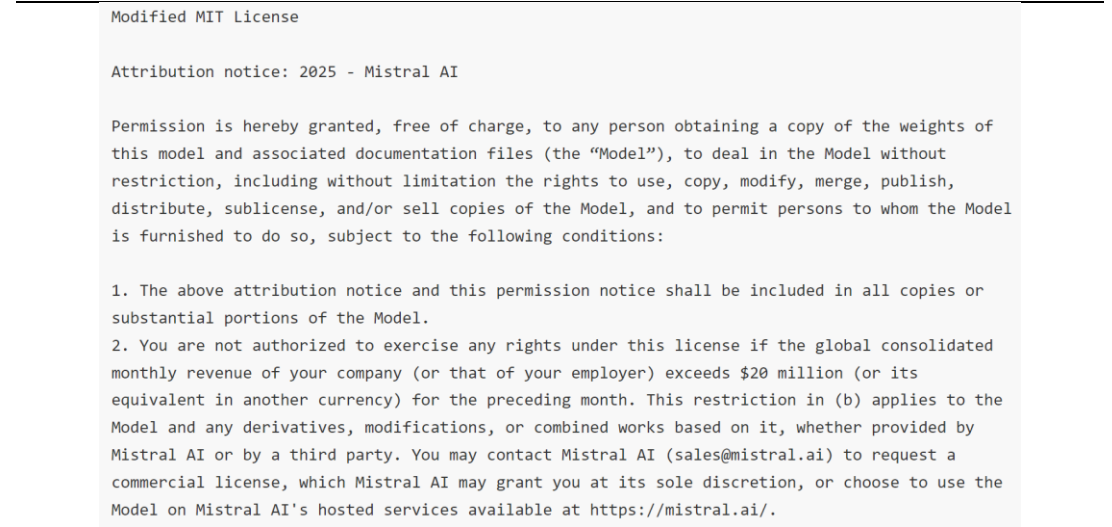


资料来源：机器之心，华鑫证券研究

值得注意的是，Devstral2 采用的许可证是一种修改版 MIT 许可证。对比标准的 MIT 许可证，可以看到一个重大差异：新增的收入限制条款。标准版 MIT 极度宽松。只要你保留版权声明，你可以将代码用于任何目的，包括商业用途、修改、分发、闭源发布，没有任何收入或公司规模的限制。Mistral 许可证则在第 2 条中增加了一个巨大的限制条件（毒丸条款）：
You are not authorized to exercise any rights under this license if the global consolidated monthly revenue of your company... exceeds \$20 million... for the preceding month. 即如果你的公司全球合并月收入超过 2000 万美元，你将无权行使本许可证下的任何权利。也就是说，超过收入门槛的公司必须联系 MistralAI 购买商业授权，或者使用他们的付费 API 服务。

并且这一限制不仅适用于原模型，还明确延伸到了 derivatives, modifications, or combined works（衍生品、修改版或结合作品）。这意味着如果你基于这个模型微调了一个新模型，大公司依然不能免费使用你的微调版。

图表 7: 修改版 MIT 许可证编码



资料来源：机器之心，华鑫证券研究

Mistral 目前正通过其 API 免费提供 Devstral2。免费期结束后，API 定价为：Devstral2 每百万 token\$0.40/\$2.00（输入/输出）；DevstralSmall2 为\$0.10/\$0.30。他们还宣布已与智能体工具 KiloCode 和 Cline 合作，将 Devstral2 带入用户现有的开发环境中。MistralVibeCLI 也已作为 Zed 的扩展程序提供，因此用户可以直接在 IDE 内部使用它。

