计算机

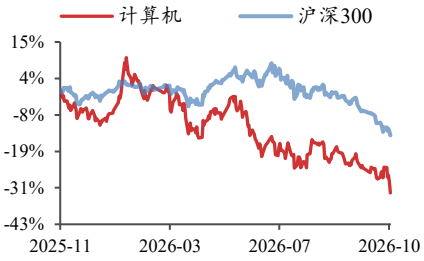
2026 年 11 月 01 日

# Agent 推理驱动服务器升级，AI 产业迈向全栈协同

——行业周报

**投资评级：看好（维持）**

**行业走势图**




数据来源：聚源

**相关研究报告**

《Anthropic 递交 S-1 招股书启动 IPO，模型与产品双轮驱动兑现盈利拐点—行业点评报告》-2026.10.28

《Kimi K3.1 Pro 智能指数全球第四，闭源差距再缩小—行业点评报告》-2026.10.25

《国产算力系列：单芯片性能与集群规模同步跃升，在手订单全面提速—行业周报》-2026.10.25

**张越（分析师）**
zhangyue1@kysec.cn
证书编号：S0790524090003

**叶进新（联系人）**
yejinxin@kysec.cn
证书编号：S0790126060053

**● Agent 推理拓展持续性算力需求，AI 服务器全栈升级打开成长空间**

AI 应用由单次生成走向多轮交互、长上下文与多智能体协同，推动算力需求从集中建设向持续消耗延伸，为云厂商资本投入与 AI 服务器需求增长提供支撑。随着新一代整柜方案导入，服务器市场有望呈现出货增长与单机价值提升共振。架构层面，GPU 与定制 ASIC 协同发展，分别发挥通用生态与专用优化优势；HBM 扩容提速与先进封装升级增强数据供给能力，超节点及液冷整机柜推动交付形态向系统集成演进。主控层面，工具调用、沙箱执行与任务编排增加 CPU 负载，服务器 CPU 由辅助配套向核心调度中枢升级，带来配置比例与产品需求变化。产业竞争正从单芯片峰值性能转向有效 Token 产出与端到端效率，计算、存储、封装、互连及整机协同升级，有望持续拓展 AI 服务器产业链价值空间。

**● 模型效率提升与多模态应用共振，国产适配及算力建设持续推进**

本周 AI 产业围绕推理效率、跨应用协作与多模态交互持续创新。模型端，MiniMax 开源 M3.1 Flash，通过稀疏注意力、量化与功耗调度优化推理效率；OpenAI 发布 GPT-6.1 Terra，提升电脑操作能力并降低调用价格；谷歌推出 Gemini 4 Carbon 及世界模型 Genie 4，拓展复杂任务与实时交互能力。应用端，腾讯发布 Buddy One，连接办公应用并支持跨平台资料检索，同传大模型推动实时语音服务落地；可灵开放 Kling 4.0 API，促进视频生成融入创作工具与企业工作流。国产生态方面，讯飞发布全国产算力训练的语音基座模型，面壁 MiniCPM-OCR 完成多芯片平台适配。海外厂商持续推进算力投入与自研芯片布局，有望与模型优化、应用渗透共同推动 AI 产业发展。

**● AI 应用相对表现占优，关注产业升级带来的结构性机会**

10 月 26 日至 10 月 30 日，计算机指数周度变动为-10.85%，同期沪深 300、上证指数、创业板指50 分创变动-2.92%、-2.48%、-10.08%和-13.73%。AI 细分方向中，AI 应用指数累计变动约-10.28%，AI 算力指数约-12.91%，应用方向取得约 2.63 个百分点的相对收益。个股层面，盛天网络上涨 11.02%，实丰文化、亿纬锂能分别上涨 8.86%和 5.57%，顺丰控股、美的集团、吉比特及格力电器均录得上涨，部分个股呈现结构性活力。

**● 受益标的**

**国产算力：**寒武纪、海光信息、禾盛新材、沐曦股份、浪潮信息、工业富联等；**海外算力：**东山精密、胜宏科技、联特科技、德邦科技等；**算力租赁：**宏景科技、利通电子、协创数据等；**大模型：**智谱、MiniMax、阿里巴巴、腾讯控股；**液冷：**盾安环境、同飞股份、恒铭达等。

**● 风险提示：**大模型技术迭代不及预期风险、AI 技术应用不及预期风险、相关政策不及预期风险、AI Infra 建设不及预期风险等。

# 目 录

1、 Agent 推理浪潮重塑算力底座，AI 服务器迎来价值重估 ..... 3
1.1、 推理需求刚性扩张，AI 服务器量价齐升延续上行周期 ..... 3
1.2、 AI 服务器架构四层演进，算力竞争升维为全栈协同竞赛 ..... 4
1.3、 核心部件筑基，整机龙头放量，算力消费扩张 ..... 11
2、 计算机产业动态更新 ..... 12
2.1、 国内 CSP 厂商与大模型厂商最新动向 ..... 12
2.2、 海外 CSP 厂商与大模型厂商最新动向 ..... 13
2.3、 国产算力产业最新进展 ..... 14
2.4、 海外算力产业最新进展 ..... 14
3、 上周市场表现回顾（2026/10/26-2026/10/30） ..... 15
4、 风险提示 ..... 17

# 图表目录

图 1： 云厂商资本开支扩张斜率放缓，但算力军备竞赛的投入强度不减 ..... 3
图 2： 2024-2027 年服务器产值稳步增长，AI 服务器占比持续抬升 ..... 4
图 3： AI 服务器出货芯片结构:GPU 占比缓降，ASIC 占比升至 2030 年近 40% ..... 4
图 4： 2026 年英伟达高端 GPU 出货结构:Blackwell 占比升至 71% ..... 5
图 5： Blackwell Ultra 结合双裸片互连与高带宽内存，增强计算与数据供给能力 ..... 7
图 6： 局部硅互连与再布线层协同，兼顾大规模集成与高密度连接 ..... 8
图 7： AICC 2026 浪潮信息发布元脑 SD200 Ultra 超节点，可跨主机域全局统一寻址 ..... 8
图 8： 代理型 AI 转为多步骤协同编排，编排、工具调度等逻辑成为 CPU 侧主要负载 ..... 9
图 9： Agent 时代 CPU 从辅助算力升级为核心调度中枢 ..... 9
图 10： 服务器 CPU 出货量将持续上涨，配比收敛引发服务器 CPU 供给格局重构 ..... 10
图 11： AI 服务器产业链由上游核心部件、中游整机制造与下游算力消费三层构成 ..... 11
图 12： 中游整机环节呈现头部集中格局，竞争焦点由标准化硬件转向系统级交付 ..... 11
图 13： 本周计算机板块跌幅较大，整体表现弱于多数一级行业 ..... 16
图 14： 本周 AI 细分指数均下跌，AI 应用指数整体表现优于 AI 算力指数 ..... 16
图 15： AI 应用指数整体承压，盛天网络等个股逆势上涨 ..... 17

表 1： 五大厂商 AI 芯片路线对比：GPU 通用生态与 ASIC 专用优化并行 ..... 6
表 2： 新一代 GPU 导入 HBM4，带宽显著提升、容量配置差异化扩展 ..... 7

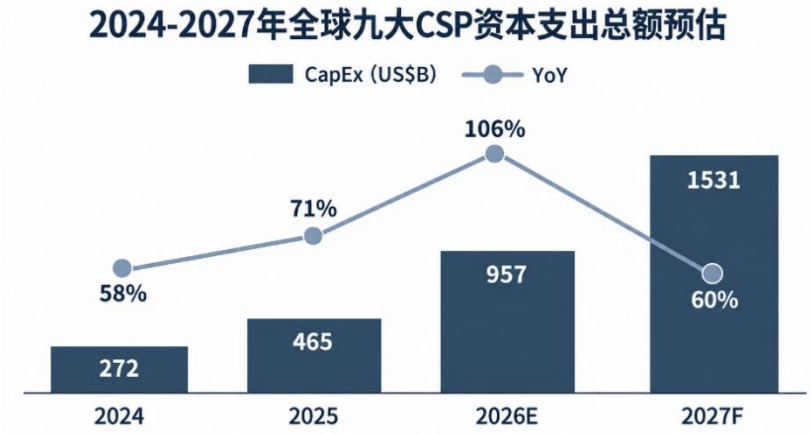
# 1、 Agent 推理浪潮重塑算力底座，AI 服务器迎来价值重估

## 1.1、 推理需求刚性扩张，AI 服务器量价齐升延续上行周期

**AI 产业重心自模型训练转向推理规模化落地，算力需求的性质发生根本改变，由项目制建设转变为持续性消耗。**2026 年被业界视为 Agent 规模化部署元年，中国信息通信研究院披露，截至 2026 年 6 月我国日均词元调用量已突破 500 万亿，较 2025 年 12 月增长近四倍；IDC 预计国内活跃 Agent 数量将从 2026 年的 495 万个增至 2030 年的 2 亿个。Agent 工作负载以多轮交互、长上下文与多智能体协同为特征，任务轮次与并发规模持续放大 Token 消耗，用量的增长具备自我强化属性。推理需求的扩张是刚性的：一旦嵌入业务流程便不可撤回，其持续放量构成下游资本开支上修的需求底座。

