仅供内部参考，请勿外传

计算机

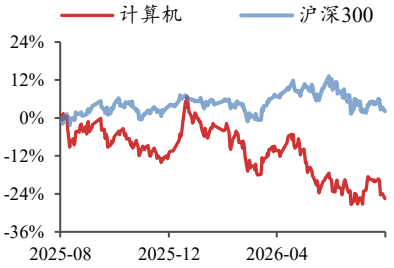
2026 年 08 月 25 日

算力供需缺口持续扩大，算力租赁行业景气度上行

——行业深度报告

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

《DeepSeek API 拟大幅提价，算力景气度高企—行业点评报告》-2026.8.9

《DeepSeek-V4-Flash 超高性价比重塑 Agent 竞争格局—行业点评报告》-2026.8.3

《AI 应用商业闭环逐步跑通，垂直场景迎来价值重估—行业点评报告》-2026.8.2

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

● AI 算力需求结构性增长，算力租赁成为灵活供给核心路径

AI 大模型从训练向推理加速演进，国内算力需求呈现快速增长态势。随着 Agent 智能体、多模态大模型的规模化落地，推理侧算力需求快速增长，成为供需缺口的核心来源。中国信通院数据显示，2026 年一季度国内 AI 算力需求同比增长 417%，有效供给增速仅 128%，供需缺口持续拉大。传统 IDC 建设周期长达 3 年以上，难以匹配 AI 算力快速迭代需求，面临“建成即落后”的扩容窘境。算力租赁以“按需付费、弹性扩容”的轻资产模式应运而生，客户无需大额资金自建 GPU 集群，可快速补足算力缺口并灵活调配规模与区域，或将成为填补供给缺口、降低企业资本开支的核心路径。

● 训练算力战略地位稳固，高端算力租赁价格攀升维持行业高景气

训练端算力仍是大模型厂商核心竞争力的根本支点，终端用户的核心决策锚点始终围绕模型的实际效果与使用体验，缺乏训练端算力支撑、产品力不足的厂商将快速被主流市场出清。受供给约束与需求增长共振，H100 GPU 一年期租赁合约价格从 2025 年 10 月的 1.70 美元/小时低点上涨至 2026 年 3 月的 2.35 美元/小时，累计涨幅近 40%；国内 B300 顶配现货口头报价站上 1450 万元，H200 整机月租同步持续上涨，大模型与智能体项目扩张下训练、微调刚需持续，B 系列新品尚不能完全填补高端训练算力缺口，优质算力进入抢货模式，现货上架即被溢价敲定。2026 年一季度算力租赁市场规模 680 亿元，同比增 62%，全年预计突破 2600 亿元，高景气度持续。

● 算力租赁综合壁垒多维构筑，GPU 获取能力成核心护城河

算力租赁行业属重资产驱动赛道，综合竞争力由 GPU 获取能力、融资能力、订单回款、成本管控及合规风控等多维构筑，行业马太效应持续凸显。GPU 获取能力为核心生存壁垒，体现为供给稳定、交付确定、规模充足、品质保障四大维度。英伟达高端 GPU 在性能、集群及 CUDA 生态上显著领先国产芯片，高阶认证厂商可获原厂直供配额，配额量相较普通伙伴多 30%-50%，交付周期缩短 6-12 个月，缺乏认证的中小厂商依赖二级市场采购，供需紧张时断供压力加大。头部厂商凭借全维度壁垒优势持续抢占高端市场份额，尾部中小厂商生存空间持续收窄。

● 受益标的

行云科技、宏景科技、利通电子、杰创智能、协创数据、航锦科技、润泽科技、东阳光、华策影视、亿田智能、平治信息、盛视科技、盈峰环境、熙菱信息等。

● 风险提示： 互联网大厂资本开支不及预期风险、行业竞争加剧风险、AI 模型发展不及预期、AI 推理需求不及预期。

仅供内部参考，请勿外传

目　录

1、 AI 算力市场重构带来算力租赁商业模式转型 3
1.1、 头部企业加码推理算力基建，国内 AI 基础设施供需持续紧张 3
1.2、 算力服务业态加速迭代，算力租赁成灵活算力供给路径 4
1.3、 算力租赁商业模式从“卖算力”转向“卖 Token” 4
2、 算力租赁供需景气持续上行，行业规模扩容释放成长弹性 7
3、 多重壁垒构筑综合竞争优势，核心资源能力分化行业梯队 11
4、 风险提示 18

