下载日志已记录，仅供内部参考

证券研究报告

2026 年 10 月 27 日

华鑫证券 CHINA FORTUNE SECURITIES

MiniMax 发布 M3.1 Flash，Anthropic 推出 Claude Duet

—计算机行业周报

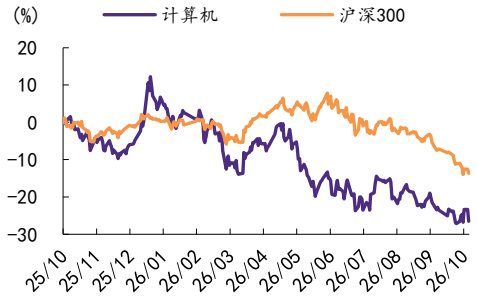
推荐(维持)

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)	-6.4	-6.5	-26.4
沪深 300	-8.8	-13.9	-14.1

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：Gemini4Argon 发布，DeepSeek 更新 V4.1-Flash-1017》2026-10-19

2、《计算机行业周报：GPT-6.1Sol 上线，Anthropic 发布 ClaudeHaiku5.5》2026-10-14

3、《计算机行业周报：GPT-6Sol 与 Luna 推出，Anthropic 发布 ClaudeOpus5.5》2026-09-30

投资要点

算力：阿里云 Meoo 接入千问 App 筑牢后台底座，MiniMax 发布 M3.1 Flash 形成双强格局

阿里云 AI 应用平台 Meoo 接入千问 App“我的应用”功能，提供覆盖数据存储、部署上线及长期更新的全托管云端服务，推动自然语言生成应用由 Demo 向生产级产品演进；同期 MiniMax 正式发布 M3.1 Flash（100 万上下文、输出 200+TPS），在长程智能体与跨应用自动化等测试中反超 Opus 5 与 DeepSeek-1017，综合成本较前代降低 70%且资费对标竞品闲时，与 DeepSeek 在轻量 Flash 段位形成国产双强格局，云端全托管治理与极致性价比轻量模型协同加速 AI 应用规模化渗透。

AI 应用：Deepl 周时长环比+2.21%，Anthropic 推出 Claude Duet

2026 年 10 月 22 日，Anthropic 推出了个人 AI 智能体应用 Claude Duet。该产品以主动式任务处理为核心定位，基于 Claude Opus 5.5 模型构建，具备连接用户邮件、日历、银行账户及企业业务系统等多类信息源的能力。与以往需要用户逐一下达指令的工具不同，Claude Duet 能够在用户尚未提出需求时主动识别待办事项，并持续跟进执行。目前，该产品已推出 iOS、Android 及网页端独立应用，首批面向美国地区 Pro 和 Max 订阅用户开放研究预览候补申请，尚未全面铺开。

AI 融资动向：Tenstorrent 完成 15 亿美元 E 轮融资，投后估值达 62 亿美元

AI 推理芯片设计商 Tenstorrent 宣布完成 15 亿美元 E 轮融资，投后估值达 62 亿美元，由 Trace Capital 与 Innovation X 领投，ARK Invest、富达、大摩等 15 家机构参投。公司主打 RISC-V 开源指令集结合大片上 SRAM 与 GDDR6 显存替代 HBM，Galaxy Blackhole 系统已实测跑通 DeepSeek R1 及 Kimi K2.6 等开源模型；商业化已斩获日本 ai&超 120 套系统大单，并向三星、LG 与现代汽车集团输出核心 IP 授权，全力布局自主可控 AI 推理算力供给体系。

投资建议

10 月 21 日，英伟达林宣布大幅扩展 Isaac 开放机器人平台，覆盖从数据生成、模型训练到部署运行的完整链

行业研究

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

路：发布开源运行时 IsaacOS，机器人学习框架 IsaacLab3.0 正式可用；上线包含 2,000 余项仿真资产的 SimReady 中央库与 Human1M 人类活动数据集，后者提供多视角视频与三维手部追踪数据；同时开放三套基于 DGXStation 的参考工作流，支持以强化学习微调 GR00T 模型，并结合 Cosmos 世界模型生成合成数据。AgileRobots 同步发布基于 IsaacGR00T 与 JetsonThor 构建的平台 AgileONESN，计划 2027 年初供货，并将集成 Halos 安全栈；ABB、库卡、NEURA、UniversalRobots 等欧洲机器人厂商均已基于英伟达开放平台开发产品。黄仁勋在会上表示，机器人行业的"ChatGPT 时刻"正在临近。

本次发布的核心信号在于物理 AI 的竞争重心正从本体与运动控制能力，转向数据、训练与部署的工具链体系。此前行业的主要瓶颈在于真实交互数据稀缺、仿真到现实的迁移成本高企，英伟达此次同时补上数据资产、学习框架与运行时三层缺口，机器人开发开始被纳入标准化的流水线，厂商之间的竞争将更多取决于各自场景的数据积累与工程化能力。我们认为，物理 AI 当前正从单机演示走向可复制、可验证的开发范式，数据闭环与工具链将成为区分厂商长期竞争力的关键变量。

中长期，建议关注专注于功率半导体及模拟芯片测试系统研发的联动科技（301369.SZ）、专注于半导体等高端制造业的罗博特科（300757.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）、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	2026-10-27 股价	EPS			PE			投资评级
			2025	2026E	2027E	2025	2026E	2027E	
300757.SZ	罗博特科	373.79	-0.40	2.13	4.49	-934.48	175.49	83.25	买入
301196.SZ	唯科科技	87.18	2.05	1.92	2.39	39.28	43.29	34.75	买入
301369.SZ	联动科技	80.03	0.48	1.08	2.62	166.73	74.10	30.55	买入
603859.SH	能科科技	41.94	0.92	1.05	1.32	45.59	39.94	31.77	买入
688615.SH	合合信息	92.56	3.24	4.22	5.25	28.57	21.93	17.63	买入

资料来源：Wind，华鑫证券研究

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

正文目录

1、 算力动态：阿里云 MEOO 接入千问 APP 筑牢后台底座，MINIMAX 发布 M3.1 FLASH 形成双强格局..... 5

1.1、 Tokens 跟踪..... 5

1.2、 算力租赁动态：成熟卡型稳中有升，新一代高端算力租金延续攀升..... 7

1.3、 云厂动态：阿里云 Meoo 平台接入千问 App，为自然语言生成应用提供全托管云端后台支撑.. 10

1.4、 产业动态：MiniMax 发布轻量模型 M3.1 Flash，低成本高吞吐与 DeepSeek 形成国产“双强”格局..... 10

2、 AI 应用动态：DEEPL 周时长环比+2.21%，ANTHROPIC 推出 CLAUDE DUET..... 13

2.1、 周流量跟踪：Deepl 周时长环比+2.21%..... 13

2.2、 产业动态：Anthropic 推出 Claude Duet，以 Opus 5.5 为底座覆盖多场景任务执行..... 13

3、 AI 融资动向：AI 推理芯片设计商 TENSTORRENT 完成 15 亿美元 E 轮融资，投后估值达 62 亿美元.... 16

4、 行情复盘..... 18

5、 投资建议.....20....