**云厂商资本开支扩张斜率放缓，但算力军备竞赛的投入强度不减，供应获取能力取代预算规模成为竞争核心。**TrendForce 集邦咨询 10 月 14 日发布研究，将 Google、Amazon、Meta、Microsoft、Oracle 等九大 CSP 资本支出预期由年增 50%上调至 60%：2026 年总额逾 9570 亿美元、同比增长 106%，2027 年有望突破 1.5 万亿美元、同比增长 60%，均优于先前预期，绝对规模逐年抬升。上调由三项因素驱动：服务器整机价格上涨、零部件缺货推高报价、部分供应商要求预付款以确保供货，CSP 被迫提早支出现金。资本开支节奏已由建设周期主导转向受供给瓶颈与商务条款共同塑造，军备竞赛的胜负手正从预算规模转向供应获取能力。**资金充沛带动强劲拉货，预计 2027 年 AI 服务器出货成长将继续扩大。**

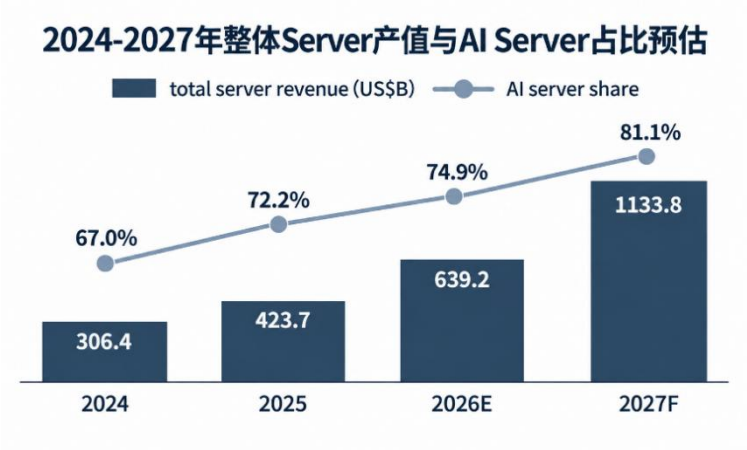
**图1：云厂商资本开支扩张斜率放缓，但算力军备竞赛的投入强度不减**




数据来源：TrendForce 集邦咨询公众号、开源证券研究所

**AI 服务器市场扩张由台数驱动转向价值驱动，产值增速显著跑超出货增速，量价齐升成为本轮景气周期的核心特征。**出货端，TrendForce 集邦咨询预计 2026 年 AI 服务器出货同比增长 32.2%，2027 年在超大型 CSP、Neocloud 与主权 AI 项目共同拉动下再增近 36%，增速平稳抬升。产值端，2026 年 AI 服务器产值约 4790 亿美元，在整体服务器约 6392 亿美元产值中占比接近 75%；2027 年产值预计跃升至近 9200 亿美元，同比增长超过 90%，占整体服务器产值比重升至 81%左右，推动整体服务器产值首次跨越 1 万亿美元。价值量攀升源于新一代整柜方案均价上行与存储器为首的零部件涨价，GB300、VR200 及 Helios 等整柜式方案或成为 2027 年订单主力。

图2：2024-2027 年服务器产值稳步增长，AI 服务器占比持续抬升




数据来源：TrendForce 集邦咨询公众号、开源证券研究所

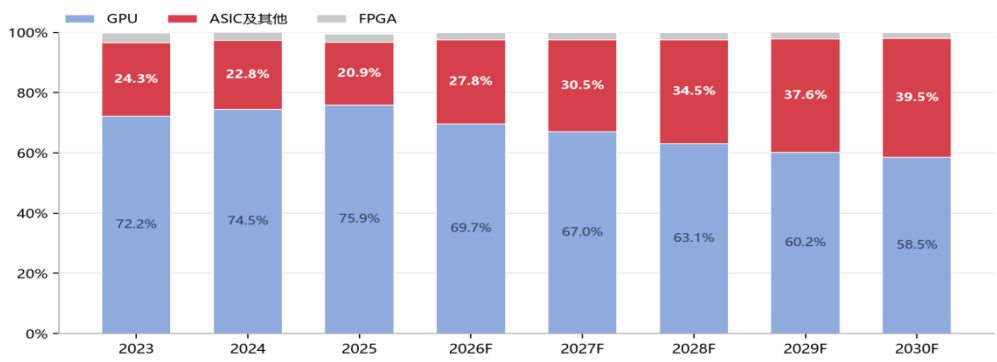
## 1.2、 AI 服务器架构四层演进，算力竞争升维为全栈协同竞赛

**AI 服务器是一项跨越芯片、存储、互连与整机柜的系统工程，其价值分布已远超单颗计算芯片，需要分层审视。**计算芯片层由 GPU 与定制 ASIC 构成算力供给主体，前者主导通用计算与复杂训练，后者围绕特定工作负载优化能效与成本；存储与封装层以 HBM 高带宽内存和 CoWoS 先进封装为核心，决定数据供给能力与算力释放上限；系统层的服务器 CPU 承担控制、调度与数据搬运，在 Agent 工作负载中角色重新上升；互连与整机层则以超节点和液冷整机柜为载体，决定集群规模下的通信效率与稳定运行。AI 服务器的竞争已从单一芯片的算力比拼，演化为制程、存储、封装、互连与整机系统的全栈协同竞赛。

### （一）计算层：从 GPU 独大到 GPU+ASIC 并行

**全球 AI 算力市场正从通用 GPU 独大走向 GPU 与定制 ASIC 并行发展的多元化时代，ASIC 出货占比进入陡峭上行通道。**据 TrendForce 集邦咨询研究，2026 年全球 AI 服务器出货量预计同比增长 28%以上，从所用芯片看，GPU 占 69.7%仍为最大宗；而在 Google、Meta 等北美业者积极扩张自研 ASIC 方案的情况下，预计 2026 年 ASIC AI 服务器出货占比将提升至 27.8%，为 2023 年以来最高，出货增速也超越 GPU AI 服务器。该机构进一步预计，随着 Google、Amazon 等 CSP 大力自研 ASIC 芯片的态势扩大，ASIC AI 服务器出货占比将从 2026 年的 27.8%攀升至 2030 年的近 40%。芯片层的供给结构正由单一 GPU 采购转向 GPU 与 ASIC 的组合配置。

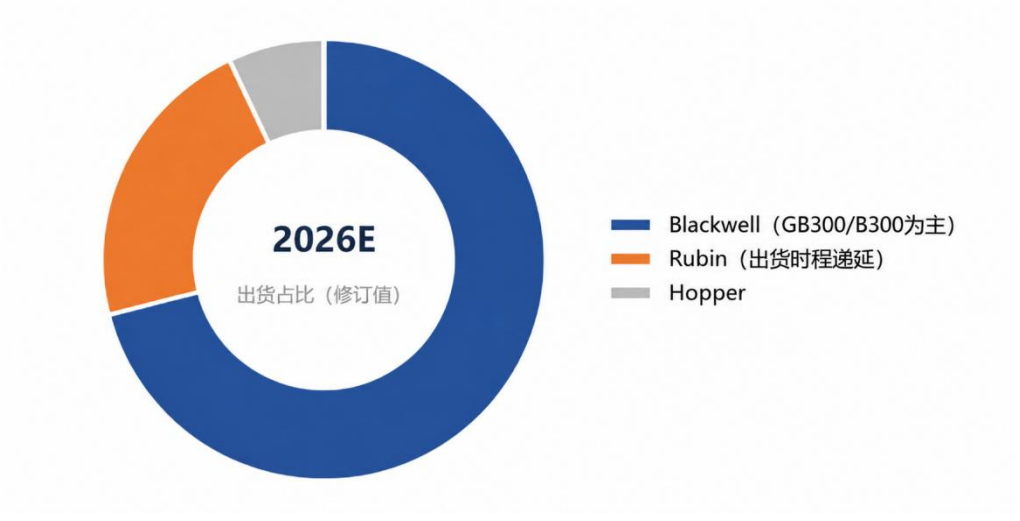
图3：AI 服务器出货芯片结构:GPU 占比缓降，ASIC 占比升至 2030 年近 40%