图表目录

图 1： 2026 年五大科技巨头 AI 资本开支将大幅增长 3
图 2： 训练和推理，本质上是两套完全不同的系统工程 5
图 3： 训练端仍以英伟达高端 GPU 为核心需求，推理侧市场格局呈现差异化竞争 6
图 4： 算力采购模式从硬件资源转向 Token 生成能力，工厂实现 Token 的高效生产、流转与治理 7
图 5： OpenAI 等全球头部基座厂商，训练研发算力成本占总算力成本比重超 50% 7
图 6： Grace CPU 与 H100 GPU 同板封装的 GH200 8
图 7： Blackwell B200 GPU 裸片与 HBM 内存封装 8
图 8： 英伟达 Rubin 中的第六代 NVIDIA NVLink 与上一代相比，将 GPU 间的通信带宽提升了 2 倍，借助最新的 AI 模型架构实现更快的训练与推理 8
图 9： 英伟达多档位 GPU 1 年期租赁价格分化，高端旗舰型号持续走高 9
图 10： H100 GPU 一年期租赁合约价格从 2025 年 10 月的 1.70 美元/小时低点上涨至 2026 年 3 月的 2.35 美元/小时 10
图 11： 国内大模型厂商参数规模持续抬升，Kimi K3 率先突破 2.8 万亿参数规模 10
图 12： 2026 年 Q1 协创数据营收同比增长 192.90% 13
图 13： 2026 年 Q1 协创数据归母净利润同比增长 343.45% 13
图 14： 服务器再制造依托回收体系对冲成本波动，盈利强劲 13
图 15： 协创数据在北京、上海、乌兰察布、廊坊、中卫等多地部署并运营算力中心 14
图 16： 2026 年 Q1 利通电子营收同比增长 41.61% 15
图 17： 2026 年 Q1 利通电子归母净利润同比增长 821.08% 15
图 18： 2026 年 Q1 宏景科技营收同比增长 7.01% 16
图 19： 2026 年 Q1 宏景科技归母净利润同比增长 39.75% 16
图 20： 行云科技布局全球跨境数字贸易、Token 工厂、液冷技术解决方案及行业 ASIC 芯片等核心业务 17
图 21： 2026 年 H1 行云科技营收同比增长 497.11% 17
图 22： 2026 年 H1 行云科技归母净利润同比增长 540.15% 17

表 1： 相比传统 IDC，算力租赁具有初期投入低、弹性伸缩快、无需运维等优势 4
表 2： AI 算力的训练和推理几乎是两种完全不同的工作负载 5
表 3： NVIDIA Partner Network（NPN）是英伟达全球唯一官方认证的合作伙伴体系 12
表 4： 协创数据持续加大算力租赁业务投入 13
表 5： 多元协同的算力服务业务模式矩阵，是公司业绩持续增长的核心驱动力 16