2、AI 应用动态：Perplexity 周时长环比+10.58%，GPT-5.2 定义专家级智能

2.1、周流量跟踪：Perplexity 周时长环比+10.58%

本期（2025.12.05-2025.12.11）AI 相关网站流量数据：访问量前三位分别为 ChatGPT（1327.0M）、Bing（780.4M）和 Gemini（391.7M），访问量环比增速第一为 Gemini（4.56%）；平均停留时长前三位分别为 Character.AI（00:18:18）、Discord（00:10:41）和 Kimi（00:08:38）；平均停留时长环比增速第一为 Perplexity（10.58%）。

图表 7：2025.12.05-2025.12.11AI 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	1327.0	-0.38%	6:32	-0.25%
Bing	搜索	微软	780.4	1.02%	7:58	0.42%
Gemini	聊天机器人	谷歌	391.7	4.56%	7:05	0.95%
Canva	在线设计	Canva	223.7	-3.45%	6:04	-0.82%
Discord	游戏社区	微软	135.2	2.89%	10:41	-1.23%
Github	代码托管	微软	114.8	1.06%	6:29	0.00%
Perplexity	AI 搜索	Perplexity	44.02	1.80%	5:03	10.58%
Character.AI	聊天机器人	Character.AI	40.94	-4.75%	18:18	-1.08%
NotionAI	文本/笔记	Notion	38.02	-1.07%	8:02	0.00%
DeepL	翻译工具	DeepL	28.97	-0.62%	1:37	-36.60%
QuillBot	释义工具	QuillBot	13.29	-4.94%	3:08	-0.53%
Kimi	聊天机器人	MoonshotAI	9.36	-4.31%	8:38	0.39%
文心一言	聊天机器人	百度	0.93	0.37%	2:52	-3.37%

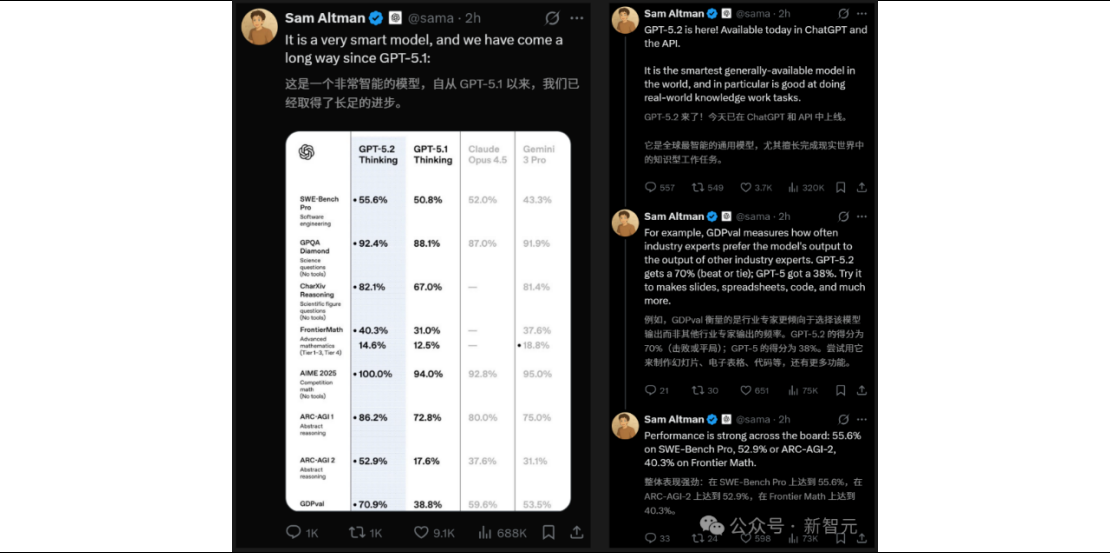
资料来源：similarweb, 华鑫证券研究

2.2、产业动态：模型实力再突破，GPT-5.2 定义专家级智能

2025 年 12 月 12 日，OpenAI 正式发布了其新一代模型系列 GPT-5.2。此次发布正值该公司成立十周年之际，标志着其在通用人工智能发展道路上又迈出了坚实一步。新推出的模型家族包括即时版 GPT-5.2Instant、思考版 GPT-5.2Thinking 以及专业版 GPT-5.2Pro，旨在全方位满足从日常办公到深度专业任务的不同需求，并在多项基准测试中展现出卓越性能，特别是在处理高难度知识型工作方面表现突出。