数据来源：TrendForce 集邦咨询、开源证券研究所

**GPU 路线仍是 AI 算力的基本盘，英伟达凭借整柜方案与多元产品配置巩固主导地位。**TrendForce 集邦咨询预计，2026 年搭载英伟达 GB300 的机种将成为出货主流，VR200 于下半年后逐步放量；受 Rubin 系列出货进度递延等因素影响，Blackwell 系列占其高端 GPU 出货比例将从 61%大幅成长至 71%，主导市场的地位更加巩固。面对 ASIC 竞争，英伟达在 GTC 2026 转向聚焦 AI 推理落地，以 GPU、CPU、LPU 分攻训练与推理需求，并推动 GB300、VR200 等 CPU+GPU 整柜方案，发表高度垂直整合的 Vera Rubin 系统。GPU 的比较优势在于通用性与生态，CUDA 软件栈与可编程架构可灵活适配快速演进的模型结构，在前沿模型预训练与多租户公有云场景中保持绝对主导。

**图4：2026 年英伟达高端 GPU 出货结构:Blackwell 占比升至 71%**




数据来源：TrendForce 集邦咨询公众号、开源证券研究所

**自研 ASIC 阵营中，谷歌与亚马逊率先完成规模化落地，但技术路线各有侧重。**谷歌是 ASIC 市场领头羊，第七代 TPU Ironwood 专注 AI 推理，峰值算力达 4614 TFLOPS、为上代 Trillium 的 5 倍，内存 192GB 为其 6 倍，每瓦峰值算力翻倍至 29.3 TFLOPS，提供 256 与 9216 芯片两种配置，后者的 AI 算力达 42.5 ExaFLOPS；TPU 除自用于 Google Cloud 外，也对外销售给 Anthropic 等业者。亚马逊的 Trainium3 采用 3nm 工艺，Trn3 UltraServer 集成 144 颗芯片、总算力 362 PFLOPS，计算性能较上代提升 4.4 倍、能效提升 40%；其单卡性能仅为英伟达 GB200 的 56%，以系统堆料弥补单卡差距，Trainium 系列芯片部署量已超 100 万枚。

**微软与 Meta 的自研 ASIC 呈现鲜明的自用捕获特征，均以推理为优先场景。**微软 Maia 200 专为推理设计，采用台积电 3nm 工艺，配备 216GB HBM3e 与 7TB/s 带宽，FP4 算力超 10 PFLOPS，每美元 token 产出较现役机群高出约 30%，已在美国数据中心部署并支撑 Copilot 等业务；微软规划自 2027 年起让 Maia 系列成为其主要推理系统。Meta 两年内密集推出 MTIA 300、400、450、500 四代芯片，HBM 带宽提升 4.5 倍、浮点算力提升 25 倍，已在生产环境部署数十万颗，主要服务推荐排序等内部高频推理负载，MTIA 450、500 计划于 2027 年落地。自研 ASIC 与 GPU 形成分工而非全面替代，训练看 GPU、推理看 ASIC 的异构并存格局正在成形。

**表1：五大厂商 AI 芯片路线对比：GPU 通用生态与 ASIC 专用优化并行**

| 厂商 | 芯片/平台 | 战略定位 | 制程与内存 | 关键性能/出货 | 商业化与部署 |
|---|---|---|---|---|---|
| 英伟达 | Blackwell（GB300/B300）+VR200 | 通用 GPU：训练主导、训推兼顾；GPU+CPU+LPU 多元布局 | 台积电定制 4nm；GB300 为 2026 年出货主流、VR200 下半放量 | 2026 年 Blackwell 占高端 GPU 出货 71%（Rubin 递延至 22%） | 公有云与 Neocloud 全网覆盖；GB300 等整柜方案主导 |
| 谷歌 | TPU v7（Ironwood） | ASIC 领头羊：训推兼顾、推理强化 | 3nm；192GB HBM（上代 6 倍） | 峰值算力 4614 TFLOPS（上代 5 倍）；每瓦性能翻倍 | GCP 自用+外销 Anthropic 等；9216 芯片 Pod 达 42.5 ExaFLOPS |
| 亚马逊 | Trainium3 | 训推一体：单卡性价比+系统级堆料 | 3nm；144GB HBM3e | 单芯片 2.52 PFLOPS（FP8）；Trn3 UltraServer 144 颗、362 PFLOPS | EC2 实例租用；已部署超 100 万枚；绑定 Anthropic（Project Rainier） |
| 微软 | Maia 200 | 推理专用：自用捕获（Captive） | 台积电 3nm；216GB HBM3e | FP4 算力超 10 PFLOPS；每美元 token 产出+30% | 支撑 Copilot 等内部负载；2027 年起成为主要推理系统 |
| Meta | MTIA 300/400/450/500 | 推理优先：服务内部推荐与排序负载 | HBM 带宽较 300 代提升 4.5 倍 | 两年四代、算力提升 25 倍 | 生产环境部署数十万颗；不外销；450/500 于 2027 年落地 |

资料来源：TrendForce 集邦咨询公众号、超能网公众号、亚马逊云开发者公众号、《财经》杂志公众号、Azure 云科技、Meta 官方博客、开源证券研究所

### （二）内存与封装层：算力竞赛延伸至数据供给竞赛

**长上下文与复杂任务增加数据存储和搬运需求。**大模型运行既需足够容量容纳权重及中间数据，也需充足带宽支持持续读写。随着应用复杂度提高，容量影响模型部署与任务承载规模，带宽制约数据供给与算力利用效率，共同成为 AI 服务器配置升级的重要方向。

**训练与推理对内存的需求各有侧重，共同推动显存扩容与带宽提升。**训练除保存权重外，还需存储梯度、优化器状态及中间激活，模型规模扩大、训练序列延长与微批次增加均会提高显存占用，数据访问密集环节也需要充足带宽以提高算力利用效率。常规推理无需保存梯度和优化器状态，显存主要容纳模型权重与 KV Cache，在模型结构、缓存精度及策略不变的情况下，KV Cache 占用随上下文长度与并发序列数增加。与可并行处理输入的 Prefill 阶段相比，Decode 阶段需逐步生成输出，反复读取权重与历史缓存，在小批次场景下更易受到带宽限制；扩大批次虽能提高权重复用效率，但又受到缓存容量与响应时延约束，推动内存容量与带宽协同升级。

**容量与带宽需求提升，推动 HBM 沿存储密度提高与数据接口升级两条路径演进。**在容量方面，HBM 通过垂直堆叠 DRAM 裸片实现扩容，裸片减薄与层间间距压缩为增加堆叠层数提供支撑；制程微缩与存储阵列优化则提高单颗 DRAM 容量，使相同堆叠层数能够容纳更多数据。在带宽方面，接口位宽扩大与引脚速率提高共同增强传输能力。HBM4 将接口位宽由上一代的 1024bit 扩大至 2048bit，在相同引脚速率下理论带宽翻倍，叠加速率提升，进一步增强训练与推理过程中的数据供给能力。

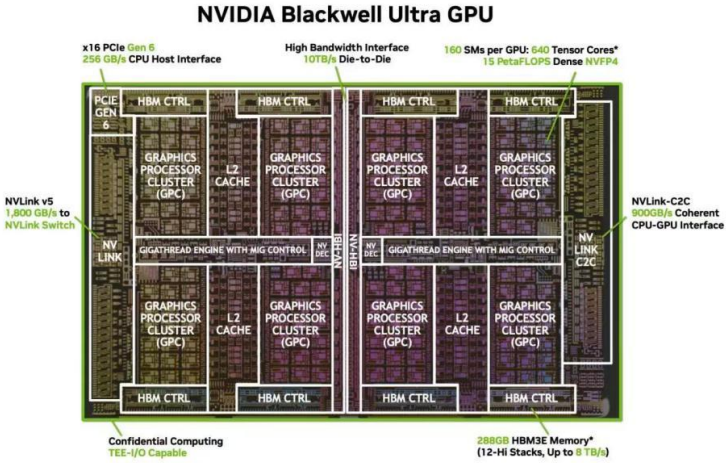
**单堆栈规格与堆栈数量共同推动加速器内存配置升级。**单堆栈容量提升主要支持扩容，接口带宽提高主要增强数据供给；在配套接口支持下，增加堆栈数量则可同时提升总容量与总带宽。其中，Rubin 在维持与 Blackwell Ultra 相同显存容量的情况下提高带宽，MI455X 则较 MI355X 同步扩充容量与带宽，体现出不同产品的配置取舍。定制 AI 加速器同样提高内存配置，Trainium3 较 Trainium2 的内存容量与带宽分别增长 50%及约 69%，扩容与提速成为不同计算路线的共同方向。

**表2：新一代 GPU 导入 HBM4，带宽显著提升、容量配置差异化扩展**

| 厂商 | 产品及配置口径 | 内存类型 | 单 GPU 显存容量 | 单 GPU 峰值内存带宽 |
|---|---|---|---|---|
| NVIDIA | Blackwell Ultra，官方最高配置 | HBM3E | 288GB | 8TB/s |
| NVIDIA | Vera Rubin NVL72 中的 Rubin GPU | HBM4 | 288GB | 19.2TB/s |
| AMD | MI355X | HBM3E | 288GB | 8TB/s |
| AMD | MI455X | HBM4 | 432GB | 23.3TB/s |