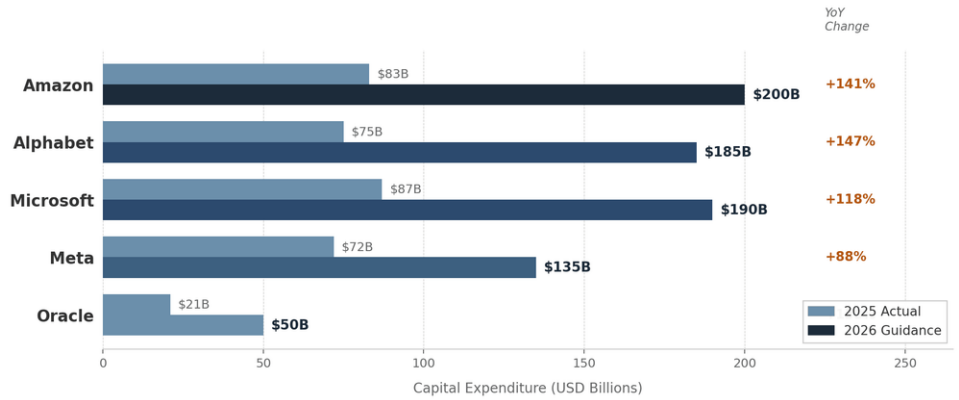
仅供内部参考，请勿外传

1、 AI 算力市场重构带来算力租赁商业模式转型

1.1、 头部企业加码推理算力基建，国内 AI 基础设施供需持续紧张

2022 年生成式 AI 兴起初期，行业竞争集中在模型训练领域。谁能训练出最强大的模型，谁就掌握了竞争优势，这促使企业投入巨额资金，不断堆叠参数和芯片资源，以求实现模型规模与能力的跨越式提升。然而，随着人工智能服务进入常规部署阶段，行业开始从训练算力建设转型为推理算力建设，各大厂积极转向推理类 AI 基础设施投入，2024 年，国内推理工作负载已经超越训练端达到 65%。2022 年，亚马逊、Alphabet、微软、Meta 及甲骨文五大科技巨头资本开支达 1620 亿美元；2025 年，五家企业资本支出总额已攀升至 4480 亿美元，预计至 2030 年，全球 AI 资本支出将达到 1.3 万亿美元。

图1：2026 年五大科技巨头 AI 资本开支将大幅增长



数据来源：paasa 官网

国内 AI 基础设施供需持续紧张已成为产业发展核心约束。随着 Agent 智能体、多模态大模型的规模化落地，国内 AI 算力需求呈现结构性快速增长的特征，推理侧算力需求或将达到训练侧的 4.5 倍以上，成为供需缺口的核心来源。中国信通院数据显示，2026 年一季度国内 AI 算力需求同比增长 417%，但同期有效供给增速仅 128%，供需缺口持续拉大。截至 2026 年，算力短缺态势将至少维持 1-2 年；受 HBM 高带宽内存、先进封装产能瓶颈制约，高端 AI 服务器整机交付周期延长。工信部在《“人工智能+信息通信”创新发展实施意见》官方顶层文件中，已明确将算力供需匹配列为三年核心攻坚任务，当前紧张格局短期内预计难以有效缓解。

在此背景下，国内云服务提供商（CSP）同步明确将持续加大算力领域资本开支。国内互联网厂商增量明显，运营商维持算力投资基本盘规模：三大运营商层面，中国移动 2026 年算力网络资本开支将同比增长 62.4%；中国联通 2026 年的 500 亿元投资额中，算力投资占比将超 35%；中国电信 2026 年算力基础设施投资达 255 亿元，同比增长 26%，AIDC 投资也将增长 28%，算力投入已成为运营商资本开支核心投向；互联网云厂商方面，腾讯 2026 年一季度资本开支 319.4 亿元，同比增长 16%，公司明确全年 AI 相关资本开支将显著增加，下半年随国产芯片交付投入将进一步提升；阿里巴巴已宣布未来三年 AI 与云计算基础设施投入计划不低于 3800 亿元，主要用于建设云和 AI 硬件基础设施，包括数据中心、AI 芯片、高性能网络、智算集群等。

仅供内部参考，请勿外传

1.2、 算力服务业态加速迭代，算力租赁成灵活算力供给路径

在 AI 大模型驱动的新一轮算力需求快速增长中，算力需求的底层逻辑已发生根本性变化。大模型训练与推理对 GPU、NPU 等异构计算资源的需求快速增长，单模型训练所需算力规模动辄数千乃至上万张高性能算力卡协同并行，且硬件迭代周期大幅压缩。传统 IDC 业务以 CPU 为核心、强调稳定托管服务的架构设计，与 AI 算力所需的弹性扩展、异构调度、快速交付之间形成结构性错配。在此背景下，市场亟需一种能够跨越重资产建设周期、直接响应 AI 算力弹性需求的新型基础设施服务模式，算力租赁应运而生。

表1：相比传统 IDC，算力租赁具有初期投入低、弹性伸缩快、无需运维等优势

特性	智能算力服务	云服务	IDC 服务
交付形式	算力词元（机时、卡时、训练任务、渲染作业等	云主机、云存储、数据库等	物理机柜、机房等
技术重心	异构计算、智能调度、高速互联	资源池化、虚拟化	土木、电力、制冷
计费逻辑	按计算任务复杂度或算力使用量	按资源规格和使用时长	按机柜/空间/带宽