6、 风险提示..... 21

图表目录

图表 1：TOKENS 规模 LEADERBOARD（TOKEN 消耗量为统计口径）..... 7

图表 2：文本类模型市场份额占据示意图（REQUEST 调用量为统计口径）..... 7

图表 3：2026.10.19 -2026.10.25 日 H100 价格（单位：美元/小时）..... 8

图表 4：2026.10.19 -2026.10.25 日 H200 价格（单位：美元/小时）..... 8

图表 5：2026.10.19 -2026.10.25 日 A100 价格（单位：美元/小时）..... 9

图表 6：2026.10.19 -2026.10.25 日 B200 价格（单位：美元/小时）..... 9

图表 7：2026.10.19 -2026.10.25 日 B300 价格（单位：美元/小时）..... 9

图表 8：MINIMAX M3.1 FLASH 与主流前沿模型多维度 BENCHMARK 评测成绩对比..... 11

图表 9：2026.10.17-2026.10.23 AI 相关网站流量..... 13

图表 10：CLAUDE DUET 发布..... 14

图表 11：CLAUDE DUET 可以提前介入用户的工作流程..... 14

图表 12：上周 AI 初创公司融资动态..... 16

图表 13：上周（2026.10.19-2026.10.23 日）指数日涨跌幅..... 18

图表 14：上周（2026.10.19-2026.10.23 日）AI 算力指数内部涨跌幅度排名..... 18

图表 15：上周（2026.10.19-2026.10.23 日）AI 应用指数内部涨跌幅度排名..... 19

图表 16：FICONTEC2025 年年中至今公告订单..... 20

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

图表 17：重点关注公司及盈利预测 21

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

1、 算力动态：阿里云 Meoo 接入千问 App 筑牢后台底座，MiniMax 发布 M3.1 Flash 形成双强格局

1.1、 Tokens 跟踪

根据 OpenRouter 公开数据，2026 年 10 月 19 日至 10 月 25 日，周度 Token 消耗量显著回升，消耗量达 188T，环比上周增加 14.63%。在 Tokens 规模 Leaderboard 前五名中，DeepSeek 的 DeepSeek V4.1 Flash 1017 以 33.6T tokens 跃居榜首；DeepSeek 的 DeepSeek V4.1 Flash 以 18T tokens 位列第二；MiniMax 的 MiniMax M3.1 Flash (free)以 13.7T tokens 位列第三；Z-ai 的 GLM 5.3 Flash 以 13.2T tokens 位列第四；Tencent 的 Hy4 preview 以 8.97T tokens 位列第五。

文本类模型的市场份额方面，deepseek 以 1.91B requests 占据 26.5%的份额，位列第一；google 以 1.45B requests 占据 20.1%的份额，位列第二；openai、z.ai、anthropic 分别以 1.34B、774M、394M requests，对应占据 18.5%、10.7%、5.5%的市场份额。

图像生成类模型调用量保持稳步增长，调用量为 8.82M requests，环比上周增加 7.17%。在 Image Leaderboard 中，Google 的 Nano Banana 2 (Gemini 3.1 Flash Image)以 1.48M requests 位居第一；OpenAI 的 GPT Image 2.5 Sunburst 以 1.24M requests 位列第二；OpenAI 的 GPT Image 2.5 Flare 以 736K requests 位列第三；Meta 的 Muse Image 以 692K requests 位列第四；Google 的 Nano Banana 2 Lite 以 664K requests 位列第五。从市场份额来看，google 以 3.69M requests 占据 41.8%的份额，位列第一；openai 以 2.87M requests 占据 32.5%，位列第二；meta、black-forest-labs、bytedance-seed 分别以 692K、407K、367K requests，对应占据 7.8%、4.6%、4.2%的市场份额。

嵌入模型调用量保持稳步增长，调用量为 426M requests，环比上周增加 5.45%。在 Embedding Leaderboard 前五名中，OpenAI 的 Text Embedding 3 Small 以 174M requests 位居榜首；Qwen 的 Qwen3 Embedding 8B 以 64.3M requests 位列第二；Google 的 Gemini Embedding 2 以 44.9M requests 位列第三；Google 的 Gemini Embedding 001 以 33.6M requests 位列第四；BAAI 的 bge-m3 以 25.3M requests 位列第五。从市场份额来看，openai 以 201M requests 占据 47.1%的份额，位居第一；google 以 81.9M requests 占据 19.2%的份额，位列第二；qwen、baai、sentence-transformers 分别以 70.4M、26.3M、19.8M requests，对应占据 16.5%、6.2%、4.7%的市场份额。

重排序模型调用量小幅增长，调用量为 13M requests，环比上周增加 2.36%。在 Rerank Leaderboard 前五名中，VoyageAI 的 rerank-2.5 以 4.61M requests 位居榜首；VoyageAI 的 rerank-3-lite 以 2.15M requests 位列第二；VoyageAI 的 rerank-2.5-lite 以 1.99M requests 位列第三；Cohere 的 Rerank v3.5 以 1.53M requests 位列第四；Qwen 的 Qwen3 Reranker 8B 以 780K requests 位列第五。从市场份额来看，voyageai 以 9.5M requests 占据 72.9%的份额，位居第一；cohere 以 2.66M requests 占据 20.4%的份额，位列第二；qwen、nvidia 分别以 780K、96K requests，对应占据 6.0%、0.7%的市场份额。

视频生成类模型调用量保持增长，调用量为 417K requests，环比上周增加 13.32%。在

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

Video Leaderboard 前五名中，Wan 的 Wan 3.0 以 70K requests 跃居第一；HeyGen 的 HeyGen Video 以 61K requests 位列第二；Google 的 Veo 3.1 Lite 以 43K requests 位列第三；ByteDance 的 Seedance 2.5 以 31K requests 位列第四；xAI 的 Grok Imagine Video 1.5 Lite 以 30K requests 位列第五。从市场份额来看，bytedance 以 101K requests 占据 24.1%的份额，位列第一；alibaba 以 82K requests 占据 19.6%的份额，位居第二；heygen、google、spacexai 分别以 72K、64K、47K requests，对应占据 17.3%、15.3%、11.4%的市场份额。

语音生成类模型调用量大幅增长，调用量为 16M requests，环比上周增加 90.70%。在 Speech Leaderboard 前五名中，Google 的 Gemini 3.8 Flash TTS 以 7.18M requests 稳居榜首；Google 的 Gemini 3.8 Flash Lite TTS 以 3.61M requests 位列第二；Elevenlabs 的 Eleven v4 Turbo 以 1.19M requests 位列第三；Drama 3 以 623K requests 位列第四；Hexgrad 的 Kokoro 82M 以 513K requests 位列第五。从市场份额来看，google 以 11.8M requests 占据 73.7%的份额，位列第一；elevenlabs 以 2.06M requests 占据 12.9%的份额，位居第二；fish-audio、hexgrad、space-x-ai 分别以 626K、513K、313K requests，对应占据 3.9%、3.2%、2.0%的市场份额。

语音转写类模型调用量大幅增长，调用量为 39.9M requests，环比上周增加 37.59%。在 Transcription Leaderboard 前五名中，OpenAI 的 Whisper Large V3 Turbo 以 15.8M requests 位列第一；OpenAI 的 Whisper Large V3 以 9.29M requests 位列第二；Microsoft 的 MAI-Transcribe 2 以 6.45M requests 位列第三；Scribe v2 以 2.34M requests 位列第四；OpenAI 的 GPT-4o Mini Transcribe 以 902K requests 位列第五。从市场份额来看，openai 以 27.6M requests 占据 69.2%的份额，位居第一；microsoft 以 6.48M requests 占据 16.3%的份额，位列第二；elevenlabs、nvidia、mistralai 分别以 2.42M、739K、637K requests，对应占据 6.1%、1.9%、1.6%的市场份额。

