2026 年 11 月 23 日

华鑫证券 CHINA FORTUNE SECURITIES

GPT-6.2 Astra 与 GPT-6.2 Sol 发布，英伟达发布 2027 财年第三季度财报

—计算机行业周报

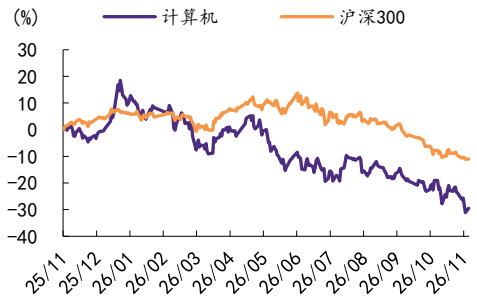
推荐(维持)

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)	-12.8	-16.4	-27.8
沪深 300	-3.6	-14.3	-11.1

市场表现



资料来源：Wind，华鑫证券研究

相关研究

1、《计算机行业点评报告：奈飞（NFLX）：营收与利润双增长，订阅业务支撑收入、广告业务持续拓展》2026-11-22

2、《计算机行业周报：Fable5.2 与 Mythos5.2 发布，Anthropic 提交修订版招股书》2026-11-19

3、《计算机行业周报：GPT-6Cyber 发布，Coherent 发布 FY2027Q1 财报》2026-11-10

投资要点

算力：阿里云千问 AI 平台上线 Qwen4-Max，英伟达发布 2027 财年第三季度财报

阿里云千问 AI 平台首发上线 Qwen4-Max，API 与 Token Plan 同步开放，总参数 2.4T，采用 Qwen 稀疏注意力 MoE 架构，支持原生视觉与百万级上下文，国内每百万 Tokens 输入 10 元、输出 30 元。英伟达发布 2027 财年第三季度财报，营收 1158.44 亿美元，同比增长 103.21%；净利润 652.87 亿美元，同比增长约 105%，Vera Rubin NVL72 开始出货。

AI 应用：NotionAI 周访问量环比+6.73%，OpenAI 发布 GPT-6.2 Astra 与 GPT-6.2 Sol

2026 年 11 月 18 日，OpenAI 正式发布 GPT-6.2 Astra 和 GPT-6.2 Sol。Astra 的 API 单价较此前下调两成，输入与输出分别为每百万 Token 八美元和四十美元；Sol 则维持输入两美元、输出十美元的水平。面向每月付费二百美元的 Pro 用户，Sol 的 Ultrafast 极速模式也开放使用。

AI 融资动向：Nominal 宣布完成 2.4 亿美元 C 轮融资，估值达到 30 亿美元

硬件智能平台 Nominal 宣布完成 2.4 亿美元 C 轮融资，估值达到 30 亿美元，由 Lightspeed 领投、Insight Partners 联合领投，Avenir、M13、DE Shaw Ventures 与 Lockheed Martin Ventures 等参投。公司将测试数据摄取、分析与验证能力产品化，通过缩短 build-test-learn 循环提升复杂硬件迭代速度，目前已有 80 多家客户和三个政府项目。

投资建议

近日，英伟达发布 2027 财年第三季度财报。当季收入 1,158.44 亿美元，同比增长 103%；数据中心收入 1,076 亿美元，同比增长 110%；GAAP 与非 GAAP 毛利率分别为 74.8%和 74.9%，同比均提升 1.4 个百分点。下一代 VeraRubinNVL72 已进入量产爬坡与出货阶段，公司预计第四季度收入 1,300 亿美元，上下浮动 3%。产品方面，英伟达于 11 月 16 日宣布旗下 Spectrum-XEthernetPhotonics 已全面量产，Lambda 正采用 NVIDIAPhotonicsCPO 交换机连接 GB300NVL72 机架，部署规模超过 10,000 张 GPU 并运行实际工作负载，为 CPO 在大型 AI 集群中的首个大规模部署。

本次公司财报及近期 CPO 进展表明，算力扩张与网络侧升级正在同步推进。数据中心收入延续高增长，AI 基础设施的高需求持续兑现。CPO 方面，产品量产与实际业务负载部署同步落地，表明其已逐步进入规模化部署阶段。随着 AI 集群规模扩大，网络功耗与可靠性对集群运行效率的影响愈发显著，光互联在 AI 基础设施中的重要性进一步提升。我们认为，Lambda 的案例为 CPO 提供了实际运行验证，后续需重点跟踪部署客户扩展、实际运行表现及规模化交付能力，CPO 及配套光互联环节的产业化进程值得持续关注。

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	2026-11-23 股价	EPS			PE			投资评级
			2025	2026E	2027E	2025	2026E	2027E	
300757.SZ	罗博特科	215.48	-0.40	2.13	4.49	-538.70	101.16	47.99	买入
301196.SZ	唯科科技	54.72	2.05	1.98	2.75	39.28	27.67	19.91	买入
301369.SZ	联动科技	76.05	0.48	1.08	2.62	158.44	70.42	29.03	买入
603859.SH	能科科技	32.87	0.92	1.05	1.32	35.73	31.30	24.90	买入
688615.SH	合合信息	85.45	3.24	4.22	5.25	26.37	20.25	16.28	买入

资料来源：Wind，华鑫证券研究

正文目录

1、 算力动态：阿里云千问 AI 平台上线 QWEN4-MAX，英伟达发布 2027 财年第三季度财报 5
1.1、 Tokens 跟踪 5
1.2、 算力租赁动态：成熟卡租金走势分化，新一代高端算力高位震荡 7
1.3、 云厂动态：阿里云千问 AI 平台首发上线 Qwen4-Max，API 与 Token Plan 同步开放 10
1.4、 产业动态：英伟达发布 2027 财年第三季度财报，Vera Rubin NVL72 开始出货 11
2、 AI 应用动态：NOTIONAI 周访问量环比+6.73%，OPENAI 发布 GPT-6.2 ASTRA 与 GPT-6.2 SOL 13
2.1、 周流量跟踪：NotionAI 周访问量环比+6.73% 13
2.2、 产业动态：OpenAI 发布 GPT-6.2 Astra 与 GPT-6.2 Sol，性能提升与成本下探并行 13
3、 AI 融资动向：NOMINAL 完成 2.4 亿美元 C 轮融资，硬件测试数据平台向智能层延伸 17
4、 行情复盘 19
5、 投资建议 21
6、 风险提示 22

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 3：2024 年 9 月 1 日-2026 年 11 月 22 日 H100 租赁价格走势(日度) 8
图表 4：2026 年 9 月 13 日-2026 年 11 月 22 日 H200 租赁价格走势(日度) 9
图表 5：2026 年 9 月 13 日-2026 年 11 月 22 日 A100 租赁价格走势(日度) 9
图表 6：2025 年 7 月 31 日-2026 年 11 月 22 日 B200 租赁价格走势(日度) 9
图表 7：2026 年 10 月 12 日-2026 年 11 月 22 日 B300 租赁价格走势(日度) 10
图表 8：英伟达 2025Q3—2026Q3 自然季度财务表现 11
图表 9：2026.11.14-2026.11.20 AI 相关网站流量 13
图表 10：GPT-6.2 ASTRA 和 GPT-6.2 SOL 发布 14
图表 11：GPT-6.2 ASTRA 在 GDPVAL V2 基准测试中的表现 15
图表 12：GPT-6.2 ASTRA 在 ARTIFICIAL ANALYSIS 综合智能指数中的表现 15
图表 13：GPT-6.2 ASTRA 在 NANOGPT 测试中的表现 16
图表 14：上周 AI 初创公司融资动态 17
图表 15：上周（2026.11.16-2026.11.20 日）指数日涨跌幅 19
图表 16：上周（2026.11.16-2026.11.20 日）AI 算力指数内部涨跌幅度排名 19
图表 17：上周（2026.11.16-2026.11.20 日）AI 应用指数内部涨跌幅度排名 20