资料来源：中国信息通信研究院、开源证券研究所

算力租赁是智算基础设施服务的核心细分业态，指服务商统筹算力硬件、数据中心场地及配套运维体系，向客户交付可直接调用的计算能力，而非单一硬件或物理空间。供给主体涵盖公有云厂商、基础电信运营商及第三方专业算力服务商，资源端通过自建自营、合建联营、设备集采托管、存量资源转租等路径整合，形成重资产自持与轻资产撮合两类运营范式。交付形态覆盖单卡、整机、专属集群等标准化产品，同时延伸出系统集成、项目制整体解决方案、定制化运维等增值服务。核心价值在于降低客户算力获取的资本开支门槛，匹配 AI 产业弹性算力需求，缓解算力供需时空错配矛盾。

中美算力租赁业态因产业发展阶段、供应链约束与客户结构差异，形成差异化发展路径。美国市场以综合云厂商与专业 AI 云服务商共同构成供给主体，头部服务商普遍重资产自建算力集群，绑定头部大模型与科技企业长周期合约，依托高杠杆实现产能快速扩张，头部客户集中度高，收入与现金流确定性强。中国市场呈现资源整合型混合业态，供给侧统筹自有、转租、IDC 托管等多渠道算力资源，交付覆盖算力直租、系统集成、项目制整体解决方案等多元形态，客户结构呈现头部集中与长尾分散并存的特征，轻重资产模式并行，适配国内算力网络调度与产业数字化落地需求。

1.3、 算力租赁商业模式从“卖算力”转向“卖 Token”

AI 算力的训练和推理几乎是两种完全不同的工作负载。AI 训练的核心，是通过海量标注数据经算法迭代求解模型最优参数。该阶段需要持续数据输入、海量参数动态更新及数周至数月不间断运行，追求规模化运算效率。这意味着训练芯片不仅要具备强悍的算力，还需配备极高的显存带宽、高效的分布式通信能力，以及万卡级集群规模下的稳定性。推理则是模型基于已训练参数对新数据进行预测、生成响应的“运用知识”阶段，是 AI 落地的核心环节。相较训练，推理更侧重低延迟、高能效与成本控制，部署覆盖云、边、端全场景，无需迭代训练即可直接调用成熟模型，在实时响应场景中高效运行。

仅供内部参考，请勿外传

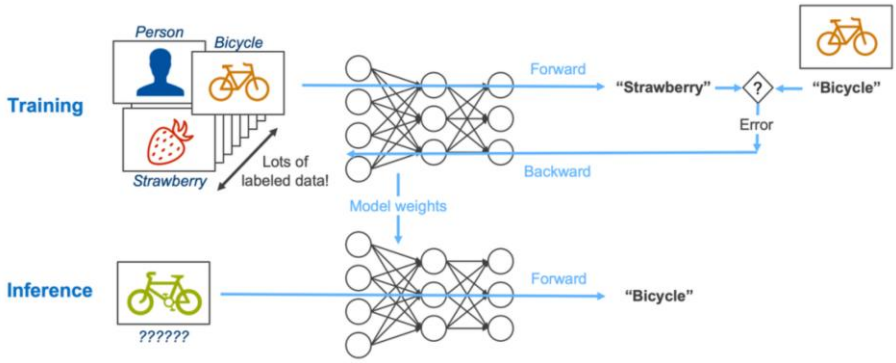
表2：AI 算力的训练和推理几乎是两种完全不同的工作负载

	训练（Training）	推理（Inference）
目的（Purpose）	学习规律	应用规律
过程（Process）	复杂迭代优化（前向+反向）	简单单向计算（仅向前）
数据（Data）	海量标注数据	少量新数据
资源（Resources）	计算量大、耗时长	计算量小、延迟低
结果（Result）	优化后的模型参数	对新数据的预测/生成结果