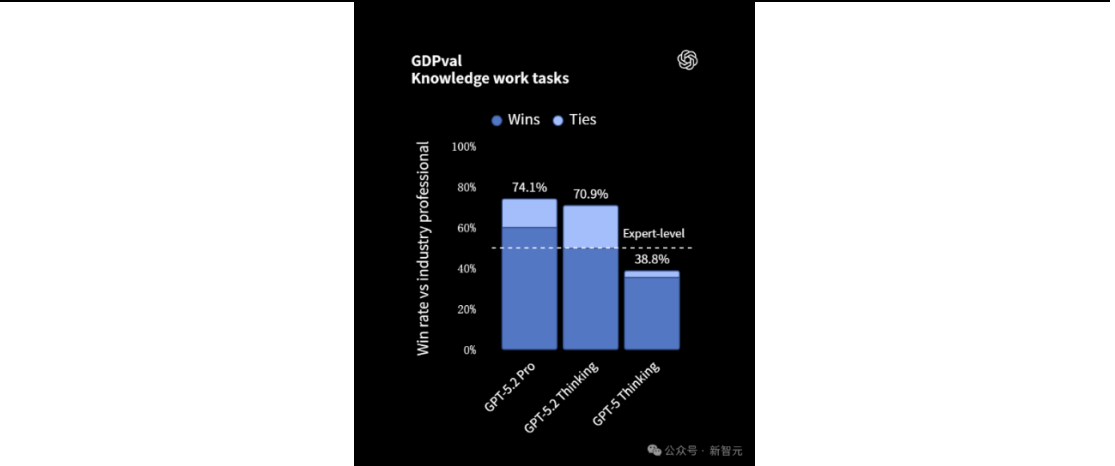
根据 OpenAI 公布的评估数据，GPT-5.2 系列在多个关键基准上实现了显著提升。在 SWE-BenchPro 测试中，模型取得了 55.6% 的得分，创造了新的记录。在 ARC-AGI-2 评估中，GPT-5.2Pro 更是以 52.9% 的优异成绩位居榜首。特别值得关注的是，在 GDPval 测试中，GPT-5.2 系列首次在 44 种职业任务上的表现超越了人类行业专家水平，在 70.9% 的情况下其产出结果被评定为优于或等同于顶尖专业人士。这些进步表明，该模型不仅能够高效协助完成诸如财务报表整理、演示文稿制作、代码编写等常规工作，还在处理需要复杂逻辑与多步骤推理的专业项目上展现出更强的可靠性与准确性。