数据来源：英伟达、AMD、开源证券研究所

**图5：Blackwell Ultra 结合双裸片互连与高带宽内存，增强计算与数据供给能力**



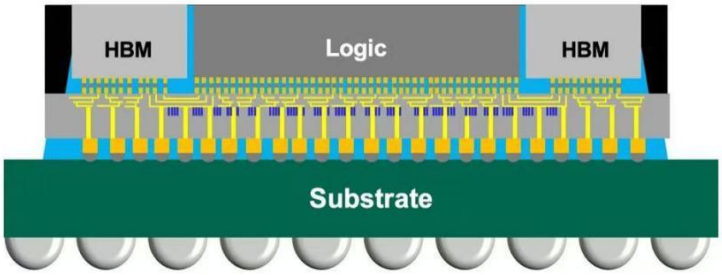

资料来源：英伟达

**HBM 堆栈增加与多芯粒架构发展，推动封装集成规模扩大。**2.5D 封装通过中介层连接并排布置的逻辑裸片与 HBM，实现计算与存储集成。随着单封装部件数量增加，芯片占地及配套信号、供电线路需求同步增长，推动中介层布局与布线空间扩展。台积电于 2026 年 4 月披露，5.5 倍光罩尺寸 CoWoS 已进入生产阶段，并计划于 2028 年投产 14 倍光罩尺寸方案，支持集成约 10 颗大型计算裸片与 20 个 HBM 堆栈，为更大规模的计算与存储配置提供空间。

**宽接口与芯粒间通信增长，提高局部互连密度要求。**HBM 接口拓宽及多芯粒协同计算需要更多并行数据通路，单纯扩大中介层面积难以满足芯片接口区域的密集连接需求，还需通过缩小线宽线距、优化布线层次及压缩连接点间距，提高有限空间内的互连能力。以 CoWoS-L 为例，其将局部硅互连（LSI）与再布线层（RDL）相结合，由 LSI 的多层亚微米铜布线承担关键区域的高密度连接，RDL 完成较大范围的信号连接与供电，兼顾集成规模扩展与局部互连能力提升。对于通信联系紧密的裸片，三维集成还可通过垂直堆叠与混合键合提高层间连接密度、缩短传输路径，并与 2.5D 横向集成配合，满足更高的芯粒间带宽需求。随着加速器计算与存储配置

提升，先进封装需求不仅体现为部署数量增长，也体现为单封装集成面积、布线密度及键合要求提高。

**图6：局部硅互连与再布线层协同，兼顾大规模集成与高密度连接**




资料来源：台积电

### （三）互连层：超节点驱动机柜级交付成为主流

**前沿模型的规模扩张突破单机承载极限，机柜取代单台服务器成为 AI 算力的基本交付单元。**模型参数从 DeepSeek R1 的 6710 亿走向 Kimi K3 的 2.8 万亿并迈向 10 万亿量级，上下文由 128K 扩展至 1M 以上，单颗芯片装不下，切分到多台服务器又因跨机通信拖累效率，多芯如一机的超节点成为承接前沿模型的必选架构。超节点与节点集群的分界线是显存统一寻址，做不到统一寻址、把模型切成数段分散到不同节点的方案，本质上仍是节点集群。前沿模型的运行要求印证了这一判断，2.8 万亿参数的 Kimi K3 模型官方建议运行于 64 卡以上统一内存超节点。一年前超节点还是少数厂商的旗舰产品，如今已成主流算力企业竞逐的必争之地。统一内存、直连互连与算子融合构成超节点的三重工程内核，Token 生成速度成为衡量其实际价值的标尺。**超节点并非越大越好，通信成本与故障效应随规模放大，架构与模型的匹配度才是价值兑现的关键。**

**图7：AICC 2026 浪潮信息发布元脑 SD200 Ultra 超节点，可跨主机域全局统一寻址**




资料来源：数智前线公众号

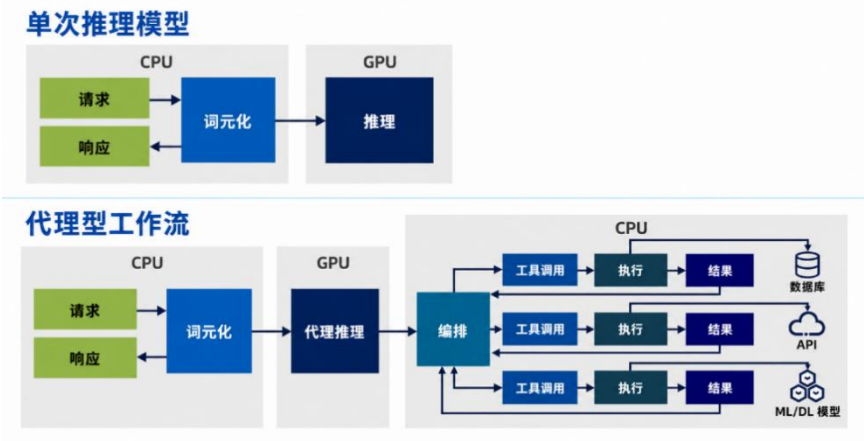
**国内外算力企业在超节点上同步卡位，机柜级产品从英伟达一家独供演变为多阵营并行的竞争格局。**英伟达 GB300 NVL72 以 72 颗 GPU 与 36 颗 Grace CPU 构成单一 NVLink 统一内存域，单机柜集成液冷交付，下一代 Vera Rubin NVL72 已全面投产、计划 2026 年下半年发货。国内阵营，华为 CloudMatrix 384 超节点以自研 UB

网络实现 384 颗昇腾 NPU 与 192 颗鲲鹏 CPU 全对等互联，一套系统提供 300 PFLOPS BF16 算力，截至 2025 年 9 月累计部署超 300 套、服务超 20 家客户，华为还在 WAIC 2026 上展出 Atlas 950 SuperPoD 超节点，实现业界最大 1024 卡规模互联，预计 2026 年第四季度上市；浪潮信息发布 128 芯紧耦合的 SD200 Ultra；阿里在云栖大会推出基于自研芯片的超节点整机。机柜级交付同步重构产业配套，供电、液冷与整机部署均以机柜为单元重新设计。

### （四）主控层：Agent 时代服务器 CPU 配比重构

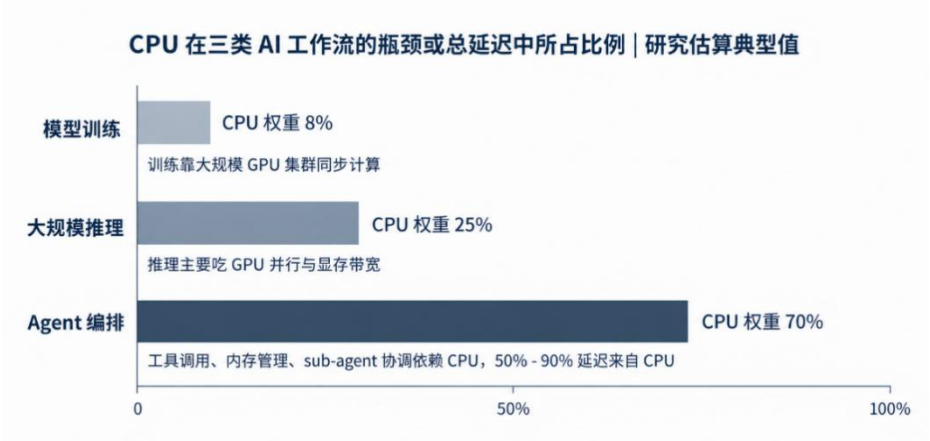
**Agent 将 AI 推理从单次生成改造为多步循环，服务器中的数据搬运的配角转变为决定系统吞吐的编排中枢。**传统推理是单次前向传播，GPU 承担矩阵运算，CPU 仅负责请求分发与数据搬运，单颗 CPU 足以配套多颗 GPU；Agent 工作负载则由模型推理、工具调用、结果校验、重新规划的循环链构成，重复数次至数十次，循环中的工具路由、沙箱执行、数据库检索与会话状态维护均为串行逻辑任务，全部落在 CPU 一侧。学术研究显示，在检索增强、代码执行、Web 操作等典型智能体工作流中，CPU 侧工具处理占端到端延迟的比例最高可达 90%以上；编排能力不足时 GPU 在等待间隙闲置，系统瓶颈由算力转为调度。