资料来源：昊源诺信公众号

大模型训练与推理的算力架构存在差异，训练算力进行推理效果较差。推理包含预填充与解码两个异构阶段：预填充属计算密集型任务，对显存带宽要求高；解码阶段则为访存密集型，计算单元利用率显著下降，瓶颈显现于内存访问速度。训练集群如果用来推理会采用同构 GPU 统一处理两阶段任务，解码阶段算力资源利用率低下，整体效率显著低于经专门优化的推理架构。从成本结构看，推理算力的经济性取决于架构与负载特性的匹配度，而非单卡峰值性能。以训练标准建设的集群承载推理任务，硬件投入与有效算力产出严重错配，单位推理成本大幅抬升，直接侵蚀运营端盈利空间，智算中心投资回报率承压。

图2：训练和推理，本质上是两套完全不同的系统工程

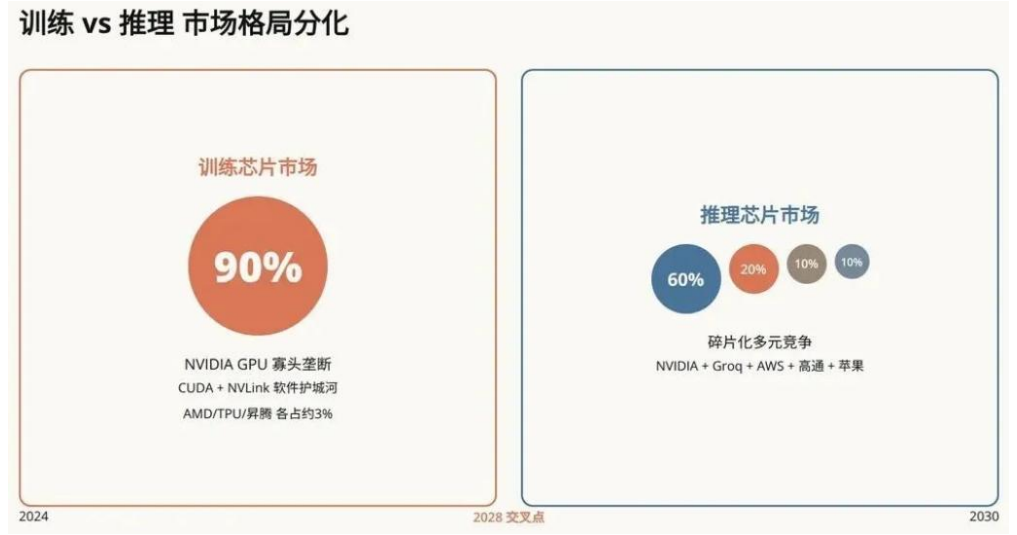


资料来源：昊源诺信公众号

推理需求快速增长正推动算力租赁行业进入结构性发展阶段。训练端对极致算力、软件生态与集群扩展能力要求极高，当前仍以英伟达高端 GPU 为核心需求，头部厂商凭借资金与供应链配额优势占据主导，订单以大额长周期、集中交付为核心特征。推理侧算力需求占比持续攀升，而推理场景对单卡极致算力要求相对宽松、更看重性价比与供应链稳定性，为国产推理卡带来规模化落地窗口。同时推理算力需求分散、波动弹性大，传统训练集群承载推理负载利用率偏低，行业红利正向具备精细化调度能力的厂商集中。

仅供内部参考，请勿外传

图3：训练端仍以英伟达高端 GPU 为核心需求，推理侧市场格局呈现差异化竞争



资料来源：驭势资本公众号

传统训练算力租赁模式中，算力租赁服务商与下游客户达成合作后，普遍签署 3-5 年期长期租赁协议，待算力集群验收交付完成后正式起算费用，按月收取租赁服务费。2026 年 5 月联想集团董事长兼 CEO 杨元庆在业绩发布会上指出，现阶段行业 GPU 服务器资源中，70%-80%算力资源集中供给大模型训练场景，仅 20%-30%用于业务推理环节。随着各类 AI 应用规模化落地，算力分配结构将迎来反转，远期推理场景算力消耗占比将突破 70%。**推理算力租赁客户分散、业务并发高、算力需求弹性大且对延迟高度敏感，按卡按时长、按项目总包已非最优结算方式。**