图表 8：GPT-5.2 在多项基准测试中的表现



资料来源：新智元，华鑫证券研究

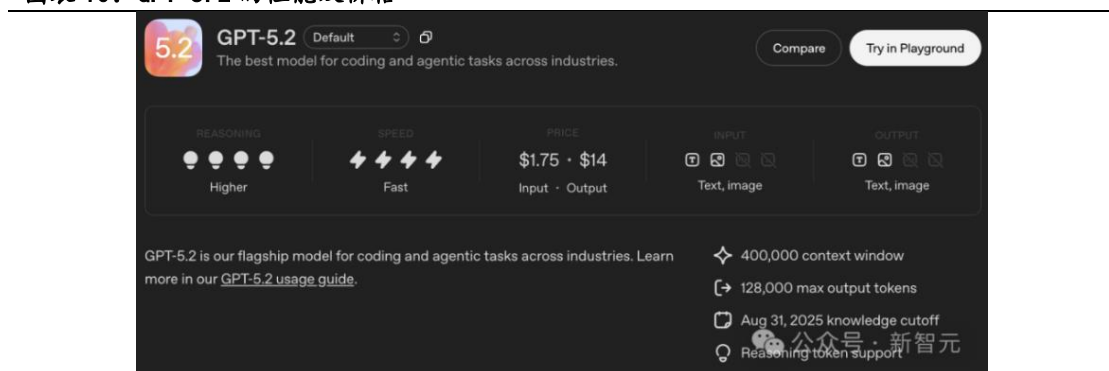
图表 9：GPT-5.2Thinking 在 GDPval 上的表现



资料来源：新智元，华鑫证券研究

在技术规格方面，GPT-5.2 系列支持高达 40 万个 Token 的上下文窗口，最大输出长度达到 12.8 万 Token，能够顺畅处理超长文本与复杂的多轮对话。其知识库已更新至 2025 年 8 月 31 日，确保了模型对世界最新动态的掌握。此外，该系列模型还增强了对视觉内容的理解能力，在图表解析、界面识别等任务上的错误率相较前代大幅降低，这对于金融、设计、工程等依赖视觉信息的工作流程具有重要意义。在长上下文推理任务中，模型在高达 256KToken 的测试中达到了接近百分之百的准确率，使其能够胜任深度文档分析、多源信息综合等复杂工作。

图表 10: GPT-5.2 的性能及价格

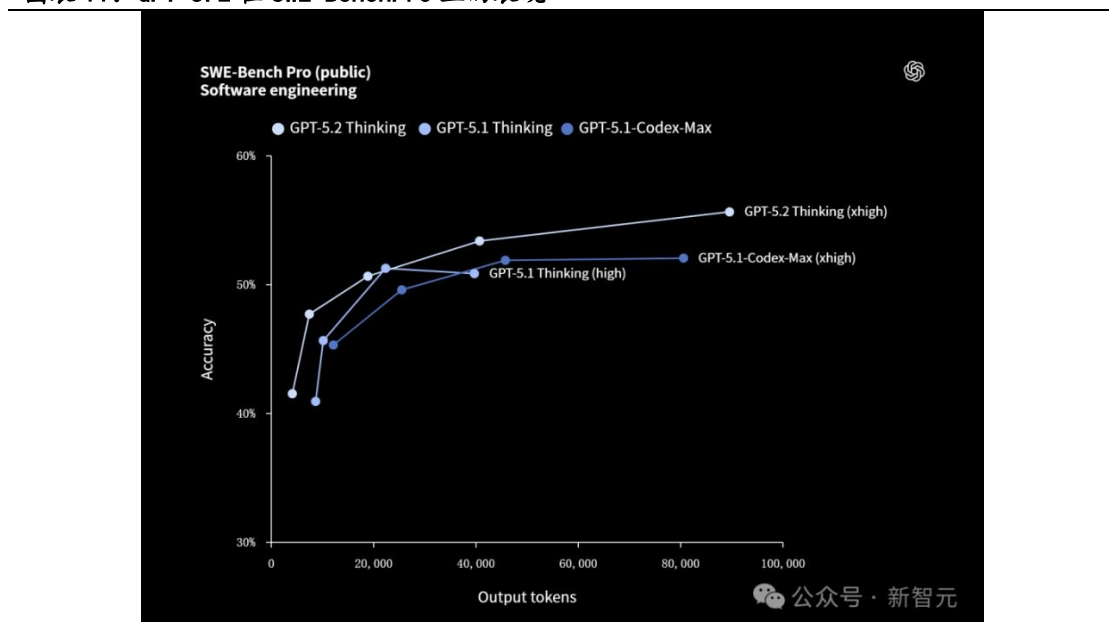


资料来源: 新智元, 华鑫证券研究

尽管模型能力全面提升, 但其使用成本也有所上升。据披露, GPT-5.2 的输入输出 Token 价格较前代提高了约 40%。然而, 由于模型在任务执行中表现出更高的 Token 效率, 因此在多项综合评价中, 其整体性价比反而被认为更具优势。目前, 付费用户可通过 ChatGPT 的相关套餐优先体验新模型, 而开发者也可通过相应的 API 进行调用。