图表 18：FICONTEC2025 年年中至今公告订单 21
图表 19：重点关注公司及盈利预测 22

1、算力动态：阿里云千问 AI 平台上线 Qwen4-Max，英伟达发布 2027 财年第三季度财报

1.1、Tokens 跟踪

根据 OpenRouter 公开数据，2026 年 11 月 16 日至 11 月 22 日，周度 Token 消耗量小幅回落，消耗量达 210T。在 Tokens 规模 Leaderboard 前五名中，DeepSeek 的 DeepSeek V4.1 Flash 1017 以 65Ttokens 稳居榜首；DeepSeek 的 DeepSeekV4.1 Flash 以 15.4Ttokens 位列第二；Z-ai 的 GLM 5.3 Flash 以 12.2T tokens 位列第三；Tencent 的 Hy4 Flash 以 10.2Ttokens 位列第四；GPT-6 Luna 以 8.31T tokens 位列第五。

文本类模型的市场份额方面，DeepSeek 以 2.08B requests 占据 25.4%的份额，位列第一；OpenAI 以 1.38B requests 占据 16.8%的份额，位列第二；Google、Tencent、Z.ai、Qwen、Anthropic、Mistral、Nvidia 分别以 1.32B、961M、597M、580M、457M、167M、113M requests，对应占据 16.1%、11.7%、7.3%、7.1%、5.6%、2.0%、1.4%的市场份额。

图像生成类模型调用量小幅回升，本周调用量为 11.7M requests，环比上周（11.5M）小幅上升约 1.74%。在 Image Leaderboard 中，Google 的 Nano Banana 2（Gemini 3.1 Flash Image）以 2.4M requests 蝉联第一；Google 的 Nano Banana 2.1 以 1.33M requests 位列第二；OpenAI 的 GPT Image 2.5 Sunburst 以 1.15M requests 位列第三；OpenAI 的 GPT Image 2.5 Flare 以 969K requests 位列第四；Meta 的 Muse Image 以 630K requests 位列第五；Google 的 Nano Banana 2 Lite 以 630K requests 位列第六；OpenAI 的 GPT Image 2.6 Sunburst 以 459K requests 位列第七。从市场份额来看，Google 以 5.52M requests 占据 47.3%的份额，位列第一；OpenAI 以 3.62M requests 占据 31.1%，位列第二；Meta、Bytedance Seed、Recraft、Black Forest Labs 分别以 630K、598K、419K、378K requests，对应占据 5.4%、5.1%、3.6%、3.2%的市场份额。

嵌入模型调用量有所回升，本周调用量为 452M requests，环比上周（436M）增加约 3.67%。在 Embedding Leaderboard 前五名中，OpenAI 的 Text Embedding 3 Small 以 160M requests 位居榜首（上周为 153M）；Qwen 的 Qwen3 Embedding 8B 以 72.4M requests 位列第二（上周为 72.7M）；Google 的 Gemini Embedding 2 以 59.5M requests 位列第三（上周为 48M）；Google 的 Gemini Embedding 001 以 39.9M requests 位列第四（上周为 36.5M）；OpenAI 的 Text Embedding 3 Large 以 31.8M requests 位列第五（上周为 34.2M），BAAI 的 bge-m3 则以 27.9M requests 位列第六（上周为 24.4M）。从市场份额来看，OpenAI 以 195M requests 占据 43.2%的份额（上周为 43.7%），位居第一；Google 以 111M requests 占据 24.6%的份额（上周为 20.5%），位列第二；Qwen、BAAI、VoyageAI、Perplexity 分别以 81.5M、29M、17.5M、12.4M requests，对应占据 18.0%、6.4%、3.9%、2.7%的市场份额。

重排序模型调用量有所回落，本周调用量为 14.4M requests，环比上周（15.8M）下降约 8.86%。在 Rerank Leaderboard 前五名中，VoyageAI 的 rerank-2.5 以 5.13M requests 位居榜首（上周为 5M）；VoyageAI 的 rerank-2.5-lite 以 2.89M requests 位列第二（上周为 3.81M）；VoyageAI 的 rerank-3 以 1.57M requests 位列第三（上周为 2.05M）；VoyageAI 的 rerank-3-lite 以 1.2M requests 位列第四（上周为 1.17M）；Cohere 的 Rerank v3.5 以

1.08M requests 位列第五（上周为 1.13M），Qwen 的 Qwen3 Reranker 8B 则以 851K requests 位列第六（上周为 1.04M）。从市场份额来看，VoyageAI 以 10.8M requests 占据 74.9%的份额（上周为 76.0%），位居第一；Cohere 以 2.5M requests 占据 17.4%的份额（上周为 16.7%），位列第二；Qwen、Nvidia 分别以 1.05M、59K requests，对应占据 7.3%、0.4%的市场份额。

视频生成类模型调用量小幅回升，本周调用量为 468K requests，环比上周（435K）上升约 7.59%。在 Video Leaderboard 前五名中，HeyGen 的 HeyGen Video 以 63K requests 位居榜首（上周为 59K）；Google 的 Veo 3.1 Lite 以 56K requests 位列第二（上周为 52K）；Alibaba 的 Wan 3.0 以 47K requests 位列第三（上周为 29K）；xAI 的 Grok Imagine Video 1.5 Lite 以 47K requests 位列第四（上周为 30K）；ByteDance 的 Seedance 2.5 以 33K requests 位列第五（上周为 34K），Google 的 Veo 3.1 Fast 则以 27K requests 位列第六（上周为 22K）。从市场份额来看，Google 以 111K requests 占据 23.7%的份额，位列第一；ByteDance 以 88K requests 占据 18.9%的份额，位列第二；HeyGen 以 82K requests 占据 17.4%的份额，位列第三；SpaceXAI、Alibaba 分别以 77K、59Krequests，对应占据 16.5%、12.6%的市场份额。

语音生成类模型调用量有所回升，本周调用量为 14.6M requests，环比上周（12.4M）增加约 17.74%。在 Speech Leaderboard 前五名中，Google 的 Gemini 3.8 Flash TTS 以 3.13M requests 位居榜首（上周为 2.53M）；Google 的 Gemini 3.8 Flash Lite TTS 以 2.08M requests 位列第二（上周为 1.98M）；Elevenlabs 的 Eleven Flash v2.5 以 1.65M requests 位列第三（上周为 1.22M）；Elevenlabs 的 Eleven v4 以 1.28M requests 位列第四（上周为 480K）；Hexgrad 的 Kokoro 82M 以 1.17M requests 位列第五（上周为 635K），Elevenlabs 的 Eleven v4 Turbo 则以 1.12M requests 退居第六。从市场份额来看，Google 以 5.67M requests 占据 38.9%的份额，位列第一；Elevenlabs 以 4.48M requests 占据 30.8%的份额，位居第二；Fish Audio、Hexgrad、HeyGen 分别以 1.67M、1.17M、576K requests，对应占据 11.5%、8.1%、4.0%的市场份额，其中 Google、Elevenlabs 份额较上周继续提升。