* 统计口径说明：以上数据均来源于 OpenRouter。当前 Rankings 采用 Token 与 Request 两套统计维度：仅 Tokens 规模 Leaderboard 以模型 Token 消耗量为统计口径，其余类目榜单及市场份额板块统一采用 Request 调用量为统计口径，两者统计维度不同，不宜直接对应或进行横向比较。

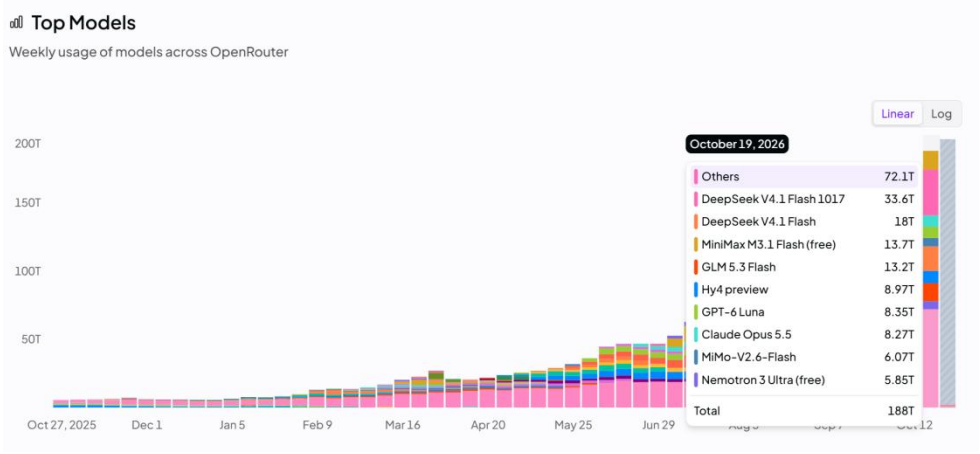
近日 AI Token 生产服务商趋境科技（Approaching.AI）宣布完成超 13 亿元 B 轮融资，2026 年内已累计募资 24 亿元，多项核心经营指标半年内实现超十倍速增长。本轮融资由浙财资本、萧山资本、工银投资、中关村科学城、星连资本及钱江世纪城共同领投，云创资本、商汤科技、网宿科技等参投，老股东尚势资本继续追投；此前公司于 7 月获河南汇融基金领投数亿元 A 轮。募集资金将重点用于高品质 AI Token 产能扩张、ATaaS 平台迭代、推理底层技术与新产品研发，并深化国产 PD 异构方案落地，构建多元、低成本的 AI 推理供给体系。



Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

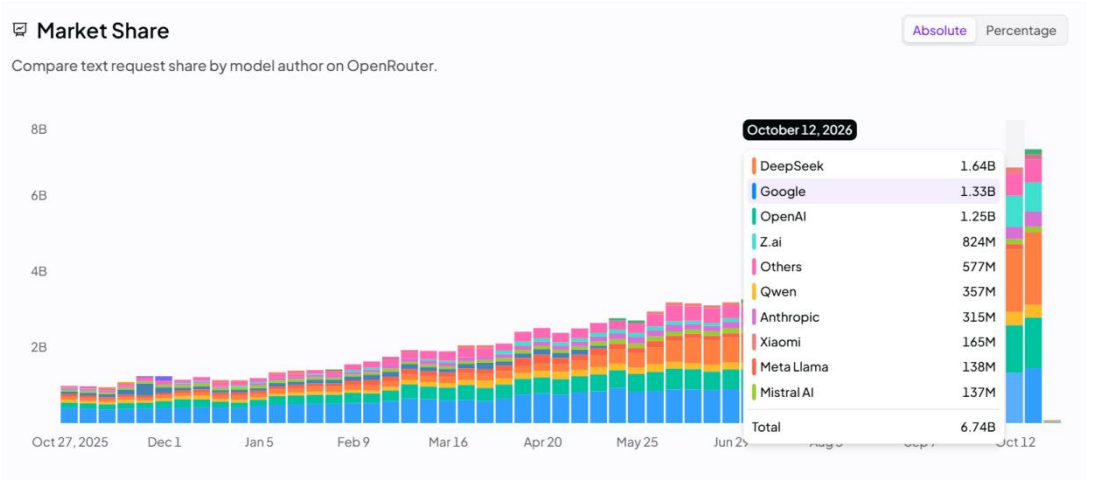
下载日志已记录，仅供内部参考

图表 1：Tokens 规模 Leaderboard（Token 消耗量为统计口径）



资料来源：OpenRouter，华鑫证券研究

图表 2：文本类模型市场份额占据示意图（Request 调用量为统计口径）



资料来源：OpenRouter，华鑫证券研究

1.2、算力租赁动态：成熟卡型稳中有升，新一代高端算力租金延续攀升

本周（2026.10.19-2026.10.25 日），算力租赁市场整体呈现成熟卡型温和走高、新一代高算力 GPU 租金持续大幅攀升态势。H100 指数/H200 指数/A100 指数/B200 指数/B300 指数日租金（美元/小时）最大值分别为 2.90/3.53/1.56/6.60/7.53，H100 指数/H200 指数/A100 指数/B200 指数/B300 指数日租金（美元/小时）最小值分别为 2.84/3.45/1.51/6.14/7.02。

对比上周（10.12-10.18）数据观察，成熟卡型价格重心小幅上移：H100、H200 与 A100 日租金最大值分别由上周的 2.85、3.51、1.53 美元/小时微增至本周的 2.90、3.53、1.56 美元/小时，日租金底部均保持稳步抬升。与此相对，新一代 Blackwell 架构算力租金出现显著连续走高：B200 日租金由周初的 6.14 美元/小时攀升至周末的 6.60 美元/小时（最大值较上周的 6.26 美元/小时进一步抬升 0.34 美元/小时）；B300 延续强劲涨势，日租金由

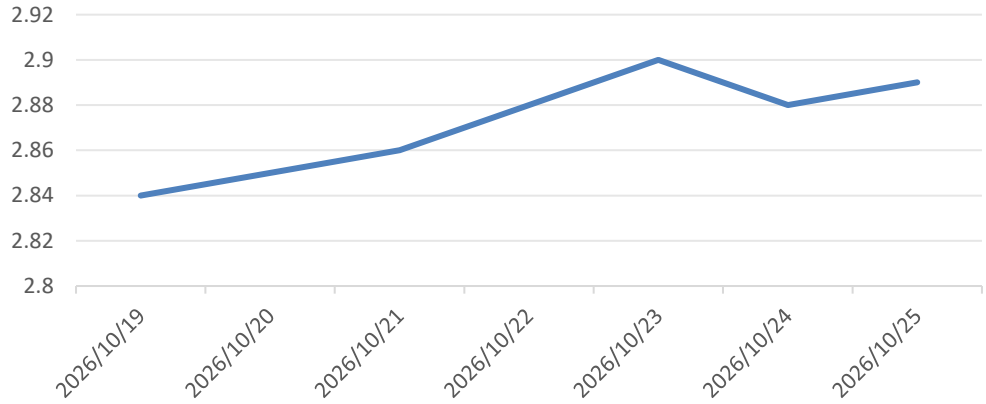
Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

周初的 7.02 美元/小时持续上行并稳定在 7.53 美元/小时（最大值突破 7.5 美元关口，较上周最大值 7.01 美元/小时上浮 0.52 美元/小时），反映出市场对新一代高性能算力的需求依旧旺盛。