在编程能力方面, GPT-5.2 系列同样树立了新的标杆。除了在 SWE-BenchPro 上的突破, 其在已验证的 SWE-bench 测试中也获得了 80% 的高分, 这意味着它能够更可靠地进行生产环境代码调试、功能实现与大型代码库重构。早期测试反馈显示, 该模型已成为全栈工程师的得力助手, 尤其擅长处理包含复杂或非常规用户界面, 如 3D 元素的前端开发任务。通过简单的提示词, 模型便能生成具备交互功能的海浪模拟网页、节日贺卡制作器或动态打字游戏等单页应用, 展示了其强大的端到端实现能力。

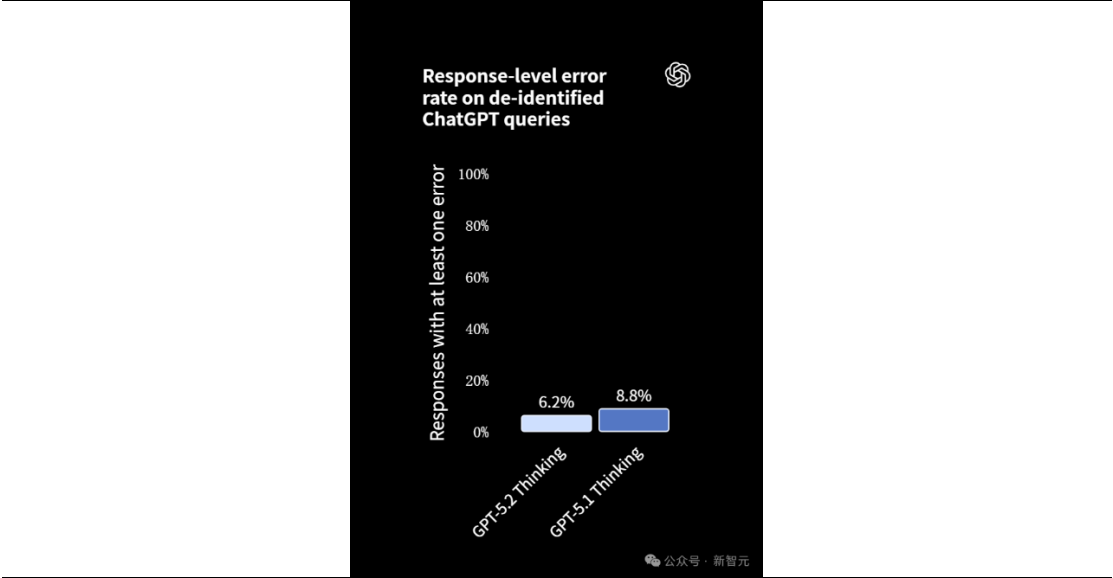
图表 11: GPT-5.2 在 SWE-BenchPro 上的表现



资料来源: 新智元, 华鑫证券研究

GPT-5.2 在减少“幻觉”方面取得了进展。在一组经过处理的真实用户查询测试中, 其产生错误回答的比例相比前代模型降低了约 30%, 这极大地增强了其在研究、分析、决策支持等严肃知识工作中的可靠性。与此同时, 模型在工具调用与端到端 workflow 管理上的能力也得到加强。在涉及多步骤、多系统协调的客户服务模拟场景中, GPT-5.2 能够更完整地处理从航班重新预订、特殊需求安排到赔偿协调的全链条任务, 减少了流程中断, 提升了执行效率。