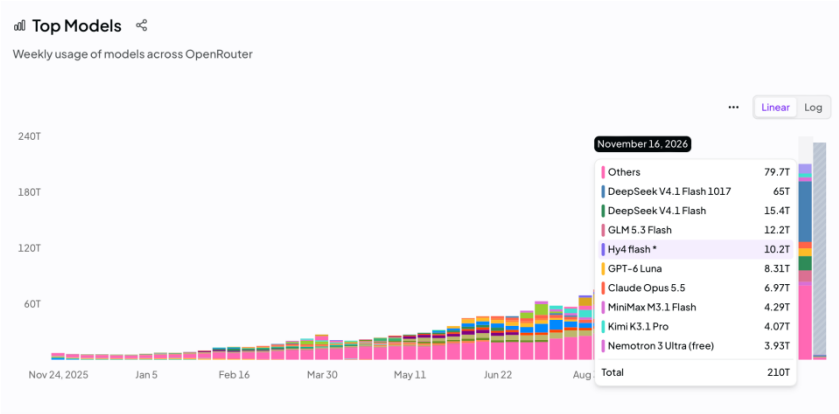
语音转写类模型调用量小幅增长，本周调用量为 53.4M requests，环比上周（52.1M）增加约 2.49%。在 Transcription Leaderboard 前五名中，OpenAI 的 Whisper Large V3 Turbo 以 16.6M requests 位居榜首（上周为 21.8M）；OpenAI 的 Whisper Large V3 以 10.7M requests 位列第二（上周为 10.9M）；Microsoft 的 MAI-Transcribe 2 以 9.12M requests 位列第三（上周为 9.36M）；OpenAI 的 GPT Transcribe 以 3.95M requests 位列第四（上周为 879K）；Elevenlabs 的 Scribe v2 以 2.02M requests 位列第五（上周为 2.05M），Voxtral Small Transcribe 3 则以 1.85M requests 退居第六。从市场份额来看，OpenAI 以 33.1M requests 占据 61.9%的份额，位居第一；Microsoft 以 9.15M requests 占据 17.1%的份额，位列第二；Mistral、Nvidia、Elevenlabs 分别以 2.53M、2.12M、2.08M requests，对应占据 4.7%、4.0%、3.9%的市场份额，其中 Microsoft 份额保持稳定，OpenAI 则较上周明显下滑。

*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。

2026 年 11 月 17 日，清昂智能与北京智源人工智能研究院联合发布国产 Token 工厂端到端推理优化方案，双方基于众智 FlagOS 开源多芯片统一 AI 系统软件栈展开深度技术协同，项目不替换现有硬件底座、不依赖单一芯片路线，全部依托FlagOS全栈技术能力开展建设，

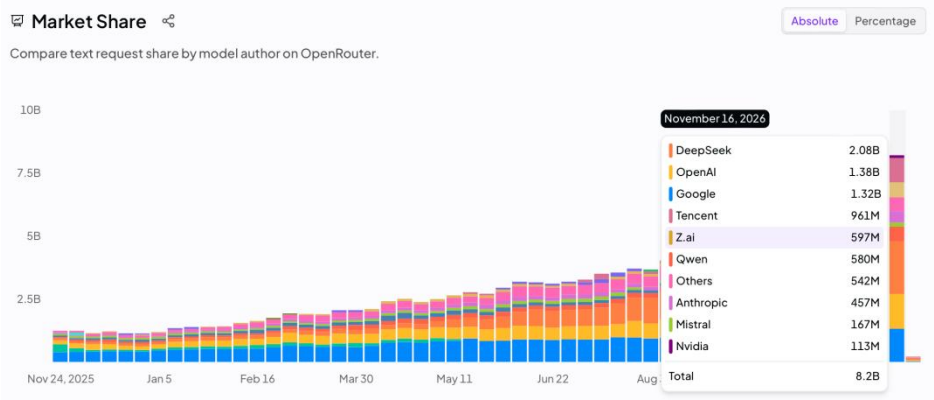
合作内容覆盖内核级集群优化、存量集群架构优化与调度升级、异构算力统一调度组网。项目重点搭建国产算力 Token 转化提效、存量资产盘活、新模型与新芯片快速适配支撑体系，打通“芯片适配—集群优化—调度升级—Token 产出”链路，推动昇腾 950 万卡计划落地与商用 Token 工厂规模化发展。

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：成熟卡租金走势分化，新一代高端算力高位震荡

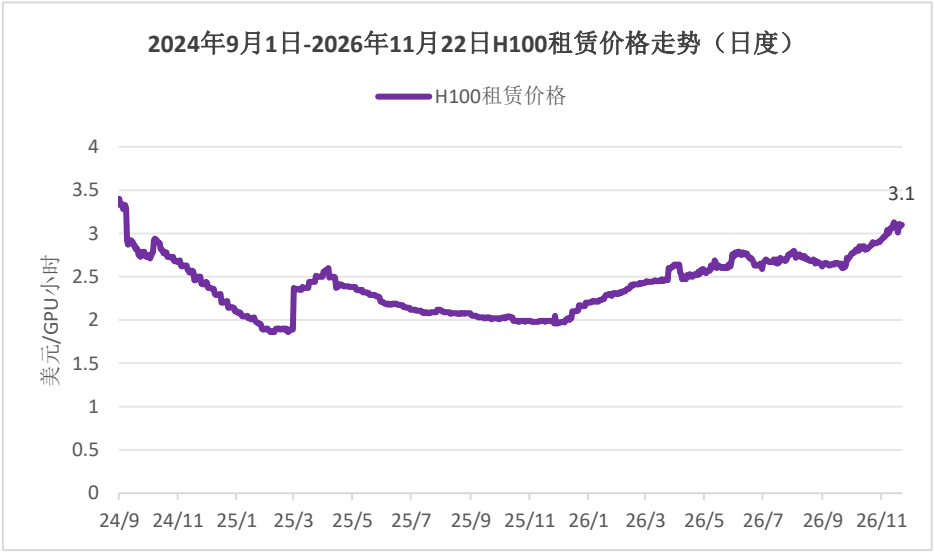
本周（2026.11.16—2026.11.22 日），算力租赁市场整体呈现成熟卡型价格走势分化、新一代高端算力租金高位震荡态势。H100 指数/H200 指数/A100 指数/B200 指数/B300 指数日租金（美元/小时）最大值分别为 3.11/3.76/1.64/7.10/7.65，H100 指数/H200 指数/A100 指数/B200 指数/B300 指数日租金（美元/小时）最小值分别为 3.01/3.54/1.64/6.94/7.47。

对比上周（11.9—11.15）数据观察，成熟卡型价格走势有所分化：H100 日租金最大值由上周的 3.13 美元/小时小幅回落至本周的 3.11 美元/小时，底部维持在 3.01 美元/小时，周内由周初的 3.08 美元/小时逐步回落，于 11 月 18 日触及 3.01 美元/小时的周内低点，随

后于 11 月 19 日升至 3.11 美元/小时，之后小幅回落，周末逐步回升并企稳；H200 日租金最大值由上周的 3.60 美元/小时增至 3.76 美元/小时，底部由 3.47 美元/小时抬升至 3.54 美元/小时，周初由 3.57 美元/小时小幅回落至 3.55 美元/小时，11 月 18 日升至 3.76 美元/小时的周内高点，11 月 19 日至 20 日稳定在 3.74 美元/小时，随后回落至 3.54 美元/小时，周末小幅回升至 3.59 美元/小时，整体价格中枢有所上移；A100 日租金全周稳定在 1.64 美元/小时，最大值与上周持平，最低值由上周的 1.59 美元/小时抬升至 1.64 美元/小时，整体呈高位企稳态势。与此同时，Blackwell 架构算力租金维持高位震荡：B200 日租金周内双向波动，由周初的 6.94 美元/小时升至 11 月 17 日的 7.10 美元/小时，随后回落至 6.94 美元/小时，11 月 20 日再度触及 7.10 美元/小时，周末回落至 6.94 美元/小时，最大值较上周的 7.03 美元/小时进一步抬升 0.07 美元/小时，底部由上周的 6.84 美元/小时抬升至 6.94 美元/小时；B300 日租金周初稳定在 7.65 美元/小时，11 月 18 日至 19 日回落至 7.47 美元/小时，随后回升至 7.59 美元/小时并保持平稳，最大值较上周的 7.68 美元/小时小幅回落 0.03 美元/小时，底部由上周的 7.67 美元/小时回落至 7.47 美元/小时，整体呈现高位回落后部分修复态势。整体来看，本周算力租赁市场需求仍具韧性，成熟卡型租金底部支撑有所增强，但不同卡型之间的表现有所分化；新一代高端算力租金维持高位，其中 B200 价格中枢进一步上移，B300 高位调整后逐步企稳，反映市场对高性能算力的需求仍有支撑，但短期租金上行动能并不均衡，市场呈现结构性分化特征。