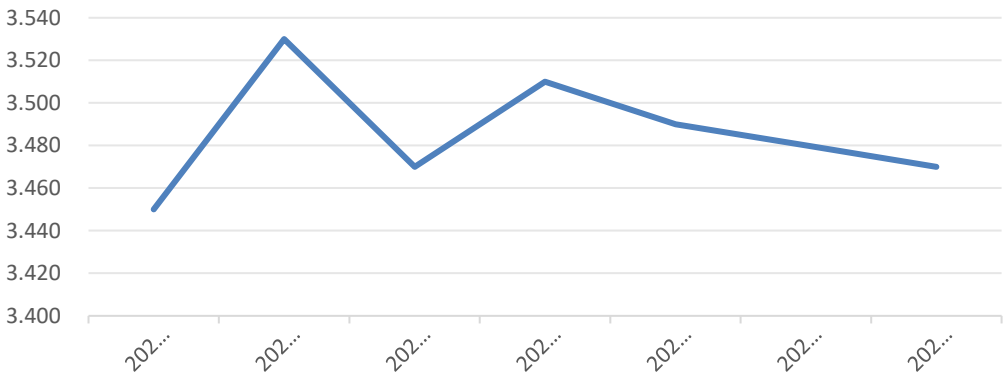
* 计算方法说明：以上数据均来源于 silicon data。每条 GPU 租赁记录首先会被转换为标准化的每 GPU 每小时费率——即单价等于总租赁价格除以（GPU 数量乘以租赁时长）。预处理后的数据会经过多阶段过滤，以剔除上报错误，同时不会忽略真实的结构性价格变动。由于单纯汇总价格会随每日配置组合的变化而波动，因此我们采用基差调整（basis adjustment）将每条记录标准化为基准等价合约，在此过程中会综合考虑租赁类型、地理位置、CPU 平台以及 GPU 型号等因素。经基差调整后的价格会在提供商层面进行汇总，最终的指数值是各提供商数据的加权平均值。

图表 3：2026.10.19 -2026.10.25 日 H100 价格（单位：美元/小时）



资料来源：silicon data，华鑫证券研究

图表 4：2026.10.19 -2026.10.25 日 H200 价格（单位：美元/小时）

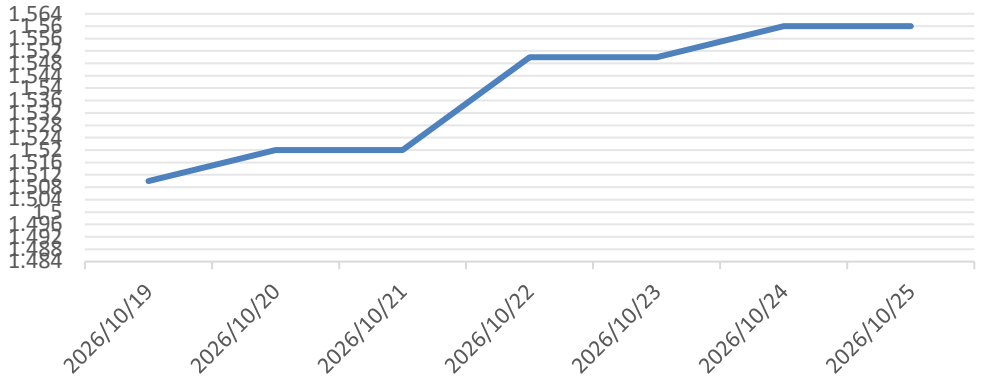


资料来源：silicon data，华鑫证券研究

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

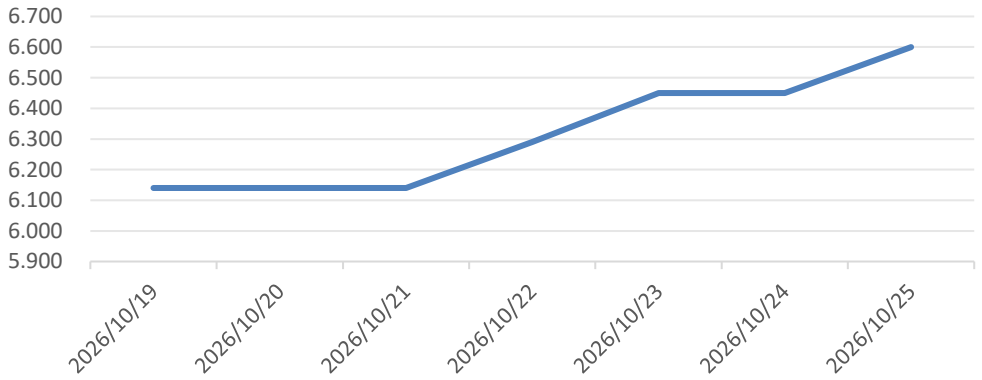
下载日志已记录，仅供内部参考

图表 5：2026.10.19 -2026.10.25 日 A100 价格（单位：美元/小时）



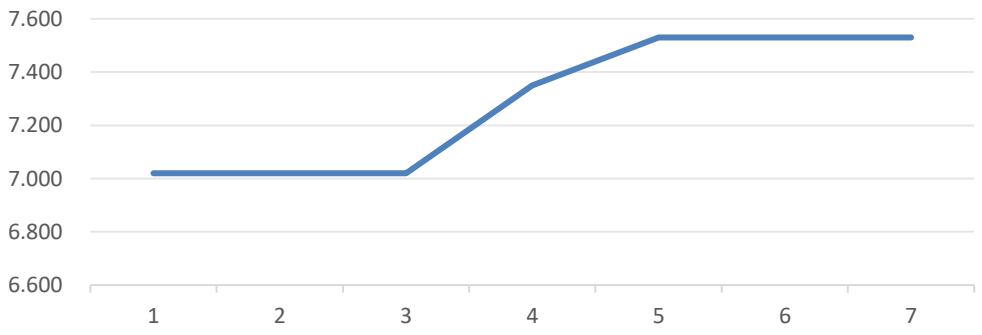
资料来源：silicon data，华鑫证券研究

图表 6：2026.10.19 -2026.10.25 日 B200 价格（单位：美元/小时）



资料来源：silicon data，华鑫证券研究

图表 7：2026.10.19 -2026.10.25 日 B300 价格（单位：美元/小时）



资料来源：silicon data，华鑫证券研究

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

1.3、 云厂动态：阿里云 Meoo 平台接入千问 App，为自然语言生成应用提供全托管云端后台支撑

2026 年 10 月 23 日，阿里云旗下 AI 应用生产平台 Meoo（秒悟）正式接入千问 App，为千问 Go 新上线的“我的应用”功能提供云端后台服务，全面覆盖 AI 应用生成、数据存储、部署上线等关键环节。用户无需编写代码，仅需使用自然语言描述业务需求即可快速创建 AI 应用；借助阿里云 Meoo 的开放云服务能力，生成的应用能够直接发布上线、分享给他人使用，并具备长期稳定运行与持续更新能力。与以往仅能用于展示基础效果的本地应用 Demo 不同，Meoo 将文件存储、数据记录、用户区分和对外发布等云端后台能力打包输出，推动 AI 应用从概念原型向真正可用、可长期运营的实用产品演进。