图表 12: GPT-5.2Thinking 与 GPT-5.1Thinking 在错误回答比例方面的对比



资料来源：新智元，华鑫证券研究

在推动科学研究方面，GPT-5.2 系列也展现出巨大潜力。在研究生水平的 GPQADiamond 基准测试中，GPT-5.2Pro 达到了 93.2% 的得分。在协助解决开放性的学术问题中，模型也已能扮演更积极的角色。例如，在一项针对统计学习理论中最大似然估计量学习曲线单调性问题的研究中，GPT-5.2Pro 在人类研究者未提供具体证明策略或大纲的情况下，独立完成了关键证明步骤，并由人类负责后续的验证与论文撰写工作。这种协作模式预示着人工智能在加速数学与科学探索方面可能开启新的范式。GPT-5.2 系列的发布，不仅是 OpenAI 在技术长征中的最新成果，也进一步强化了人工智能作为强大辅助工具，在提升专业工作效率、激发科学创新乃至重塑工作流程方面的关键地位。

3、 AI 融资动向：Fal.ai 完成 1.4 亿美元融资，巩固 AI 内容生成设施地位

2025 年 12 月 10 日，多模态 AI 平台 Fal.ai 宣布完成 1.4 亿美元 D 轮融资，公司估值达到 45 亿美元。本轮由红杉资本(Sequoia)领投，KleinerPerkins 跟投，新进投资者 AlkeonCapital 和 NVIDIA 的风险投资部门 NVentures 参投，现有投资者包括 AndreessenHorowitz(a16z)、KindredVentures、Meritech、BessemerVenturePartners、NotableCapital、ShopifyVentures 和 SalesforceVentures 继续参投支持。该轮融资巩固了 fal 作为 AI 生成内容基础设施层在开发者、企业和创意生态系统中的地位。

Fal.ai 成立于 2021 年，由 BurkayGur 和 GorkemYurtseven 共同创立，专注于 AI 生成音频、视频和图像领域。其云平台提供超过 600 个优化用于生成图像、音频和视频的 AI 模型，其中包括 OpenAIGroupPBC 的 Sora2 和 GoogleLLC 上个月发布的 NanoBananaPro 图像生成器，以及大量开源算法。

目前，Fal.ai 已拥有数百家企业客户，平台用户超过 200 万开发者。截至 2025 年 10 月，公司 ARR 已超过 2 亿美元。自 2024 年以来，公司收入增长超过三倍，平台用户数量持续保持在 200 万开发者以上。自 7 月份完成 C 轮融资后，公司在四个月内运营收入翻了一番。每月提供数十亿实时生成的资产，涵盖图像、视频、音频和 3D 内容。

开发者可通过统一的 API 访问 Fal 云平台的模型，这一设计免除了学习各类算法技术细节的繁琐，简化了模型切换流程。将应用程序从某一 Fal 托管模型迁移至另一模型时，无需进行大量代码修改，因为 API 始终保持一致。若开发者的需求无法通过 Fal 目录内的算法得到满足，他们可在名为 FalServerless 的服务上部署自定义模型。该服务能近乎即时地启动配备数千个图形处理单元的集群，确保客户工作负载能够迅速扩展，以应对突发的使用高峰。

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
Fal.ai	多模态 ai 平台	Sequoia	D 轮	1.4 亿美元	5.38 亿美元	45 亿美元
深度智耀	ai 制药	鼎晖百孚	D 轮	5000 万美 元	5241 万美 元	-