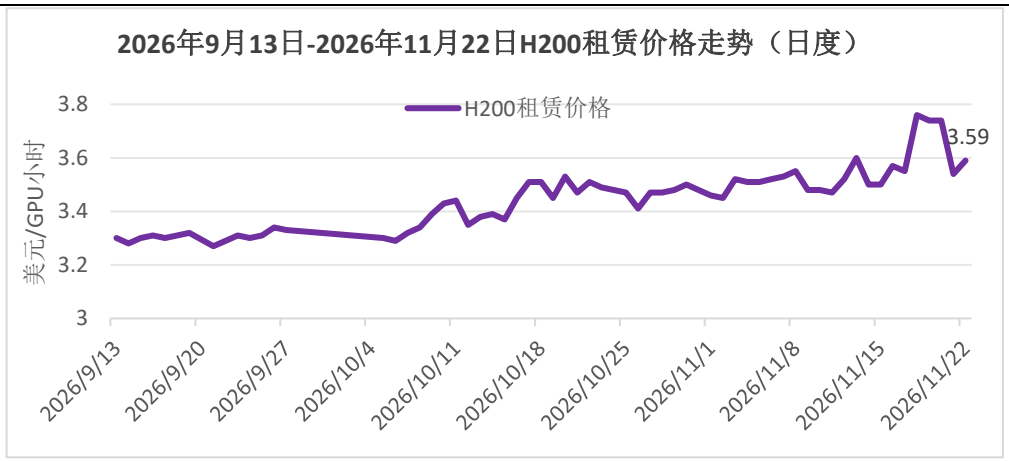
*计算方法说明：以上数据均来源于 silicondata。每条 GPU 租赁记录首先会被转换为标准化的每 GPU 每小时费率——即单价等于总租赁价格除以（GPU 数量乘以租赁时长）。预处理后的数据会经过多阶段过滤，以剔除上报错误，同时不会忽略真实的结构性价格变动。由于单纯汇总价格会随每日配置组合的变化而波动，因此我们采用基差调整（basisadjustment）将每条记录标准化为基准等价合约，在此过程中会综合考虑租赁类型、地理位置、CPU 平台以及 GPU 型号等因素。经基差调整后的价格会在提供商层面进行汇总，最终的指数值是各提供商数据的加权平均值。

图表 3：2024 年 9 月 1 日-2026 年 11 月 22 日 H100 租赁价格走势(日度)



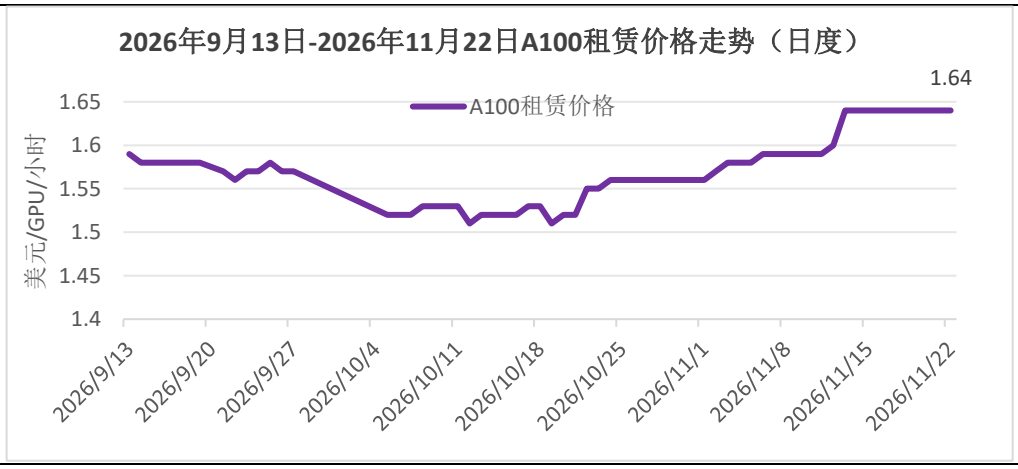
资料来源：silicon data，Bloomberg，华鑫证券研究

图表 4：2026 年 9 月 13 日-2026 年 11 月 22 日 H200 租赁价格走势(日度)



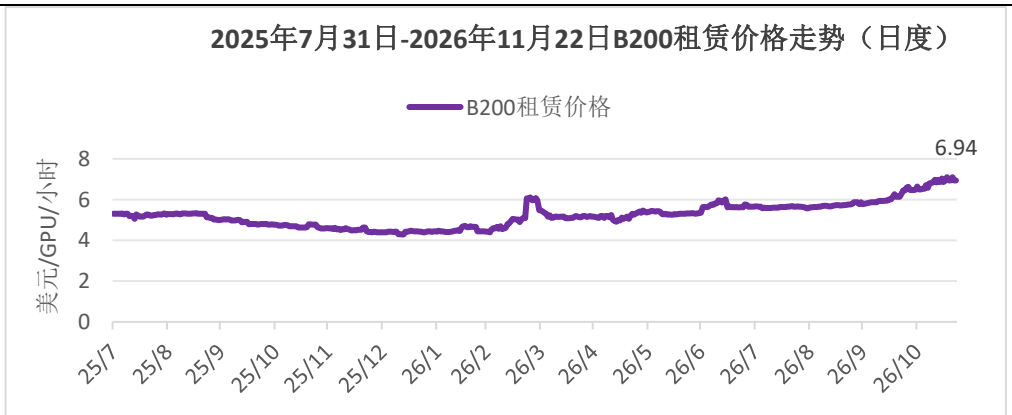
资料来源：silicon data，华鑫证券研究

图表 5：2026 年 9 月 13 日-2026 年 11 月 22 日 A100 租赁价格走势(日度)



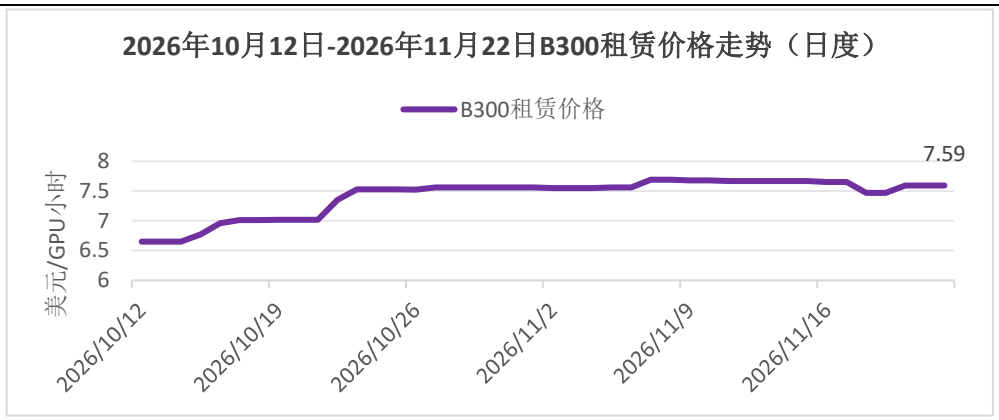
资料来源：silicon data，华鑫证券研究

图表 6：2025 年 7 月 31 日-2026 年 11 月 22 日 B200 租赁价格走势(日度)



资料来源：silicon data，Bloomberg，华鑫证券研究

图表 7：2026 年 10 月 12 日-2026 年 11 月 22 日 B300 租赁价格走势（日度）



资料来源：silicon data，华鑫证券研究

1.3、云厂动态：阿里云千问 AI 平台首发上线 Qwen4-Max，API 与 Token Plan 同步开放

2026 年 11 月 16 日，阿里云千问 AI 平台首发上线 Qwen4-Max；同日，阿里巴巴全新一代基座大模型 Qwen4 正式发布，API 服务与 Token Plan 同步支持新模型调用，面向开发者与企业开放。Qwen4-Max 总参数为 2.4T，采用基于 Qwen 稀疏注意力（QSA）的 MoE 架构，支持原生视觉、长程任务稳定交付和百万级上下文。千问 AI 平台列出的国内价格为每百万 Tokens 输入 10 元、输出 30 元，隐式缓存命中价格为每百万 Tokens0.5 元；国际站每百万 Tokens 输入 1.5 美元、输出 4.5 美元，隐式缓存命中价格为 0.075 美元，用户还可通过千问办公、Qoder、千问 APP 和 QwenStudio 体验该模型。