业务场景落地与后台能力支撑方面，Meoo 为千问 Go 生成的各类实用工具提供了底层数据库与文件管理支撑。以教育场景为例，有家长依托千问 Go 搭建了“亲子导览”应用，通过选定路线与年龄生成“行前学习包”并设置打卡任务；在该应用运行背后，Meoo 负责存储“行前学习包”文件，通过数据库记录访问痕迹与打卡进度，并支持将应用生成为独立的可分享链接，在应用上线后支持用户随时修改功能且持续保留和积累历史数据。此外，依托 Meoo 的后台能力，用户还可快速搭建如健身教练专属约课工具（支持学员自主预约课程）、朋友圈“周末搭子”投票工具（发起游玩项目并投票统计）以及社群“活动报名页”（统一完成报名信息收集）等多元化轻量实用小工具。

开放生态与测试进展方面，此次接入千问 Go 系 Meoo 开放平台能力的一次标杆落地。Meoo 依托 OpenAPI 向各类 AI 产品与平台输出应用预览、发布上线、数据存储等一站式云服务能力，合作伙伴无需自建复杂的云端后台与运维体系，即可在自有产品中无缝集成 AI 应用创建能力，并可按需自主选择是否展示 Meoo 品牌。目前，千问 Go“我的应用”功能，以及支持直接操作云手机、云电脑的“云端 PRO”能力正处于邀测阶段，用户可在千问 App 内的邀测页面申请体验，千问团队正加速向申请用户开放测试权限。

1.4、 产业动态：MiniMax 发布轻量模型 M3.1 Flash，低成本高吞吐与 DeepSeek 形成国产“双强”格局

2026 年 10 月 19 日，MiniMax 正式认领此前在社区火爆的匿名模型“Space Bunny”，并正式发布面向 Coding 与 Agent 优化的高效轻量级模型 M3.1 Flash。该模型具备 100 万 Token 的超长上下文窗口，在匿名测试期便已在 OpenCode 调用量榜单稳居第二并在 OpenRouter 上一度登顶第一。在模型架构与推理效能方面，MiniMax 团队进一步优化了 M3.1 Flash 的推理效率，采用更激进的稀疏注意力架构，日常使用场景下的输出解码速度可达 200+ TPS（Tokens per Second），中位数（P50）稳定达到 170+ TPS。此次发布距离 DeepSeek 于 10 月 17 日上线 DeepSeek V4.1 Flash-1017 仅相隔约 48 小时，两款国产轻量高性价比模型的集中落地，推动国内大模型在 Flash 段位形成了正面对决的“双强”并立格局。

基准测试与多任务性能方面，M3.1 Flash 在 Agentic 任务、编程、Computer Use 及网络安全等领域实现大幅提升，多项核心指标追平甚至超越行业旗舰大模型。在考察智能体长程、多步骤真实任务的 Agents' Last Exam 评测中，M3.1 Flash 跑出了 29.6%的通过率和

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

739 元的任务成本，处于“低成本-高通过率”的最前沿区间，性能远超同价位模型；在考察真实业务流程跨应用自动化的 Automation-Bench v1.0.6 评测中，M3.1 Flash 取得 62.2 分，反超 DeepSeek V4.1 Flash-1017 的 60.5 分，并将 Claude Opus 5 甩在身后。在代码编程与修复领域，双方各有侧重：在 DeepSWE v1.1 软件工程测试中，DeepSeek V4.1 Flash-1017 以 77.0 分位居榜首，M3.1 Flash 取得 73.7 分（距离 Opus 5 仅差 0.3 分）；而在代码修复测试 SWE Refore Bench 中，M3.1 Flash 拿到 23.0 分，成绩达到 DeepSeek 的两倍以上。在专业任务与安全领域，M3.1 Flash 在金融工具调用测试 BankerToolBench 中以 76.2 分与 Opus 5 并列第一；在网络安全测试 CyberGym 中以 82.2 分与 GLM-5.3 同处第一梯队；在无工具学术推理测试 HLE w/o tools 中取得 43.4 分，略高于 DeepSeek V4.1 Flash-1017。

图表 8：MiniMax M3.1 Flash 与主流前沿模型多维度 Benchmark 评测成绩对比

		MiniMax M3.1 Flash	DeepSeek -V4.1-Flash-1017	DeepSeek-V4.1-Flash	Kimi-K3	GLM-5.3	GP
CoWork	Automation-Bench v1.0.6	**62.8**	60.5	54.8	46.7	48.8	
	BankerToolBench	**76.2**	-	73.7	69.1	74.5	
	GDP.pdf AA	20.8	-	12.8	22.0	11.2	
	Terminal Bench 4.0	38.1	42.3	31.2	12.6	37.9	
	LOCA-Bench 256K	**60.3**	-	-	49.7	50.3	
Computer Use	Agents' Last Exam	29.6	-	-	28.3	-	
	OSWorld 2.0	61.7	-	54.3	58.3	-	
Coding	PostTrain Bench v1.1	39.3	-	30.5	36.6	37.5	
	DeepSWE v1.1	73.7	**77.0**	74.2	67.5	66.9	
	NL2Repo-Bench	61.2	68.2	65.4	58.0	58.0	
	SWE Refactor Bench	23.0	-	11.8	19.5	10.4	
网络安全	CyberGym	82.2	-	-	80.0	**84.5**	
通用能力	BabyVision (w/tools)	91.0	92.5	89.6	85.7	-	
	HLE (w/o tools)	43.4	44.0	39.2	43.5	42.3	
	CharXiv RQ	82.5	-	69.0	84.8	-	

资料来源：智东西，华鑫证券研究

商业化定价与成本结构方面，M3.1 Flash 进一步打穿了同类模型的成本底线，其标准输入价格为每百万 Token 1 元，输出价格为每百万 Token 4 元，缓存读取价格为每百万 Token 0.1 元；据 MiniMax 团队发布数据，在 Agentic 类任务中，该模型的综合使用成本较前代模型大幅降低了 70%。对比来看，其输入与输出资费约为 DeepSeek V4.1 Flash-1017 高峰时段价格（未命中输入 2 元/百万 Tokens、输出 8 元/百万 Tokens）的一半，与 DeepSeek 空闲时段价格持平。在实际落地中，两款模型呈现差异化互补：开发者面对代码生成效率

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

与缓存成本控制时可调度 DeepSeek V4.1 Flash-1017；而在处理长程 Agent 任务、跨应用自动化或多模态驱动流程时，M3.1 Flash 则成为高性价比的首选工具。Flash 系列模型逐步打破“低预算折中方案”的固有标签，正在以接近旗舰的综合性能、极高的推理速度和极致的性价比成为企业级生产环境的关键选型。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

2、 AI 应用动态：Deepl 周时长环比+2.21%，Anthropic 推出 Claude Duet

2.1、 周流量跟踪：Deepl 周时长环比+2.21%

本期（2026.10.17-2026.10.23）AI 相关网站流量数据：访问量前三位分别为 ChatGPT（1371.0M）、Bing（852.6M）和 Gemini（602.6M），访问量环比增速第一为文心一言（129.79%）；平均停留时长前三位分别为 Character.AI（00:14:24）、Discord（00:10:36）和 NotionAI（00:07:28）;平均停留时长环比增速第一为 Deepl（2.21%）。

图表 9：2026.10.17-2026.10.23 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	器人	enAI	1371.0	-0.94%	6:34	-0.51%
Bing	搜索	微软	852.6	-0.57%	7:01	-0.71%
Gemini	聊天机器人	谷歌	602.6	-0.72%	6:22	0.26%
Canva	在线设计	Canva	218.5	0.64%	5:34	-0.30%
Github	代码托管	微软	174.5	2.83%	6:29	0.78%
Discord	游戏社区	微软	145.9	-0.95%	10:36	0.16%
Character.AI	聊天机器人	Character.AI	33.36	-4.74%	14:24	-1.71%
Deepl	翻译工具	DeepL	21.59	-2.17%	2:19	2.21%
Perplexity	AI 搜索	Perplexity	19.94	-1.34%	4:19	-0.38%
QuillBot	释义工具	QuillBot	8.322	1.13%	3:01	0.00%
Kimi	聊天机器人	Moonshot AI	4.368	-5.64%	5:34	-0.30%
NotionAI	文本/笔记	Notion	1.35	-20.06%	7:28	-3.24%
文心一言	聊天机器人	百度	0.01	129.79%	0:00	