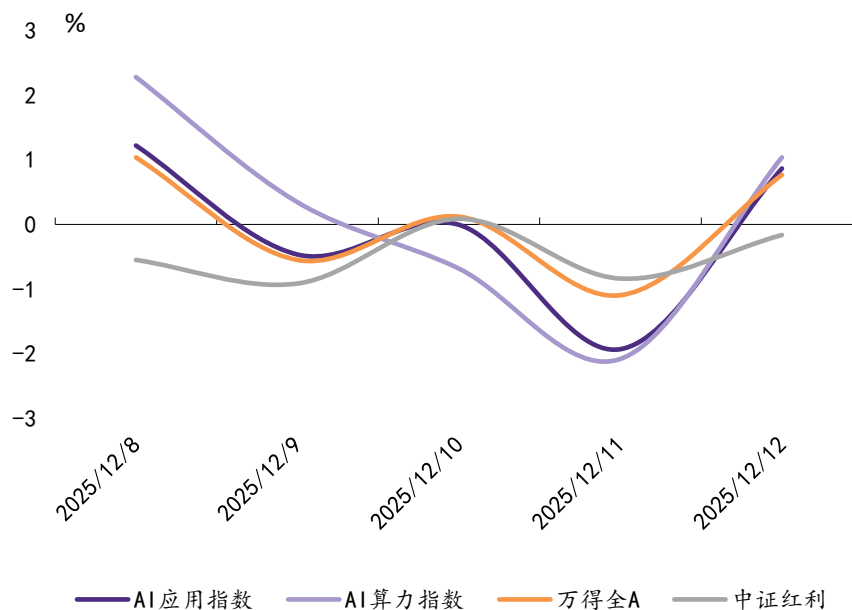
Bobyard	ai 建筑科技	8VC	A 轮	3500 万美 元	-	1 亿美元
---------	---------	-----	-----	--------------	---	-------

资料来源：Saasverse，wind，36 氪，华鑫证券研究

4、行情复盘

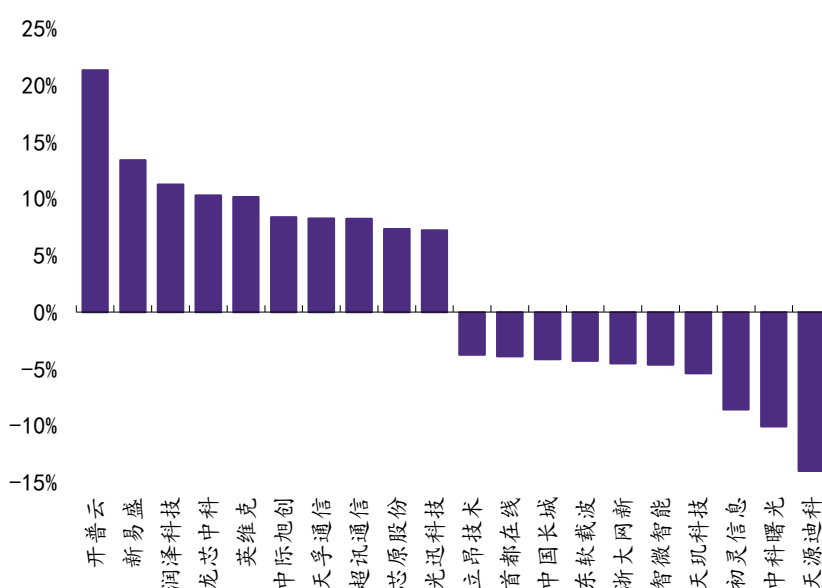
上周（12.8-12.12日），AI应用指数/AI算力指数/万得全A/中证红利日涨幅最大值分别为1.22%/2.28%/1.04%/0.09%，AI应用指数/AI算力指数/万得全A/中证红利日跌幅最大值分别为-1.93%/-2.08%/-1.10%/-0.92%。AI算力指数内部，开普云以+21.30%录得上周最大涨幅，天源迪科以-14.01%录得上周最大跌幅。AI应用指数内部，淳中科技以+13.90%录得上周最大涨幅，恒银科技以-13.62%录得上周最大跌幅。