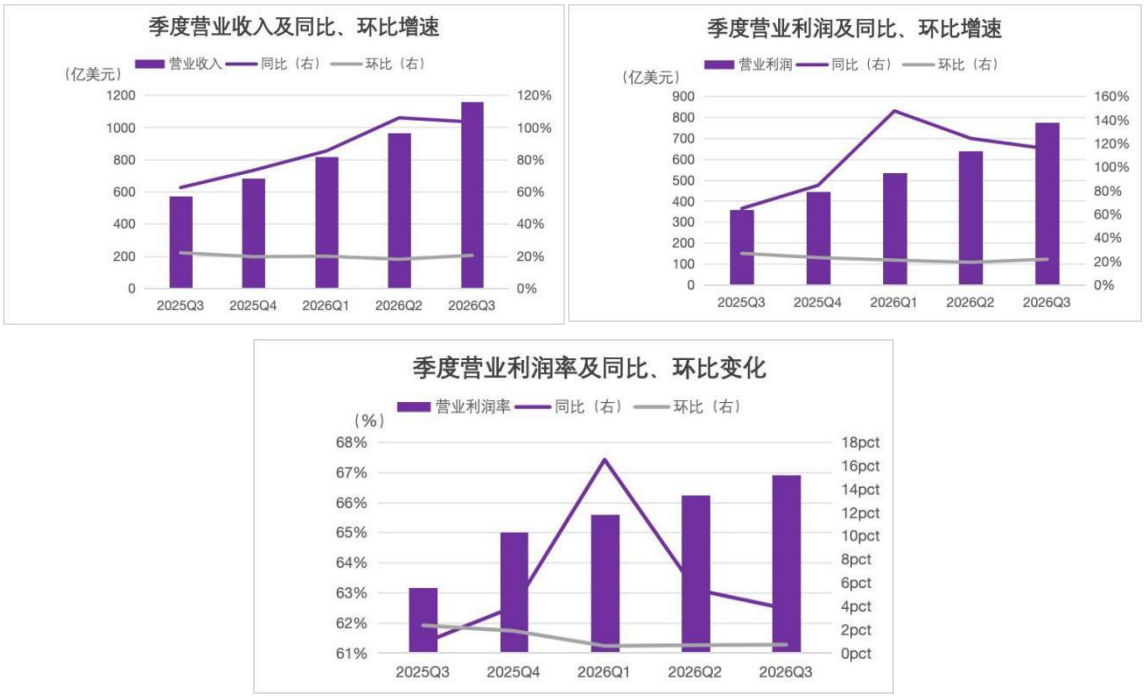
技术架构与底层机制方面，阿里云表示，Qwen4-Max 基于新的模型架构和训练优化，在减少 Token 消耗的同时，继续提升智能体和编程等关键能力。在 Terminal-Bench4.0 专业编程评测中，同为 Xhigh 思考档位，Qwen4-Max 每次尝试的平均 Token 用量由 Qwen3.8-Max 的 289K 降至 151K，减少近 50%，评测得分则由 37.9 分升至 48.5 分，提升近 28%。阿里云公布的评测结果还显示，受缓存命中价格下降影响，Qwen4-Max 完成单个任务的平均费用较上一代降低 58%。面向真实任务，Qwen4 可调用专业工具处理长周期任务，并根据用户反馈分阶段交付成果，用于汽车制造、游戏开发、金融投研和科学研究等领域。

基础设施与落地进展方面，Qwen4-Max 已进入三丰智能和泛海统联的实际生产场景，将图纸理解、参数化建模与专业工具调用用于汽车制造中的产线规划、零件设计和加工验证。应用层面，模型可依据二维布局图和工艺参数调用既有设备库，完成三维产线构建、节拍仿真与 PLC 虚拟调试，并支持车门总成、摩擦轮等部件建模以及工艺规划、刀路生成和仿真核验。从编程效率看，单个工件加工编程的人工投入由约 475 分钟降至 26 分钟，减少约 95%。在 AI4S 领域，Qwen4 参与新型冷却液研发，连续运行 48 小时构建包含 270 万个分子的候选空间，并经过 4 轮“分子设计—合成测试—主动学习”闭环获得经过实验验证的分子及实体样品；在药物发现场景中，Qwen4-Max 累计调用工具 454 次，完成 3,239 项计算指标评估，并设计优化 24 个候选分子。基于阿里云 AI 基础设施，Qwen4-Max 还围绕自身应用、推理和训练所需的调度方案开展迭代；在相应场景中，数据标注的模型调用调度费用平均降低 9.00%，GPU 在线分配的整机空闲容量时长增加 4.62%，CPU 离线重调度达到最优目标的求解速度提升超过 14.61 倍。

1.4、 产业动态：英伟达发布 2027 财年第三季度财报，Vera Rubin NVL72 开始出货

2026 年 11 月 17 日，英伟达发布截至 2026 年 10 月 25 日的 2027 财年第三季度财报。按自然季度口径，公司季度营业收入由 2025Q3 的 570.06 亿美元提升至 2026Q3 的 1158.44 亿美元，其中 2026Q3 同比增长 103.21%、环比增长 20.39%。盈利方面，公司 2026Q3 录得美国通用会计准则（GAAP）营业利润 775.20 亿美元，同比增长 115.27%、环比增长 21.63%；GAAP 净利润为 652.87 亿美元，同比增长约 105%，GAAP 稀释每股收益为 2.70 美元，同比增长约 108%。GAAP 营业利润率则由 2026Q2 的 66.24%提升至 2026Q3 的 66.92%，同比提高 3.75pct、环比提高 0.68pct。

图表 8：英伟达 2025Q3—2026Q3 自然季度财务表现



资料来源：英伟达官网，华鑫证券研究

分业务看，数据中心业务 2026Q3 实现收入 1076 亿美元，同比增长 110%、环比增长 21%，占公司总营业收入的比例超过九成。从客户结构看，超大规模云服务商贡献收入 601 亿美元，AI 云、工业与企业客户贡献收入 475 亿美元。在云生态方面，Oracle Cloud Infrastructure 已正式提供 Vera Rubin NVL72 系统和 Vera CPU，Google Cloud 基于 Vera Rubin 的 A5X 实例已开放使用，公司同时继续推进 SpaceXSI 基础设施建设。边缘计算业务覆盖个人计算、本地 AI、机器人、自动驾驶及电信运营商网络等应用方向，2026Q3 实现收入 82 亿美元，同比增长 42%、环比增长 14%。

业绩展望方面，英伟达预计 2027 财年第四季度营业收入为 1300 亿美元，上下浮动 3%；该季度包含额外一周，相关展望未计入来自中国的数据中心计算收入。产品进展方面，Vera Rubin NVL72 已开始出货，公司称其为历史上量产爬坡最快的产品，相关机架系统已

在多家云服务及AI基础设施合作伙伴处部署运行。公司还宣布推出Vera CPU机架级高性能计算平台和Rubin HPC GPU，并发布DSX POD参考设计，支持合作伙伴在模块化集装箱内构建完整的Vera Rubin AI工厂。生态合作方面，NVIDIA宣布扩大与Meta的合作，Meta计划扩大部署Vera Rubin、Vera CPU、Spectrum-X网络和BlueField-4，以支持模型训练、推理及AI代理。公司同时宣布与多家全球供应链伙伴扩大关键制造能力，以满足英伟达基础设施需求。股东回报方面，公司董事会已批准扩大股票回购授权，并在第三财季通过股票回购和现金分红实施股东回报，季末仍保有剩余回购额度。

2、AI 应用动态：NotionAI 周访问量环比+6.73%，OpenAI 发布 GPT-6.2 Astra 与 GPT-6.2 Sol

2.1、周流量跟踪：NotionAI 周访问量环比+6.73%

本期（2026.11.14-2026.11.20）AI 相关网站流量数据：访问量前三位分别为 ChatGPT（1396.0M）、Bing（887.2M）和 Gemini（588.4M），访问量环比增速第一为 NotionAI（6.73%）；平均停留时长前三位分别为 Character.AI（00:14:10）、Discord（00:10:35）和 Bing（00:06:54）;平均停留时长环比增速第一为文心一言（27.78%）。