资料来源：similarweb,华鑫证券研究

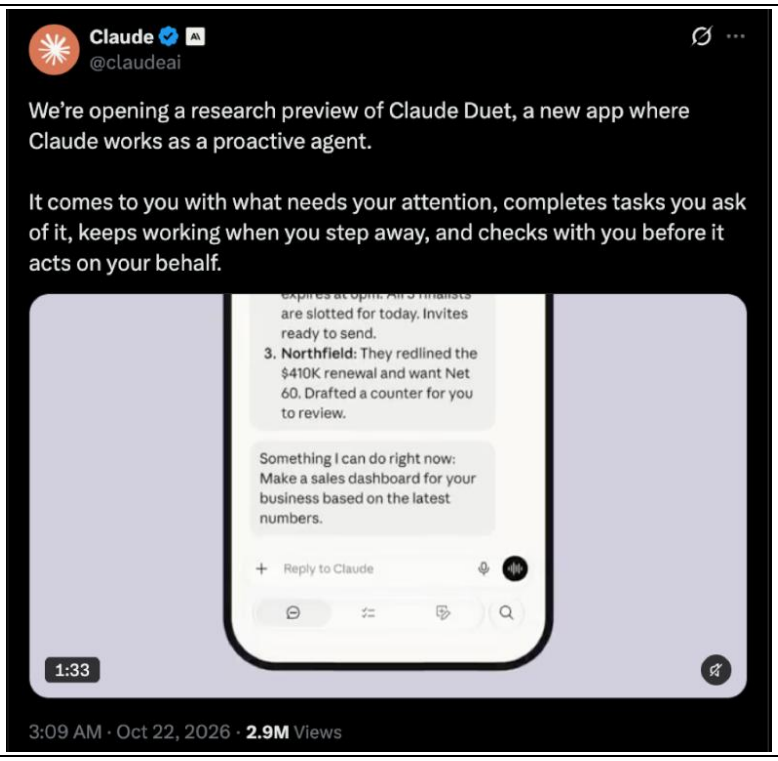
2.2、 产业动态：Anthropic 推出 Claude Duet，以 Opus 5.5 为底座覆盖多场景任务执行

2026 年 10 月 22 日，Anthropic 推出了个人 AI 智能体应用 Claude Duet。该产品以主动式任务处理为核心定位，基于 Claude Opus 5.5 模型构建，具备连接用户邮件、日历、银行账户及企业业务系统等多类信息源的能力。与以往需要用户逐一下达指令的工具不同，Claude Duet 能够在用户尚未提出需求时主动识别待办事项，并持续跟进执行。目前，该产品已推出 iOS、Android 及网页端独立应用，首批面向美国地区 Pro 和 Max 订阅用户开放研究预览候补申请，尚未全面铺开。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

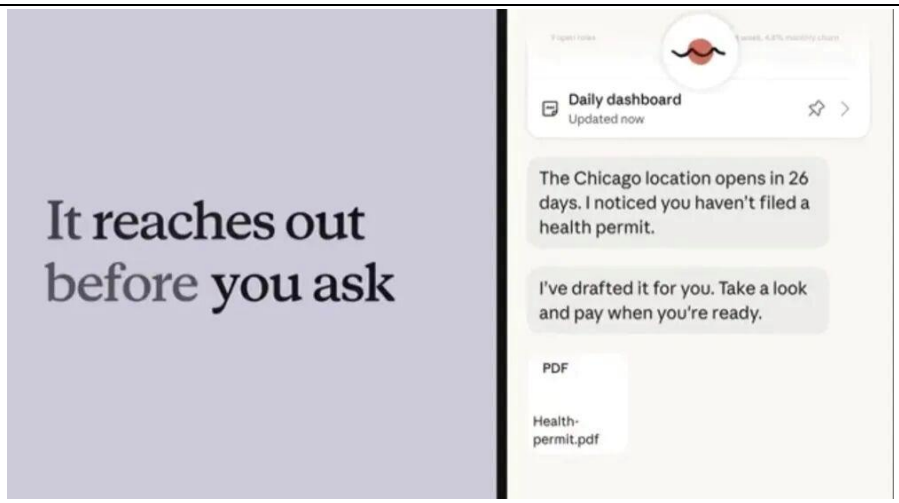
图表 10：Claude Duet 发布



资料来源：智东西，华鑫证券研究

从功能设计来看，Claude Duet 的特点是提前介入用户的工作流程。该应用每天清晨会主动推送一份工作简报，从已连接的各类应用中汇总当日重要事项和需要尽快作出的决策。例如，一位经营咖啡公司的用户收到了 Claude Duet 的工作简报，它不仅列出事项，还预先准备了许可证申请文件、面试安排和合同还价草案等配套材料。当用户授权后，它能够自行从销售终端、线上商店、客户关系管理系统、客服系统、薪资系统及银行账户中提取数据，生成销售仪表盘，免去用户逐个平台下载整理的繁琐操作。除此以外，它还会主动提示时间节点。在执行过程中，Claude Duet 能够根据进展调整优先级，对暂不紧急的事项给出延后建议。招聘场景中，它可在数小时内跟进面试进程，汇总面试官反馈，指出首选候选人及其手持其他录用通知的情况，并在用户确认后调取薪酬方案、起草录用邮件。

图表 11：Claude Duet 可以提前介入用户的工作流程



资料来源：智东西，华鑫证券研究

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

在工具接入与运行机制方面，Claude Duet 可调用 Claude Marketplace 中超过六千个连接器和插件，覆盖企业日常使用的多种软件服务。它能够从 Gmail 和 Slack 中查找遗漏的重要信息并起草回复，通过 QuickBooks 或 Stripe 核查未付发票并准备催款提醒，在 Gusto 填写工时、奖金及报销信息，还可起草招聘信息经用户审核后发布至领英，利用客户关系管理记录撰写跟进邮件，通过 Docusign 准备协议，借助 Canva 将产品发布计划转化为社交媒体营销活动。对于经营网店的商家，它能够连接 Shopify 协助店铺设计与管理。支付环节则通过数字钱包完成授权，使用虚拟卡付款，避免直接向商家提供真实银行卡信息。值得关注的是，Claude Duet 拥有独立的虚拟机和浏览器，用户交代任务后，它可自行判断调用哪些应用、执行哪些操作，并在后台持续运行，即使设备关闭，任务也不会中断，完成或需审阅时会向用户发送通知。该应用还继承了用户既有账户中的记忆、技能和已连接应用，支持文字与语音输入，用户可在同一条持续对话中安排任务、查看进展并补充要求。不过，目前它尚不能向用户展示浏览器操作过程。