图表 14：上周（12.8-12.12日）指数日涨跌幅



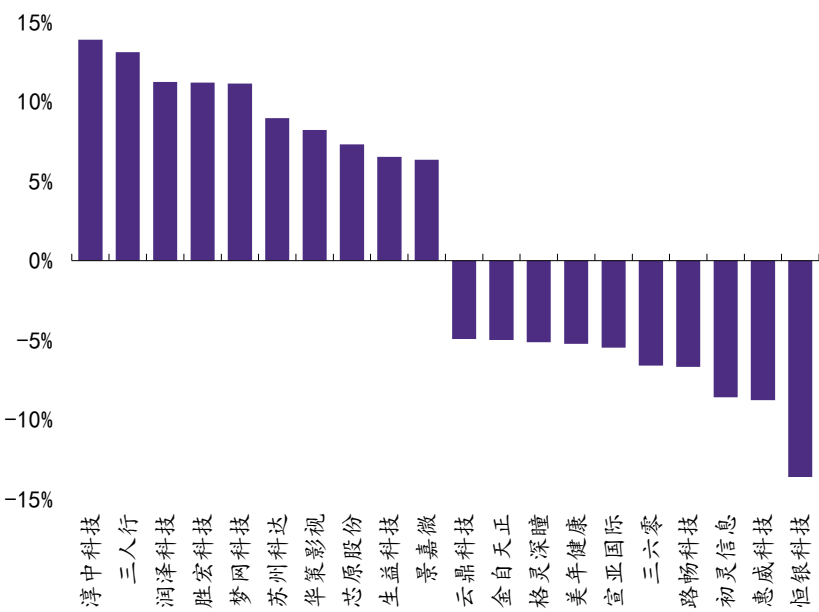
资料来源：wind, 华鑫证券研究

图表 15：上周（12.8-12.12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（12.8-12.12 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

2025 年 12 月 12 日，博通 CEO 陈福阳在 2025 财年第四季度财报电话会议中披露，Anthropic 已累计向博通订购谷歌 TPUIronwood 机架，订单总额达 210 亿美元，折合人民币约 1481.8 亿元。财报同时显示，博通该季度实现营收 180.2 亿美元，同比增幅达 28.2%；其中 AI 芯片业务表现尤为亮眼，销售额同比激增 74%，累计未完成订单规模攀升至 730 亿美元，同时成功拓展第五家 XPU 定制芯片客户。当前谷歌 TPU 由谷歌主导架构设计，博通则承担核心工程实现工作，且该代 TPU 产品能效比较前代实现六倍提升。目前，谷歌正持续推进 TPU 产品的商业化布局，逐步向云服务客户及 Meta 等科技企业开放供应。在 AI 算力领域电力瓶颈愈发突出的行业背景下，谷歌与博通联手打造的 TPU 产品，有望成为破解行业算力能效困境的重要替代方案。此次博通财报的超预期表现，结合其与谷歌在 TPU 芯片领域的商业化落地推进及供应链扩张，为整个 AI 算力板块注入了强劲增长动能。从行业基本面看，博通 AI 芯片销售额的高速增长及未来 18 个月 730 亿美元的未完成订单储备，充分印证了 AI 算力基础设施领域需求的刚性增长态势，彰显了行业的高景气度。谷歌与博通的 TPU 合作凭借双方技术协同优势，有望加速全球算力基础设施的迭代升级，进一步巩固算力作为 AI 产业核心底座的地位。叠加行业需求确定性与持续技术迭代，算力板块的业绩增长具备坚实支撑，我们持续看好 AI 算力板块。

中长期，建议关注加快扩张算力业务的精密零部件龙头迈信林（688685.SH）、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）。

图表 17：重点关注公司及盈利预测

公司代码	名称	2025-12-16 股价	EPS			PE			投资评级
			2024	2025E	2026E	2024	2025E	2026E	
301196.SZ	唯科科技	76.15	1.76	2.53	3.34	43.27	30.10	22.80	买入
603859.SH	能科科技	38.97	0.78	0.96	1.18	49.96	40.59	33.03	买入
688615.SH	合合信息	213.60	4.01	3.37	4.11	53.27	63.38	51.97	买入
688685.SH	迈信林	50.63	0.31	1.64	2.26	163.32	30.87	22.40	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业务资格。
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义务。

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。