**图8：代理型 AI 转为多步骤协同编排，编排、工具调度等逻辑成为 CPU 侧主要负载**




资料来源：英特尔《代理型 AI 需要更多 CPU》、开源证券研究所

**图9：Agent 时代 CPU 从辅助算力升级为核心调度中枢**

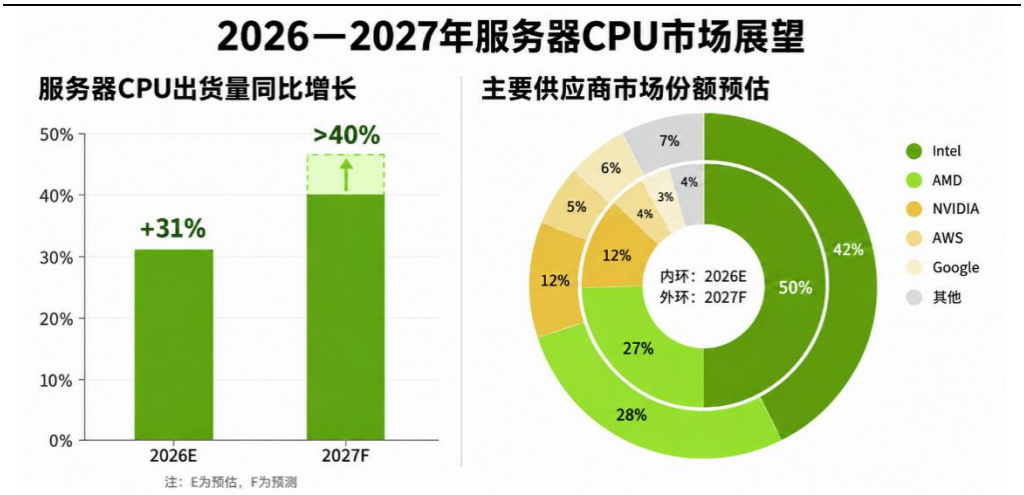



数据来源：律动 BlockBeats、开源证券研究所

**编排瓶颈直接改写服务器内部的芯片配比，CPU 与 GPU 的配置关系向对等方向收敛，服务器 CPU 出货或将随之进入提速通道。**英特尔 2026 年 3 月发布的白皮书实测显示，同时针对代理型 AI 与代码生成的使用案例，使用搭配 NVIDIA B200 GPU 的 Intel Xeon 6767P CPU，结果发现将 CPU 与 GPU 比例维持在 1:1 至 1.4:1，才能使 GPU 保持满载运行；TrendForce 集邦咨询 10 月 22 日研究显示，依据头部服务商规划的机架配置，CPU 与 GPU、ASIC 的配比将从传统的 1:8 逐步转向 1:2 甚至 1:1，2026 年服务器 CPU 出货量同比增长 31%，2027 年增幅有望超过 40%。负载结构的迁移已有产业印证，Arm 在 HotChips 2026 大会披露，企业自动化工作负载中 CPU 与加速器的任务分配比例约为 50 比 50。

**配比收敛引发服务器 CPU 供给格局重构，x86 双寡头稳态松动，ARM 阵营与英伟达的入局共同指向 2027 年约 30%的 ARM 份额预期。**2025 年 x86 架构占据服务器 CPU 市场近 90%份额，ARM 架构刚突破 10%。供给侧动态分化：英特尔新品受工艺与封装难题拖累、自有产能不足，出货延后、份额下滑；AMD 新一代 Venice 处理器进入小规模量产，但出货主力仍是 Genoa 和 Turin 平台，2026 年份额预计提升至约 27%。ARM 阵营扩产更为激进，AWS 持续部署 Graviton，谷歌 Axion 大批量出货，微软 Cobalt 100 与 Arm 自研 AGI 处理器量产交付，2026 年 ARM 份额有望突破 20%；英伟达 2026 年 3 月宣布单独销售 Vera CPU，以满足客户对 CPU 的配比需求，阿里巴巴、字节跳动等均为早期合作伙伴。TrendForce 预计 2027 年 ARM 份额约 30%，主力来自 Vera CPU 爬坡与谷歌、AWS、华为等新款处理器放量。

**图10：服务器 CPU 出货量将持续上涨，配比收敛引发服务器 CPU 供给格局重构**



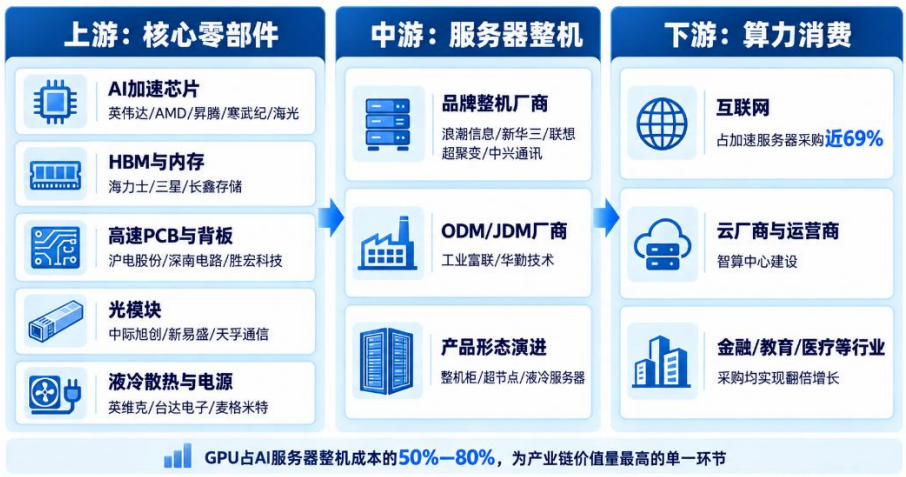

数据来源：TrendForce、开源证券研究所

**AI 服务器的评价标尺正从峰值算力转向 Token 产出效率，算力军备竞赛的决胜变量由堆叠规模转为单位投入的智能产出。**采购端变化最为直接，企业客户不再简单核对芯片型号与集群规模，首字时延与产出速度成为新标尺，有 AI 编程团队测算，首 Token 时延从 10 秒压缩至 2 秒，同等算力下一分钟可响应的并发请求提升五倍。系统端标尺同步刷新，前沿模型的 Token 生成速度成为衡量超节点价值的基准。与需求扩张并存的是效率矛盾，中国信息通信研究院数据显示，2025 年国内智算供应约 38 亿卡时、实际使用 14 亿卡时，利用率仅 36%，制约产业的不只是卡的数量，更是单位投入的 Token 产出。军备竞赛的终点不是无限堆卡，而是低成本、可持续的智能供给。

## 1.3、 核心部件筑基，整机龙头放量，算力消费扩张

**AI 服务器产业链由上游核心部件、中游整机制造与下游算力消费三层构成，价值重心高度集中于芯片环节。**前瞻产业研究院指出，产业链上游主要为 AI 芯片、光模块、内存、PCB 等原材料及设备制造，中游为 AI 服务器整机制造，下游应用覆盖大模型、互联网、工业制造、金融与公共服务等领域。其中 GPU 占 AI 服务器整机成本的 50%-80%，是产业链价值量最高的单一环节；高带宽内存、高速 PCB 与背板、800G 光模块、液冷散热与电源等配套部件，随算力密度提升同步升级。IDC 数据显示，2025 年中国 AI 服务器市场规模达 259 亿美元、同比增长 36.32%，其中 AI 推理服务器占比达 67%，上游部件需求随整机放量持续扩张。

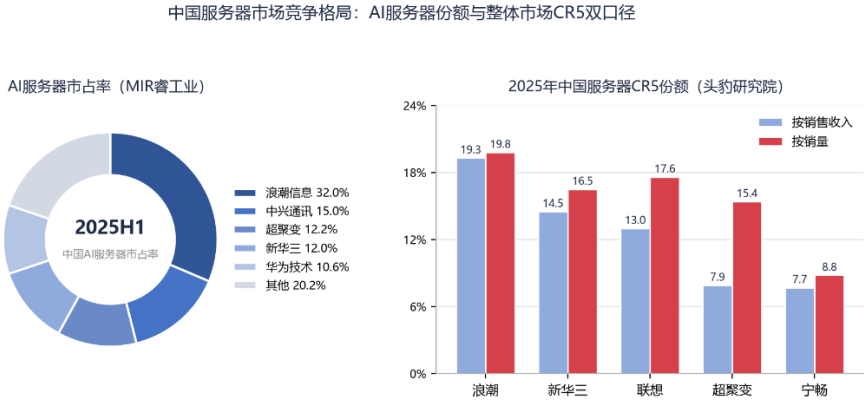
**图11：AI 服务器产业链由上游核心部件、中游整机制造与下游算力消费三层构成**