安全与成本控制是 Claude Duet 产品设计中的重要考量。由于涉及邮件、银行及企业应用等敏感连接，该产品不会直接接触用户密码或登录令牌。在读取信息、起草文件和准备操作方面，Claude Duet 可自主进行，但涉及对外发送消息、付款、签署文件或删除内容时，必须先展示即将执行的操作并等待用户确认；部分操作，例如证券交易和账户间资金转移，即使获得批准也不会执行。针对敏感信息外泄风险，系统配备了专用分类器和安全提示词，对信用卡号、政府签发身份识别号码等最高敏感级别信息拒绝分享请求，健康、个人财务等信息仅在用户明确指示时允许传递。其安全基础设施已接受内部红队和外部安全公司的渗透测试，并纳入公开漏洞悬赏计划。在主动通知机制上，Claude Duet 并非不间断监控所有连接账户，而是每隔数小时检查一次，主要原因是持续检查会更快消耗用户套餐额度，用户也可自行设置通知范围和频率。该产品与 Claude 主产品共享 Pro 和 Max 套餐的使用额度，后台定期检查会消耗额度，制作财务模型、演示文稿等复杂任务因调用更多轮模型推理而产生更高消耗。此次预览版仅面向美国 Pro 和 Max 用户开放候补申请，需等待官方邮件通知方可获得使用资格，产品目前以独立应用形式提供，尚未整合进 Claude 主应用。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

3、AI 融资动向：AI 推理芯片设计商 Tenstorrent 完成 15 亿美元 E 轮融资，投后估值达 62 亿美元

2026 年 10 月 19 日，AI 推理芯片设计商 Tenstorrent 宣布完成 15 亿美元 E 轮融资，投后估值达 62 亿美元。本轮由 Trace Capital 与 Innovation X 领投，ARK Invest、富达、大摩、T. Rowe Price 及 Tether Investments 等 15 家机构参投。此前在 5 月，彭博社曾披露 Intel 与 Qualcomm 就潜在收购进行谈判（估值或超 50 亿美元），6 月 Qualcomm 谈判估值达 80 至 100 亿美元；CEO 吉姆·凯勒确认在跑通自研系统后回绝收购意向并坚持独立融资，本轮估值较 2024 年 12 月 D 轮上涨约 2.4 倍。商业化方面，日本 ai&已部署超 120 套 Galaxy 系统构建主权 AI 基建，Virtu Financial 在本地部署 Galaxy 运行交易 Agent，Equinix 与 Smallest.ai 亦已接入；IP 授权业务涵盖三星、LG（车载与电视芯片）、现代汽车集团，并向日本 Rapidus 2nm 项目提供 RISC-V CPU 小芯片。

技术架构上，Tenstorrent 主打以 RISC-V 开源指令集结合大片上 SRAM 与 GDDR6 显存替代高昂的 HBM，从底层破解大模型解码阶段的显存带宽瓶颈。芯片基于完全可编程的 Tensix 核心，旗舰 Blackhole 采用三星 4nm 小芯片工艺，集成 140 核、32GB GDDR6 及 768 个 RISC-V 核心；Galaxy 服务器采用 6U 机箱集成 32 颗 Blackhole 芯片，提供 23 PFLOPS 的 Block FP8 算力，实现 Prefill 与 Decode 同构处理并以以太网替代高昂专有互联。实测方面，Galaxy 服务器运行 DeepSeek R1 输出达 350+ tokens/s（EE Times 实测为 255 tokens/s），Token 成本为 6 美元对标竞品 30 美元，后续实测 DeepSeek-R1-0528 单用户超 400 tokens/s、Kimi K2.6 达 900 tokens/s/user（达 GPU 的 3 倍）。挑战方面，受制于早期内核软件优化滞后，实测显示 Blackhole 仍存在 76 个核心闲置，团队正加速推进 A0 芯片出货后的软件栈除错与算力释放。

图表 12：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
Tenstorrent	AI 推理芯片研发	Trace Capita l& Innovation X	E 轮	15 亿美元	暂未披露	62 亿美元

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

Flapping Airplanes	GPU 内核框架	Index Ventures & Lux Capital	A 轮	10 亿美元	暂未披露	60 亿美元
Warp Sciences	AI 自动化 CRO	Lightspeed Venture Partners	A 轮	2.15 亿美元	暂未披露	8.65 亿美元

资料来源：Saasverse，华鑫证券研究

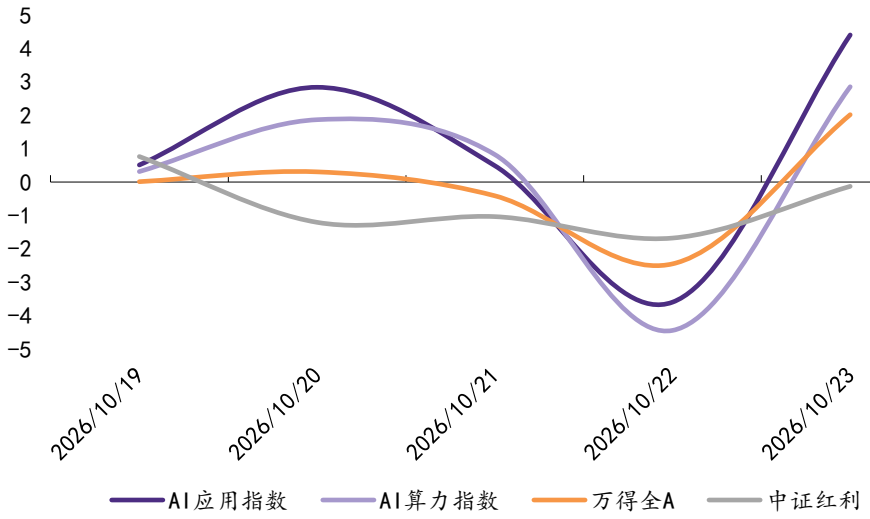
Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

4、 行情复盘

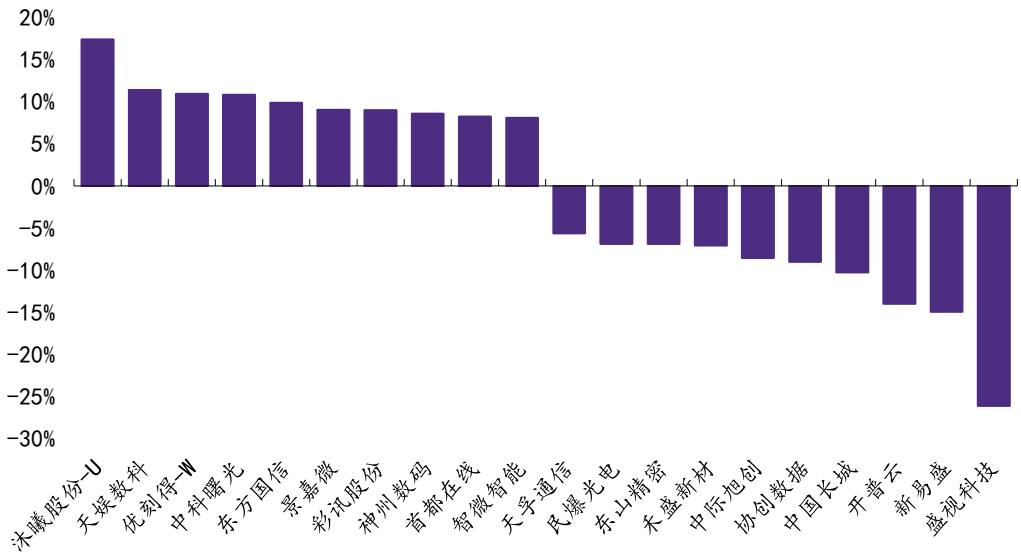
上周（2026.10.19-2026.10.23 日），AI 应用指数/ AI 算力指数/万得全 A/中证红利日涨幅最大值分别为 4.41%/2.85%/2.02%/0.76%，AI 应用指数/AI 算力指数/万得全 A/中证红利日跌幅最大值分别为-3.6%/-4.45%/-2.45%/-1.68%。AI 算力指数内部，沐曦股份-U 以 17.38%录得上周最大涨幅，盛视科技以-26.12%录得上周最大跌幅。AI 应用指数内部，海天瑞声以 79.98%录得上周最大涨幅，华盛昌以-16%录得上周最大跌幅。