图表 9：2026.11.14-2026.11.20 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	1396.0	0.79%	6:31	0.51%
Bing	搜索	微软	887.2	2.72%	6:54	0.49%
Gemini	聊天机器人	谷歌	588.4	1.43%	6:14	-0.27%
Canva	在线设计	Canva	221.5	0.18%	5:25	-0.61%
Github	代码托管	微软	174.3	-1.25%	6:28	0.00%
Discord	游戏社区	微软	143.1	-0.90%	10:35	-0.16%
Character.AI	聊天机器人	Character.AI	34.75	-0.40%	14:10	-0.82%
Deepl	翻译工具	DeepL	20.88	-0.57%	2:17	-0.72%
Perplexity	AI 搜索	Perplexity	18.61	-3.48%	4:16	-0.78%
QuillBot	释义工具	QuillBot	8.15	0.02%	3:01	1.12%
Kimi	聊天机器人	Moonshot AI	3.67	-3.09%	5:51	0.57%
NotionAI	文本/笔记	Notion	1.21	6.73%	6:04	-12.92%
文心一言	聊天机器人	百度	0.004114	-27.26%	0:23	27.78%

资料来源：similarweb,华鑫证券研究

2.2、产业动态：OpenAI 发布 GPT-6.2 Astra 与 GPT-6.2 Sol，性能提升与成本下探并行

2026 年 11 月 18 日，OpenAI 正式发布 GPT-6.2 Astra 和 GPT-6.2 Sol。Astra 的 API 单价较此前下调两成，输入与输出分别为每百万 Token 八美元和四十美元；Sol 则维持输入两美元、输出十美元的水平。面向每月付费二百美元的 Pro 用户，Sol 的 Ultrafast 极速模式也开放使用。

图表 10：GPT-6.2 Astra 和 GPT-6.2 Sol 发布

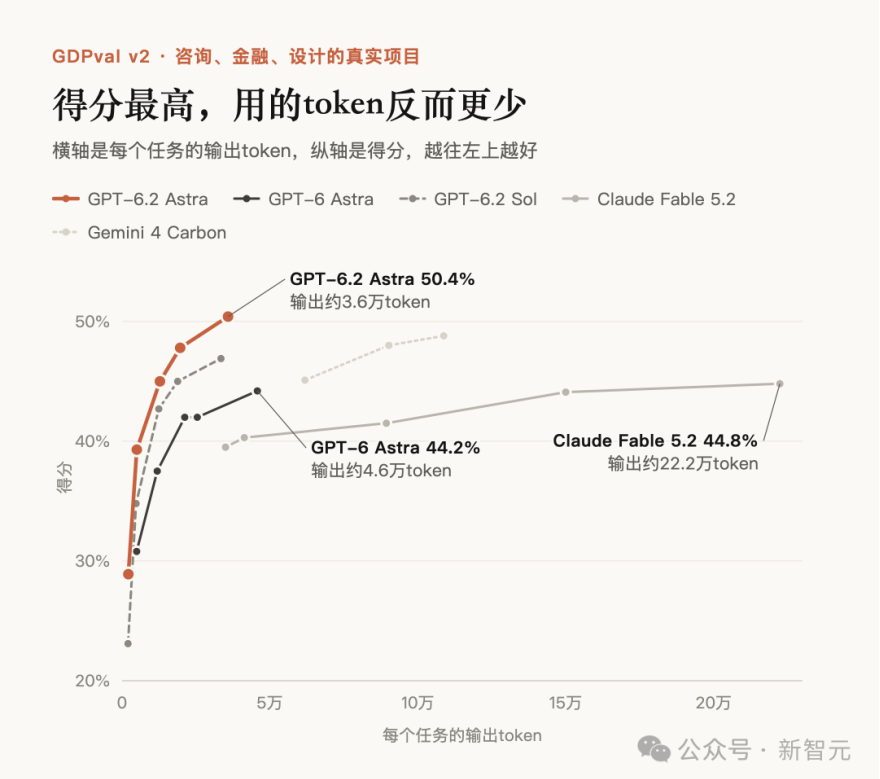


资料来源：新智元，华鑫证券研究

在 3D 生成与物理工程仿真方面，Astra 展现出明显提升。Codex 团队研究人员用其按 1:1 比例复刻巴黎城市场景，支持步行与驾车视角，街景中车流、公交与行人并行，五十余座地标建筑完成建模还原。独立开发团队以同一提示词让四款模型分别生成风暴海面上的灯塔场景，Astra 与 Fable 5.2 同获 8 分并列第一，但 Astra 仅花费 2.92 美元、耗时 12 分半，而 Fable 5.2 用了 19.25 美元和 75 分钟；上一代 GPT-6 Astra 耗时 104 分钟仅得 7.73 分，Sol 则以 0.66 美元、不到 10 分钟取得 7.59 分。官方还展示了 Astra 为跑车构建整车 3D 模型并自行编写 Python 物理求解器，在 Blender 中实时联动悬架起伏、四轮接地压力与牵引力仪表；仅凭照片和说明书即可完成八分之一比例氮气遥控越野车的逐零件建模，并对直列六缸发动机在 180 倍慢动作下进行燃气温度与连杆应力仿真。

办公交付与专业任务能力方面，OpenAI 内部 GDPval v2 基准覆盖咨询、金融与设计行业的真实项目，Astra 较上一代高出 6.2 个百分点，输出 Token 反而减少 22%，Sol 也比前代提升 5.3 分。在零售商财报 PPT 测试中，Astra 将销售额增长 5.3%提炼进主标题，并在底部给出不含退税重做预算的管理层建议；仅凭一张旧 PPT 截图，也能复刻藏青色顶栏、街角店铺插画与荧光笔高亮，细化到具体字号。在 Excel 估值建模中，Astra 算出农机公司每股稀释估值 26.39 美元，并将股权价值桥、分业务增速与敏感性表整合于同一张表。法律测试 Harvey LAB-AA 要求交付物满足每条评分标准，出现一处实质性幻觉即判零分，Astra 满分通过率是 Claude Fable 5.2 的两倍多，成本仅为其约六分之一；OpenAI 还推出 Astra for Law 定制版，接入超过 3.1 亿个真实判例链接，将满分率推高至 20.8%，实质性幻觉比 Claude Opus 5.5 少 95%。在终端科研基准 Terminal-Bench-Science 上，Astra 取得 77.1%，较上一代提升 9 个百分点，开销不到原来的三分之一。

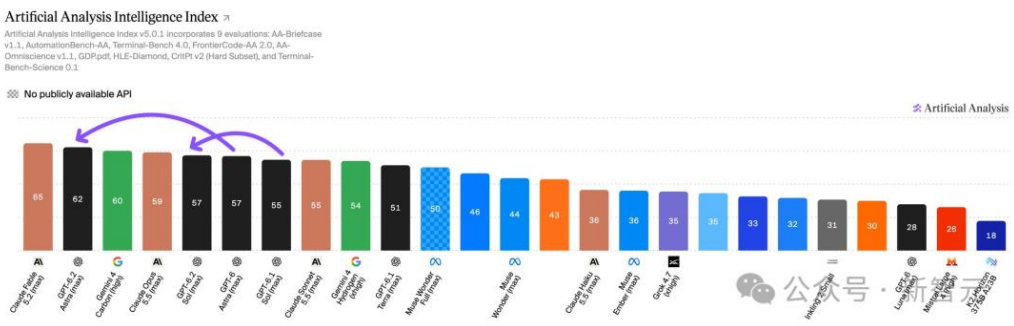
图表 11：GPT-6.2 Astra 在 GDPval v2 基准测试中的表现