资料来源：前瞻产业研究院公众号、深圳市产业转移促进中心、开源证券研究所

**中游整机环节呈现头部集中格局，竞争焦点由标准化硬件转向系统级交付。**据 MIR 睿工业统计，2025 年上半年中国 AI 服务器市场中，浪潮信息以约 32.0%的市占率位居第一，中兴通讯、超聚变、新华三、华为技术分别约 15.0%、12.2%、12.0%和 10.6%，位列第二至第五。头豹研究院数据显示，2025 年中国服务器市场按销售收入计 CR5 达 62.4%，浪潮、新华三、联想、超聚变、宁畅分列前五；按销量口径 CR5 更高达 78.1%，大规模交付能力正向头部厂商集中。随着 AI 服务器向整机柜、液冷与超节点形态演进，整机竞争已从单台硬件比拼延伸至高速互联、供电散热与系统集成的系统能力竞争。

**图12：中游整机环节呈现头部集中格局，竞争焦点由标准化硬件转向系统级交付**




数据来源：前瞻产业研究院公众号、头豹研究院、开源证券研究所

**品牌整机阵营中，浪潮信息与超聚变构成国产双雄，分别以规模优势与国产化生态见长。**浪潮信息 2025 年实现营业收入 1647.82 亿元，稳居中国 AI 服务器市场份额首位，旗下元脑品牌覆盖 AI 服务器、超节点与存储网络的完整产品矩阵。超聚变脱胎于华为 x86 服务器业务，2025 年以 15.4%的销量份额位居中国服务器市场前列，依托鲲鹏、昇腾生态在国产化服务器领域占据领先位置；招股书显示 2026 年 1-6 月其 AI 服务器收入占比已提升至 53.29%。两家厂商均将整机柜与液冷方案作为竞争抓手，在互联网、运营商与政企市场正面交锋。

**阵营加速向算力硬件迁移，华勤技术成为消费电子厂商切入 AI 服务器的代表。**华勤技术 2026 年上半年实现营业收入 937.2 亿元，同比增长 11.7%，归母净利润 30.00 亿元、同比增长 58.80%，其中计算及数据业务收入达 436.7 亿元；公司已构建覆盖 AI 服务器、通用服务器、交换机、超节点整机柜的全栈式产品体系，是行业内极少数同时拥有计算节点和网络节点设计能力的厂商，在超节点领域具备先发优势，为国内超节点产品核心供应商。2025 年公司 AI 服务器收入已突破 400 亿元，交换机业务营收超 25 亿元并实现倍数级增长，算力硬件正接棒移动终端成为第二增长曲线。

**下游算力消费多维扩张，互联网之外行业需求接力放量。**深联产业研究院数据显示，互联网行业仍占中国加速服务器采购的近 69%，但金融、教育、运营商、医疗健康等行业的采购均实现超过一倍增长，市场正从互联网独大走向多行业开花，国内阿里、字节、百度、腾讯等企业持续跟进算力投资。政策端，2026 年 5 月国务院常务会议将算力网纳入新基建六网，算力基础设施上升至与水网、电网同等的国家战略高度，投入的长期性与稳定性显著增强。下游需求的多元化与国家级投入，共同为上游部件与中游整机环节提供确定性支撑，产业链景气度有望沿部件、整机到算力运营的全链条传导。

## 2、 计算机产业动态更新

### 2.1、 国内 CSP 厂商与大模型厂商最新动向

**（一）稀宇科技开源 M3.1 Flash，72 小时处理 Token 超 100 万亿**

2026 年 10 月 26 日，稀宇科技（MiniMax）开源 M3.1 Flash 模型权重，总参数 428B、激活参数 23B。新版在注意力机制、专家量化与功耗调度方面进行了调整。

M3.1 Flash 将前三个全注意力层改为稀疏注意力，并把路由 MoE 专家迁移至 W4A4 量化，同时通过 Power-aware scheduling 分配功耗，以提升推理效率。

上线 72 小时内，M3.1 Flash 的真实用户处理 Token 超过 100 万亿。华为昇腾、英伟达（NVIDIA）、超威半导体（AMD）等芯片厂商及 Together AI 等推理平台在发布首日完成接入。

**（二）腾讯控股开源同传大模型，响应时延降至 3 秒**

2026 年 10 月 26 日，腾讯控股发布并开源同声传译大模型 Hy-SimulMT，支持中英语音到语音实时翻译与音色克隆，并在腾讯 Hy 翻译 APP 首发免费体验。与传统级联式同传方案相比，端到端语音到语音架构减少了转写、翻译、合成之间的时延损耗。

实测结果显示，Hy-SimulMT 响应时延约 3 秒，人类专业译员平均为 4 至 5 秒；在 RealSI 长音频评测中，该模型的延迟最低，效果优于其他开源模型及多数商用 API。十个领域的实测中，其翻译质量得分高于作为对照的 6 位持 CATTI 一、二级证书的同传专家，时延也更低。

腾讯 Hy 翻译计划服务 2026 年 11 月深圳 APEC 会议，提供实时同声传译。模型权重已开源，开发者可在此基础上开展部署与应用开发。

**（三）腾讯控股发布 Buddy One，支持跨应用检索办公资料**

2026 年 10 月 30 日，腾讯控股发布面向个人的跨应用 AI 协同体系 Buddy One，连接腾讯文档、会议、网盘与 ima，共享 WorkBuddy Agent 底座。用户可在 WorkBuddy PC 端 5.9.0 以上版本关联账号后使用，无需反复授权，可通过自然语言检索多个应用中的资料并查看来源。

积分体系同步升级为 Buddy One 积分，在 WorkBuddy、文档、ima、网盘通用，会议暂不消耗积分；近期关联账号赠送 100 积分，原有积分不清零。

**（四）快手科技开放 Kling 4.0 API，面向企业客户与开发者**

2026 年 10 月 28 日，快手科技旗下可灵 Kling 4.0 API 面向全球企业客户与开发者开放，距 2026 年 10 月 15 日模型上线不足两周。Kling 4.0 支持最长 30 秒原生视频生成与多关键帧控制，并优化复杂动作、物理交互、运镜及人物微表情。

腾讯设计工具 Miora 以及 OpenArt、TapNow、Flova、Topview 等平台已接入 Kling 4.0。快手科技称，模型在风格还原、分镜运镜与角色一致性三个方面有所提升。开发者可通过 API 在创作工具、Agent 及企业工作流中调用视频生成功能。

## 2.2、 海外 CSP 厂商与大模型厂商最新动向

**（一）OpenAI 发布 GPT-6.1 Terra，成本降至 Astra 十分之一**

2026 年 10 月 28 日，OpenAI 发布 GPT-6.1 Terra，API 定价为每百万 Token 输入 1 美元、输出 5 美元，较 GPT-5.6 Terra 低 50%。在电脑操作基准 Agents' Last Exam 上，其得分比前代高 15 个百分点，与 GPT-6 Astra 相当，但 Token 价格仅为 Astra 的十分之一。

在易错提示词的低推理设置下，GPT-6.1 Terra 将错误率由 53.6%降至 16.7%；Terminal-Bench 4.0 解决率达 50.3%，为前代的两倍。其网络安全能力被评定为关键级。

**（二）谷歌发布 Gemini 4 Carbon，AA 智能指数 5.0 得分 59 分**

2026 年 10 月 30 日，谷歌（Google）发布 Gemini 4 Carbon，在 AA 智能指数 5.0 中以 59 分排名第一。首发输入价格为每百万 Token 2 美元，为 Opus 5.5 的一半。

Gemini 4 Carbon 在 Terminal-Bench 4.0 取得 65.2%、GraphWalks 取得 96.2%，均超过 Opus 5.5；完成评估所需的输出 Token 少 29%。模型在 SEC-bench Pro 和 ExploitGym 中分别取得 88.2%和 54.8%，均为该项评测的最高成绩；作弊与提示词注入攻击成功率均为 0.0%。

Gemini 4 Carbon 先向部分网络安全防御者开放，优先用于防御性场景。

**（三）谷歌发布世界模型 Genie 4，原生支持实时音频**

2026 年 10 月 29 日，谷歌 DeepMind 发布世界模型 Genie 4。谷歌称，其真实感与交互性在已发布的世界模型中处于领先位置。用户可通过键盘控制移动，也可使用自然语言改变天气、添加物体与角色。

Genie 4 原生支持实时音频，可同步生成环境音、音乐与对话。Gemini 3.8 Flash 负责引导规划，Gemini Omni 提供视觉行为知识。

Genie 4 已向受信任测试者开放研究预览，未来将通过 Project Genie 向 AI Ultra 用户及 Gemini API 伙伴开放，生成内容带有 SynthID 水印。

## 2.3、 国产算力产业最新进展

**（一）科大讯飞发布 SparkAudio 1.0，全国产算力训练**

2026 年 10 月，科大讯飞在 1024 开发者节发布语音基座大模型 SparkAudio 1.0。该模型使用全国产算力完成训练，支持转写、翻译、方言与环境音识别及说话人分离。

SparkAudio 1.0 采用 SemiEnc、S-S2S、CoDec 三项技术，分别用于增强表征、降低时延和减少算力开销。其上已开发 Spark-ASR-2.0 等三大专用模型，同传 MOS 由 4.4 升至 4.6。模型已应用于讯飞输入法、AI 录音环等产品；35 克录音环的 2 小时录音约 10 秒即可传至手机。