图表 13：上周（2026.10.19-2026.10.23 日）指数日涨跌幅



资料来源：wind，华鑫证券研究

图表 14：上周（2026.10.19-2026.10.23 日）AI 算力指数内部涨跌幅度排名

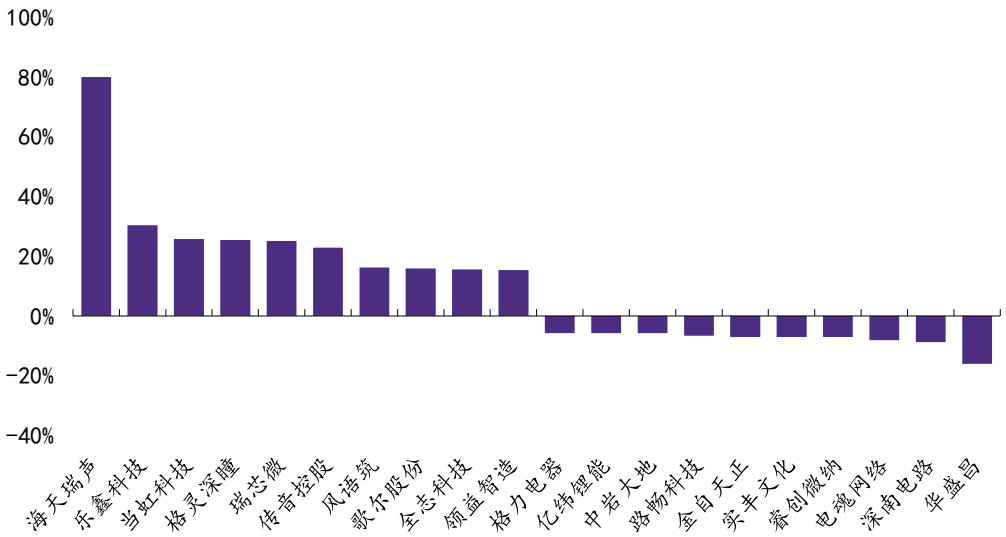


资料来源：wind，华鑫证券研究

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

图表 15：上周（2026.10.19-2026.10.23 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

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

5、 投资建议

10月21日，英伟达在GTC柏林宣布大幅扩展Isaac开放机器人平台，覆盖从数据生成、模型训练到部署运行的完整链路：发布开源运行时 IsaacOS，机器人学习框架 IsaacLab3.0 正式可用；上线包含 2,000 余项仿真资产的 SimReady 中央库与 Human1M 人类活动数据集，后者提供多视角视频与三维手部追踪数据；同时开放三套基于 DGXStation 的参考工作流，支持以强化学习微调 GR00T 模型，并结合 Cosmos 世界模型生成合成数据。AgileRobots 同步发布基于 IsaacGR00T 与 JetsonThor 构建的平台 AgileONESN，计划 2027 年初供货，并将集成 Halos 安全栈；ABB、库卡、NEURA、UniversalRobots 等欧洲机器人厂商均已基于英伟达开放平台开发产品。黄仁勋在会上表示，机器人行业的"ChatGPT 时刻"正在临近。

本次发布的核心信号在于物理 AI 的竞争重心正从本体与运动控制能力，转向数据、训练与部署的工具链体系。此前行业的主要瓶颈在于真实交互数据稀缺、仿真到现实的迁移成本高企，英伟达此次同时补上数据资产、学习框架与运行时三层缺口，机器人开发开始被纳入标准化的流水线，厂商之间的竞争将更多取决于各自场景的数据积累与工程化能力。我们认为，物理 AI 当前正从单机演示走向可复制、可验证的开发范式，数据闭环与工具链将成为区分厂商长期竞争力的关键变量。

中长期，建议关注专注于功率半导体及模拟芯片测试系统研发的联动科技（301369.SZ）、专注于半导体等高端制造业的罗博特科（300757.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）、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 16：ficonTEC2025 年年中至今公告订单

签约日期	客户/描述	业务类型	金额	折合人民币
2025/6/20	美国某头部公司 A 及其子公司	光电子封测设备	约 1,710 万欧元	约 1.36 亿元
2025/7/11	美国某头部公司 B 及其子公司	光电子封测设备	约 1,418 万美元	约 0.98 亿元
2025/9/3	瑞士某头部公司 C 的子公司	全自动硅光子封装整线设备或服务	约 946.50 万欧元	约 0.75 亿元
2025/10/21	武汉驿路通科技股份有限公司	光纤预制及组装线相关自动化设备	约 900 万美元	约 0.62 亿元
2026/1/6	瑞士某头部公司 C 的子公司	第二条全自动 OCS（光交换机）封装整线设备及服务	约 770.00 万欧元	约 0.61 亿元
2025/9/24-2026/1/26	以色列的纳斯达克上市的头部公司 E	单面晶圆测试设备及服务	约 921.60 万美元	约 0.64 亿元
2026/3/13	暂未披露	双面晶圆测试设备及服务	约 608.09 万欧元	约 0.48 亿元

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

2026/3/19-2026/3/25	纳斯达克上市的公司 F 及其子公司	耦合设备及服务（可用于可插拔硅光高速光模块封装制程核心环节的量产）	约 6 亿元人民币	约 6 亿元
2026/4/1	纳斯达克上市的公司 F	耦合设备及服务（可用于可插拔硅光高速光模块封装制程核心环节的量产）	约 3,570 万美元	约 2.46 亿元
2026/4/8-2026/5/1	纽约证券交易所上市的公司 B 的子公司	耦合设备及相关服务	约 2680 万美元	约 1.83 亿元
2026/4/8-2026/5/1	纳斯达克上市的公司 F	视觉检测设备、高精度激光 bar 条封装设备及相关服务	约 3226 万美元	约 2.20 亿
2026/7/23	公司 H 的子公司	光学元器件量产的自动化制造设备及服务	约 1674 万欧元	约 1.29 亿
	总金额			约 19.22 亿元

资料来源：Wind，公司公告，华鑫证券研究

图表 17：重点关注公司及盈利预测

公司代码	名称	2026-10-27	EPS			PE			投资评级
		股价	2025	2026E	2027E	2025	2026E	2027E	
300757.SZ	罗博特科	373.79	-0.40	2.13	4.49	-934.48	175.49	83.25	买入
301196.SZ	唯科科技	87.18	2.05	1.92	2.39	39.28	43.29	34.75	买入
301369.SZ	联动科技	80.03	0.48	1.08	2.62	166.73	74.10	30.55	买入
603859.SH	能科科技	41.94	0.92	1.05	1.32	45.59	39.94	31.77	买入
688615.SH	合合信息	92.56	3.24	4.22	5.25	28.57	21.93	17.63	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。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

中小盘&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

何春玉：金融学士、理学硕士，2023 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。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster

下载日志已记录，仅供内部参考

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。

Remove Demo Watermark from http://www.wancaiinfo.com/pdfmaster