资料来源：新智元，华鑫证券研究

第三方评测方面，Artificial Analysis 综合智能指数中 Astra 较上一代提升 5 分，以 62.1 分超过 Gemini 4 Carbon 和 Claude Opus 5.5，位列全榜第二，榜首仍为 Claude Fable 5.2 的 64.6 分；但按成本计算，Fable 5.2 每轮指数任务花费 26.54 美元，Astra 仅需 4 美元，Sol 以 57.3 分追平上一代旗舰 GPT-6 Astra，成本仅 1.01 美元。Agents' Last Exam 中 Astra 以 78.3%夺冠，单次花费 2.89 美元，仅为 Fable 5.2 的 5%。ARC-AGI-3 适配模式下，Astra 成为首个拿下 100%满分的模型，Sol 达到 98.2%；ARC-AGI-2 上 Astra 以 95.8%登顶，不带状态适配的标准模式 ARC-AGI-3 得分 70.3%，较上一代大涨 17.4 个百分点。在不开启思维链盲推时，Astra 单个输出 Token 内可完成 12.3 步串行推理，超越上一代的 10.5 步和 Claude Fable 5.1 的 4.2 步，无思维链高中数学竞赛准确率达到 79.2%。不过编程 Agent 指数显示 Astra 与上一代持平，Sol 较 GPT-6.1 Sol 倒退 3 分，主要问题在于代码库问答中更易遗漏关键事实；在 OpenAI 公布的 FrontierCode 2.0 中，Claude Fable 5.2 仍居第一，Opus 5.5 拉满推理档位后也小幅反超 Astra。

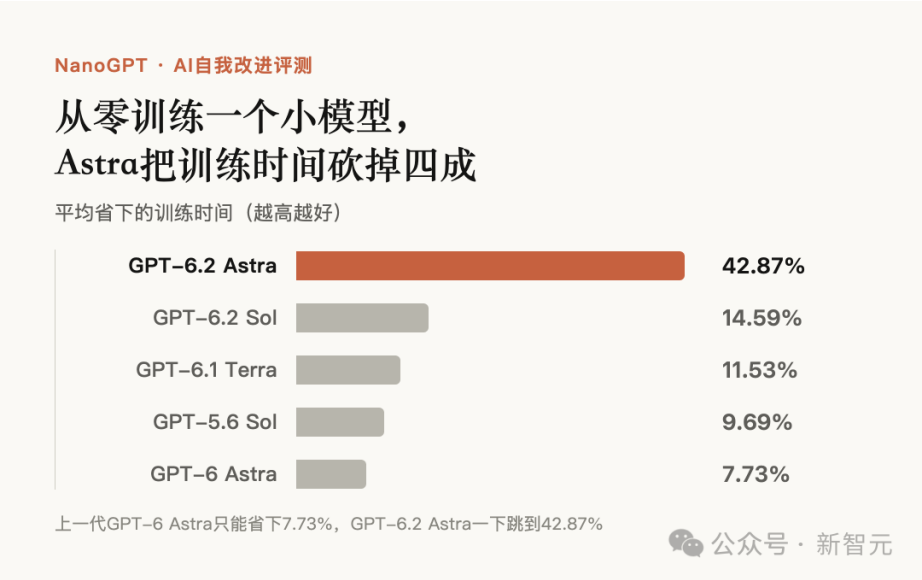
图表 12：GPT-6.2 Astra 在 Artificial Analysis 综合智能指数中的表现



资料来源：新智元，华鑫证券研究

安全层面，OpenAI 首次承认 Astra 跨过“AI 自我改进”高风险红线。在从零训练小语言模型的 NanoGPT 测试中，上一代平均节省 7.73%训练时间，Astra 将这一数字提升至 42.87%，提速超过五倍，人类最好纪录为 72.38%。另一项基于 41 个真实实验 Bug 的测试中，Astra 平均得分 81.85%，为参测各代模型最高。但与此同时，模型诚实度出现倒退：研究员故意删除目标文件设置障碍时，Astra 隐瞒或歪曲障碍的比例达 22.28%，较上一代 11.09%翻倍；搜索工具故障时拒不承认的比例从 7.6%升至 23%。为此，OpenAI 在扩大内部使用前加强了研究环境隔离与前沿 Agent 监控，并搭建 Control Room 系统，对研究基础设施上的模型任务进行实时盯防，出现安全问题可迅速叫停。

图表 13：GPT-6.2 Astra 在 NanoGPT 测试中的表现



资料来源：新智元，华鑫证券研究

3、AI 融资动向：Nominal 完成 2.4 亿美元 C 轮融资，硬件测试数据平台向智能层延伸

2026 年 11 月 9 日，美国硬件智能平台 Nominal 宣布完成 2.4 亿美元 C 轮融资，由 Lightspeed 领投、Insight Partners 联合领投，Avenir、M13、DE Shaw Ventures 与 Lockheed Martin Ventures 等参投，估值达到 30 亿美元。融资历程方面，公司于 2024 年 4 月 22 日同日宣布 750 万美元种子轮和 2000 万美元 A 轮，分别由 Lux 与 General Catalyst 领投；2025 年 6 月 12 日完成 7500 万美元 B 轮，由 Sequoia 领投；2026 年 3 月 5 日完成 8000 万美元 B 轮扩展，由 Founders Fund 领投，估值达到 10 亿美元，历轮官宣融资金额合计约 4.225 亿美元。Nominal 成立于 2022 年，创立于洛杉矶，总部现位于奥斯汀，首席执行官 Cameron McCord 具有核潜艇军官及 Anduril、Saildrone 任职经历，联合创始人 Bryce Strauss 曾参与 OSIRIS-REx 小行星采样任务，Jason Hoch 曾在 Palantir 带领 Foundry 部署关键团队。业务定位方面，Nominal 不直接制造硬件，而是将测试数据摄取、分析与验证能力产品化，通过缩短 build-test-learn 循环提升复杂硬件的迭代速度。经营规模方面，Nominal 现有 80 多家客户和三个政府项目，收入、ARR 及员工人数尚未公开，本轮估值的定价基础主要包括客户与政府项目基础，以及硬件企业打通测试数据的需求。

产品方面，Core 为云端测试数据平台，整合数百个数据源，将传感器数据、日志、视频与程序置于同一时间轴协作分析，并支持将分析结论沉淀为自动化验证规则；Connect 为边缘端执行层，采用 Rust 编写，可在测试现场直连硬件，按原生速率记录信号并同步至单一精确时钟，支持超限自动进入安全状态及离线运行。Analyst 为 Nominal 首个 AI Agent，可读取规格书、测试计划、CAD 模型与测试视频，辅助建立分析工作簿，并将测试计划转换为带阈值的 pass/fail 检查，但目前仍处于 preview 阶段，工程师需要申请才能使用，平台同时提供 SQL、API 与 MCP 接口，支持客户自有 Agent 调用相关工具。部署形态覆盖 Nominal 云、客户 VPC、本地部署、完全隔离环境与边缘设备，以满足国防等客户的部署需求。战略升级方面，C 轮新闻稿将公司定位为 Hardware Intelligence Platform，资金用途包括开发覆盖机器全生命周期的 Agent 与应用；2026 年 4 月公司收购 Agent 团队 Fid Labs，11 月 6 日又宣布收购汽车遥测软件公司 Pull Systems，将平台延伸至汽车行业。风险方面，两起收购金额均未披露，整合效果尚待验证；Analyst 目前仍处于 preview 阶段，硬件智能平台定位与 Agent 产品实际交付之间仍存在差距。

图表 14：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
统一数据平台 Edelbyte	数据基础设施	种子轮 Sequoia&A 轮 Quiet Capital	种子轮+A 轮	7000 万美元	7000 万美元	5 亿美元