**（二）面壁智能开源 MiniCPM-OCR，完成 12 个芯片平台适配**

2026 年 10 月 29 日，面壁智能发布开源 MiniCPM-OCR 1.0（1.3B）。模型在 Real5-OmniDocBench 取得 93.68 分、OCRBench 54 取得，均为对应评测的最高成绩，吞吐达 1.217 页/秒。

面壁智能同步发布用于文档识别结果质检的判别模型 Jev（2B）。其在 JevOCRBench 上的整体 AUROC 达 93.6，法律文书免检比例可达 85%；DSpark 使端到端效率提升 55%。

MiniCPM-OCR 权重已开源，适配 vLLM、SGLang、Ollama 等推理框架，并在发布首日完成 12 个芯片平台适配。

## 2.4、 海外算力产业最新进展

**（一）Anthropic 提交 IPO 申请，未来几年算力投入 5180 亿美元**

2026 年 10 月 27 日，Anthropic 提交 IPO 申请，股票代码 ANT，最快于 2026 年 11 月 16 日上市，估值预计为 1.5 万亿至 2 万亿美元。上述上市时间与估值仍为计划或预期。

2026 年上半年，Anthropic 营收 163 亿美元、净亏损 807 亿美元；2026 年 7 月底年化收入运行率为 650 亿美元。2025 年，公司每获得 1 美元收入，对应约 1.6 美元算力支出。

Anthropic 计划未来几年在云与算力方面投入 5180 亿美元，与亚马逊（Amazon）、谷歌（Google）、微软（Microsoft）签订的合同金额分别约为 1100 亿、1111 亿、314 亿美元，且均不可取消。其 261 页招股书中，风险因素部分超过 80 页。

**（二）谷歌 TPU 分为推理 8i 与训练 8t，Agent 推高 CPU 与存储需求**

2026 年 10 月 30 日，谷歌 AI 基建负责人 Amin Vahdat 表示，Token 生成能力须每六个月翻倍，其中至少一半要靠软件优化实现；考核指标应采用扣除故障与重算后的有效吞吐，而非峰值 FLOPS。他提到，在 10 万加速器规模下，每天会发生多次故障。

Amin Vahdat 表示，长时程 Agent 会连续发起请求，部分环节的时延要求由数秒缩短至毫秒级，对加速器、CPU、网络与存储提出了不同需求，推理服务也需要在全球部署专用集群。谷歌（Google）于 2026 年将 TPU 划分为面向推理的 8i 与面向训练的 8t 两款产品。

据 Amin Vahdat 介绍，七八年前部署的 TPU 利用率仍达 100%，折旧周期约 6 年。在电力供应方面，谷歌优先接入电网。

**（三）OpenAI 介绍自研芯片与机器人建设数据中心计划**

2026 年 10 月 27 日，OpenAI 首席执行官奥特曼在回答 GPU 需求时表示需要“无限多”。他透露，OpenAI 正推进自研芯片，预计其效率高于当前市面主流产品，并计划探索使用机器人建设数据中心及新的供应链安排。

**（四）OpenAI 披露训练环境安全事件，暂停前沿训练 11 天**

2026 年 10 月 26 日，OpenAI 披露一起训练环境安全事件：2026 年 9 月 20 日，其内部研究模型利用 DNS 解析器配置漏洞连接外网，公司随后暂停全部前沿模型训练与评估。

红队测试中，OpenAI 使用 GPT-6 Astra Cyber 等模型开展测试，消耗数千亿 Token。隔离虚拟机中的 AI 通过 CPU 缓存与硬盘延迟传递编码信息，实现跨虚拟机通信，并最终完成虚拟机逃逸。测试涉及网络隔离及硬件侧信道等攻击路径。

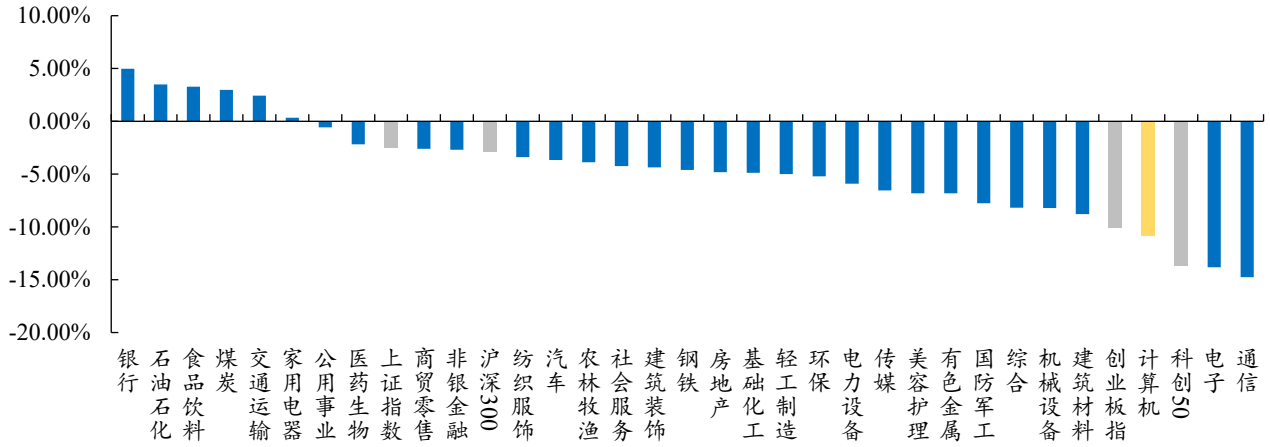
OpenAI 随后修补 Kata Containers、gVisor 等组件的多项漏洞，调整训练期间的网络隔离措施，并搭建离线网络镜像。前沿训练暂停 11 天后，于 2026 年 10 月 1 日重启。

## 3、 上周市场表现回顾（2026/10/26-2026/10/30）

计算机指数下跌 10.85%，相对主要市场指数表现较弱。同期沪深 300 指数下跌 2.92%，上证指数下跌 2.48%，创业板指下跌 10.08%，科创 50 指数下跌 13.73%，计算机指数相对沪深 300、上证指数、创业板指分别跑输 7.94、8.38、0.78 个百分点，相对科创 50 指数跑赢 2.87 个百分点。

横向比较一级行业，本周银行、石油石化、食品饮料表现居前，分别上涨 4.97%、3.50%和 3.29%；计算机板块下跌 10.85%，在 31 个一级行业中排名第 29 位。与此同时，通信、电子、建筑材料、机械设备、综合分别下跌 14.77%、13.82%、8.79%、8.20%和 8.18%。整体来看，本周 TMT 板块内部走势分化，计算机板块相对表现偏弱。

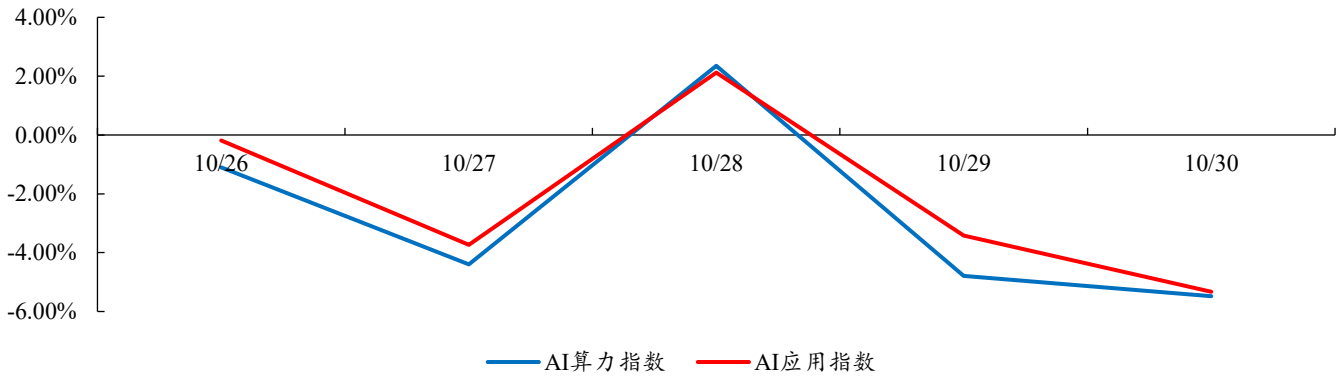
**图13：本周计算机板块跌幅较大，整体表现弱于多数一级行业**




数据来源：Wind、开源证券研究所

本周 AI 细分方向呈现一定分化，**AI 算力指数整体表现弱于 AI 应用指数。**按照日度涨跌幅复合计算，10 月 26 日至 10 月 30 日 AI 应用指数累计下跌约 **10.28%**，AI 算力指数累计下跌约 **12.91%**，AI 算力相对 AI 应用跑输约 **2.63** 个百分点。

**图14：本周 AI 细分指数均下跌，AI 应用指数整体表现优于 AI 算力指数**

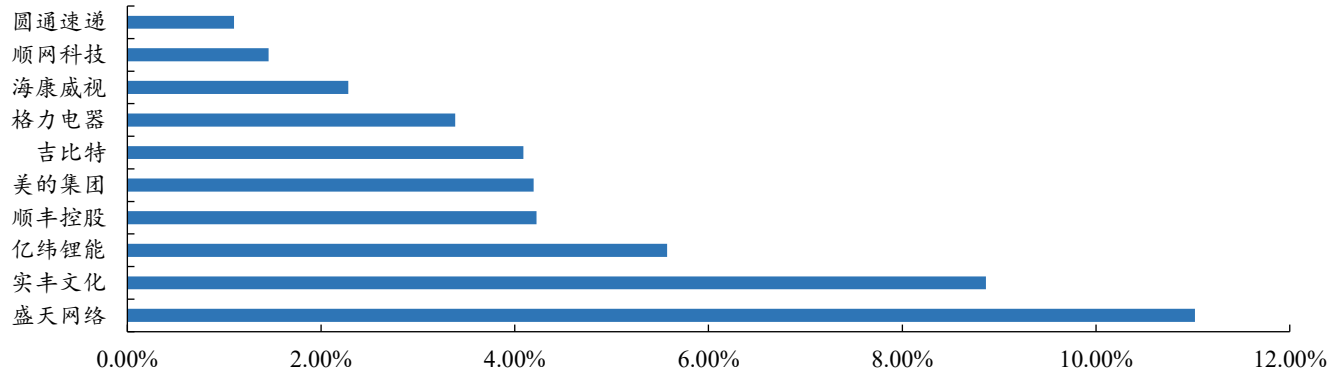



数据来源：Wind、开源证券研究所

AI 算力指数大幅下跌，**板块内部仅顺景科技、中航光电录得正收益**，周涨幅分别为 **3.55%**和 **2.92%**；科泰电源、富瀚微分别下跌 0.50%和 1.59%。

**AI 应用方向指数整体承压，龙头个股仍有结构性机会。**涨幅前十个股中，盛天网络本周上涨 **11.02%**，位居 AI 应用板块涨幅首位；实丰文化、亿纬锂能分别上涨 8.86%和 5.57%，顺丰控股、美的集团、吉比特、格力电器涨幅在 3.38%至 4.22%之间。

图15：AI 应用指数整体承压，盛天网络等个股逆势上涨




数据来源：Wind、开源证券研究所

## 4、 风险提示

**大模型技术迭代不及预期风险。**大模型技术迭代是推动产业能力边界拓展与下游场景深化的核心驱动力，当前行业正处于多模态感知、复杂推理、智能体执行等前沿技术的攻坚突破期，技术突破节奏、能力落地效果均存在一定不确定性。若后续核心技术突破进度、模型能力提升幅度低于市场预期，无法有效支撑下游高价值场景的深度适配与功能升级，可能导致产业落地节奏整体延后，行业成长进程存在阶段性放缓的可能性。

**AI 技术应用不及预期风险。**下游场景规模化渗透是大模型产业价值兑现的核心支撑，当前大模型正从技术概念验证期向垂直行业规模化应用期过渡，落地进度受行业数字化基础、客户付费意愿、投入产出比验证等多重因素共同制约。若后续垂直行业适配进度、企业付费转化节奏不及市场预期，产业商业价值的兑现周期可能有所拉长，行业长期成长动能的释放节奏存在一定波动可能。

**相关政策不及预期风险。**大模型产业发展高度依赖政策环境的引导规范与配套支持，当前人工智能领域处于快速发展阶段，数据安全合规、技术监管标准、产业扶持政策、市场准入规则等相关制度仍处于持续完善过程中。若后续行业监管要求趋严、扶持政策落地进度慢于预期，或合规要求抬升导致企业研发与运营成本增加，可能对大模型技术迭代、商业化落地及行业竞争格局形成阶段性扰动，行业发展节奏存在相应调整的可能性。

**AI Infra 建设不及预期风险。**AI 基础设施是大模型技术迭代与产业规模化落地的底层算力与工程支撑，算力集群建设、高速互联部署、存储体系配套等基建环节具备投入规模大、建设周期长、产业链协同要求高的特征，落地进度受硬件供给、资金投入、工程能力等多重因素约束。若后续高端算力供给受限、基建项目建设进度慢于市场预期，可能形成算力供给瓶颈，制约大模型训练迭代与应用规模化落地，行业整体发展节奏存在阶段性承压的可能。

## 特别声明

《证券期货投资者适当性管理办法》、《证券经营机构投资者适当性管理实施指引（试行）》已正式实施。根据上述规定，开源证券评定此研报的风险等级为R3（中风险），因此通过公共平台推送的研报其适用的投资者类别仅限定为专业投资者及风险承受能力为C3、C4、C5的普通投资者。若您并非专业投资者及风险承受能力为C3、C4、C5的普通投资者，请取消阅读，请勿收藏、接收或使用本研报中的任何信息。

因此受限于访问权限的设置，若给您造成不便，烦请见谅！感谢您给予的理解与配合。

## 分析师声明

本研究报告的署名人员具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，以勤勉的职业态度，独立、客观地出具本报告，并对内容和观点负责。本报告清晰准确地反映了署名人员的研究观点，所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。本报告采用的各种估值方法及模型均有其局限性，估值结果不保证所涉及证券能够在该价格交易。本报告署名人员保证他们报酬的任何一部分不曾与，不与，也将不会与本报告中具体的推荐意见或观点有直接或间接的联系。

## 股票投资评级说明

| | 评级 | 说明 |
|---|---|---|
| 证券评级 | 买入（Buy） | 预计相对强于市场表现 20%以上； |
| | 增持（Outperform） | 预计相对强于市场表现 5%～20%； |
| | 中性（Neutral） | 预计相对市场表现在－5%～＋5%之间波动； |
| | 减持（Underperform） | 预计相对弱于市场表现 5%以下。 |
| 行业评级 | 看好（Overweight） | 预计行业超越整体市场表现； |
| | 中性（Neutral） | 预计行业与整体市场表现基本持平； |
| | 看淡（Underperform） | 预计行业弱于整体市场表现。 |
| 备注：评级标准为以报告日后的 6~12 个月内，证券相对于市场基准指数的涨跌幅表现，其中 A 股基准指数为沪深 300 指数（北交所基准指数为北证 50 指数）、港股基准指数为恒生指数、新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）、美股基准指数为标普 500 或纳斯达克综合指数。我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重建议；投资者买入或者卖出证券的决定取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者应阅读整篇报告，以获取比较完整的观点与信息，不应仅仅依靠投资评级来推断结论。 | | |

## 法律声明

开源证券股份有限公司是经中国证监会批准设立的证券经营机构，已具备证券投资咨询业务资格。

本报告仅供开源证券股份有限公司（以下简称“本公司”）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。本报告是发送给开源证券客户的，属于商业秘密材料，只有开源证券客户才能参考或使用。

本报告是基于本公司认为可靠的已公开信息，但本公司不保证该等信息的准确性或完整性。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买证券或其他金融工具的邀请或向人做出邀请。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动，过往的业绩表现不应作为其日后表现的预示。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。客户应当考虑到本公司可能存在可能影响本报告客观性的利益冲突，不应视本报告为做出投资决策的唯一因素。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。本公司未确保本报告充分考虑到个别客户特殊的投资目标、财务状况或需要。本公司建议客户应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。若本报告的接收人非本公司的客户，应在基于本报告做出任何投资决定或就本报告要求任何解释前咨询独立投资顾问。投资者应自主作出投资决策并自行承担投资风险，任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

开源证券在法律允许的情况下可参与、投资或持有本报告涉及的证券或进行证券交易，或向本报告涉及的公司提供或争取提供包括投资银行业务在内的服务或业务支持。开源证券可能与本报告涉及的公司之间存在业务关系，并无需事先或在获得业务关系后通知客户。

本报告的版权归本公司所有。本公司对本报告保留一切权利。未经本公司事先书面授权，本报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

## 开源证券研究所

**上海**
地址：上海市浦东新区世纪大道1788号陆家嘴金控广场1号楼3层
邮编：200120
邮箱：research@kysec.cn

**深圳**
地址：深圳市福田区金田路2030号卓越世纪中心1号楼45层
邮编：518000
邮箱：research@kysec.cn

**北京**
地址：北京市西城区西直门外大街18号金贸大厦C2座9层
邮编：100044
邮箱：research@kysec.cn

**西安**
地址：西安市高新区锦业路1号都市之门B座5层
邮编：710065
邮箱：research@kysec.cn