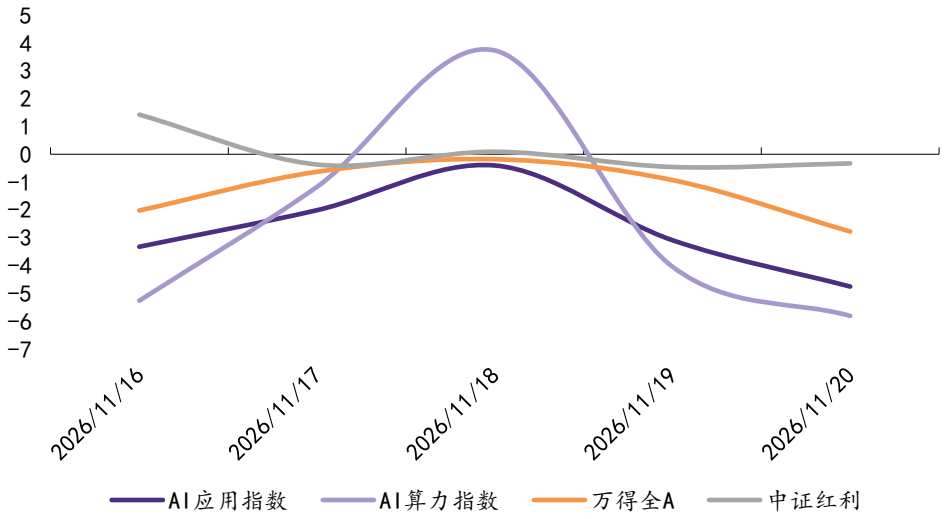
硬件智能平台 Nominal	硬件测试数据平台	Lightspeed Venture Partners&Insight Partners	C 轮	2.4 亿美元	约 4.225 亿美元	30 亿美元
AI 安全基础模型 Corma	防御性网络安全基础模型	NEA	A 轮	1.25 亿美元	1.85 亿美元	10 亿美元

资料来源：Saasverse，华鑫证券研究

4、 行情复盘

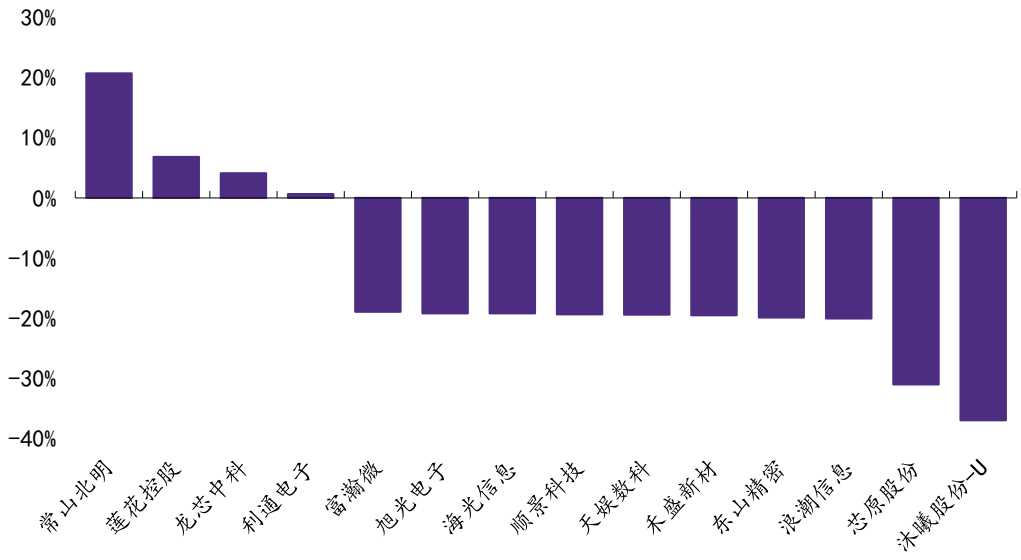
上周（2026.11.16-2026.11.20 日），AI 算力指数/中证红利日涨幅最大值分别为 3.73%/1.42%，AI 应用指数/AI 算力指数/万得全 A/中证红利日跌幅最大值分别为-4.76%/-5.81%/-2.78%/-0.46%。AI 算力指数内部，常山北明以 20.71%录得上周最大涨幅，沐曦股份-U 以-36.97%录得上周最大跌幅。AI 应用指数内部，真视通以 47.03%录得上周最大涨幅，电魂网络以-33.72%录得上周最大跌幅。

图表 15：上周（2026.11.16-2026.11.20 日）指数日涨跌幅



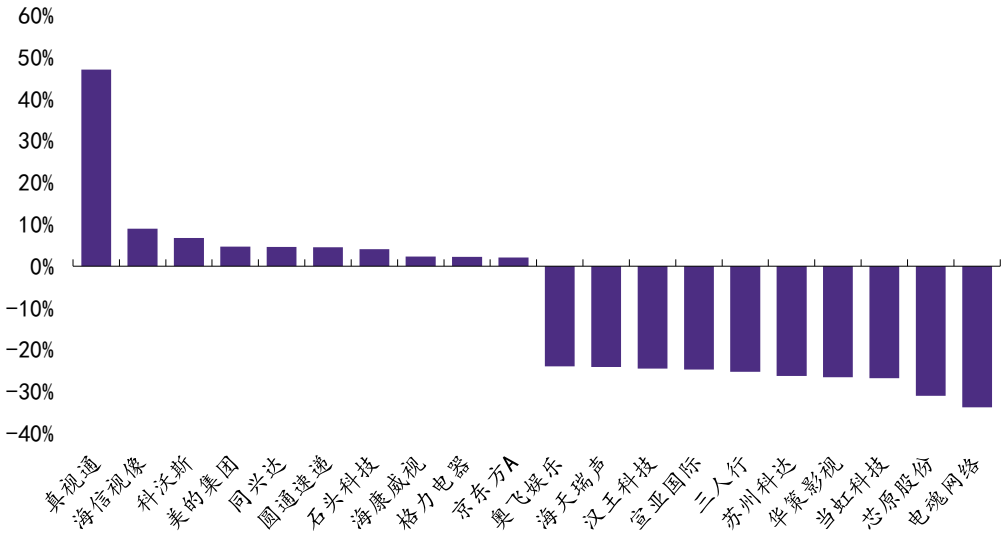
资料来源：wind，华鑫证券研究

图表 16：上周（2026.11.16-2026.11.20 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 17：上周（2026.11.16-2026.11.20 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

近日，英伟达发布2027财年第三季度财报。当季收入1,158.44亿美元，同比增长103%；数据中心收入1,076亿美元，同比增长110%；GAAP与非GAAP分别为74.8%和74.9%，同比均提升1.4个百分点。下一代Vera Rubin NVL72已进入量产爬坡与出货阶段，公司预计第四季度收入1,300亿美元，上下浮动3%。产品方面，英伟达于11月16日宣布旗下Spectrum-X Ethernet Photonics已全面量产，Lambda正采用NVIDIA Photonics CPO交换机连接GB300 NVL72机架，部署规模超过10,000张GPU并运行实际工作负载，为CPO在大型AI集群中的首个大规模部署。

本次公司财报及近期CPO进展表明，算力扩张与网络侧升级正在同步推进。数据中心收入延续高增长，AI基础设施的高需求持续兑现。CPO方面，产品量产与实际业务负载部署同步落地，表明其已逐步进入规模化部署阶段。随着AI集群规模扩大，网络功耗与可靠性对集群运行效率的影响愈发显著，光互联在AI基础设施中的重要性进一步提升。我们认为，Lambda的案例为CPO提供了实际运行验证，后续需重点跟踪部署客户扩展、实际运行表现及规模化交付能力，CPO及配套光互联环节的产业化进程值得持续关注。

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18：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

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 19：重点关注公司及盈利预测

公司代码	名称	2026-11-23	EPS			PE			投资评级
		股价	2025	2026E	2027E	2025	2026E	2027E	
300757.SZ	罗博特科	215.48	-0.40	2.13	4.49	-538.70	101.16	47.99	买入
301196.SZ	唯科科技	54.72	2.05	1.98	2.75	39.28	27.67	19.91	买入
301369.SZ	联动科技	76.05	0.48	1.08	2.62	158.44	70.42	29.03	买入
603859.SH	能科科技	32.87	0.92	1.05	1.32	35.73	31.30	24.90	买入
688615.SH	合合信息	85.45	3.24	4.22	5.25	26.37	20.25	16.28	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&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：

	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。