为应对推理算力带来的新需求，算力租赁行业的商业模式正迎来结构性转变。行业盈利载体或将由传统裸算力资源出租，逐步过渡至面向终端客户的 Token 计费服务。当前行业高端算力供给持续偏紧，算力租赁商议价能力显著增强，推理算力比例持续提升，业务形态逐步脱离单纯硬件算力出租，延伸出模型托管、Token 流量分成等增值服务体系，有望完成从“卖算力”到“卖 Token”的转型。该商业模式升级能够显著增厚服务商盈利水平并适应未来算力发展。

Token 指标的意义在于推动算力服务从“资源供给”走向“能力交付”。 Token 指标的价值，不只是一个新的统计单位，更重要的是为大模型时代的算力服务建立了一套贴近业务结果的表达方式。对于采购方来说，Token 相关指标有助于从“买资源”转向“买能力”，更清楚地比较不同平台、不同集群、不同服务方案的实际交付效果；对于服务商来说，Token 评价有助于展示自身在部署优化、资源调度、稳定运行和成本控制方面的综合能力。对于产业发展而言，Token 评价体系能够在芯片、整机、云服务、模型应用、平台运营和测试验证之间建立共同语言，减少单纯围绕峰值算力、设备数量或单点性能进行比较带来的偏差，引导行业更加关注真实业务场景下的服务能力、资源效率和持续运营能力。

仅供内部参考，请勿外传

图4：算力采购模式从硬件资源转向 Token 生成能力，Token 工厂实现 Token 的高效生产、流转与治理



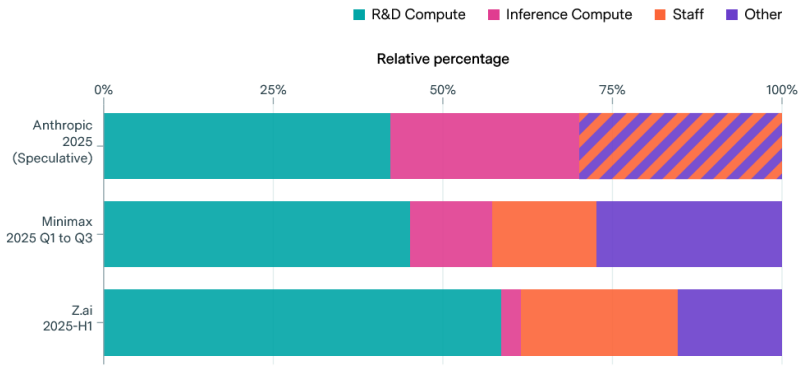
资料来源：云计算与数字化研究所公众号

云服务商与 AI 模型企业的算力储备规模已成为制约整体营收增长的核心要素。即便模型具备顶尖的算法性能，若无充足且稳定的算力资源作为底座，也难以完成商业化落地与规模化推广。在此背景下，算力租赁厂商的价值属性进一步提高：其提供的算力资源能够直接赋能下游客户增收创收。基于上下游深度绑定的合作关系，行业后续有望普及 Token 收益分成模式，算力供应商可根据下游客户 AI 服务的实际调用体量，按约定比例共享收益，打造互利共赢的新型盈利模式。

2、 算力租赁供需景气持续上行，行业规模扩容释放成长弹性

尽管推理侧算力需求随大模型商业化落地持续扩容，但训练端算力投入仍是大模型厂商构建核心竞争力的根本支点。推理算力对应模型落地后的业务调用与商业化变现环节，其需求增长更多由用户规模、调用频次与场景覆盖范围驱动；而训练端算力是支撑模型参数规模、决定迭代效率、影响能力边界的核心硬件基础，是大模型产品力的核心来源与底层支撑。当前通用大模型赛道竞争趋于白热化，行业呈现“赢家未必通吃、落后者加速淘汰”的竞争格局，终端用户的核心决策锚点始终围绕模型的实际效果与使用体验，缺乏训练端算力支撑、产品力不足的厂商将快速被主流市场出清。据 Epoch AI 统计，OpenAI、Anthropic 等全球头部基座厂商，训练对应的研发算力成本占总算力成本比重超 50%，训练侧算力的战略优先级并未因推理需求快速增长而出现下移。

图5：OpenAI 等全球头部基座厂商，训练研发算力成本占总算力成本比重超 50%



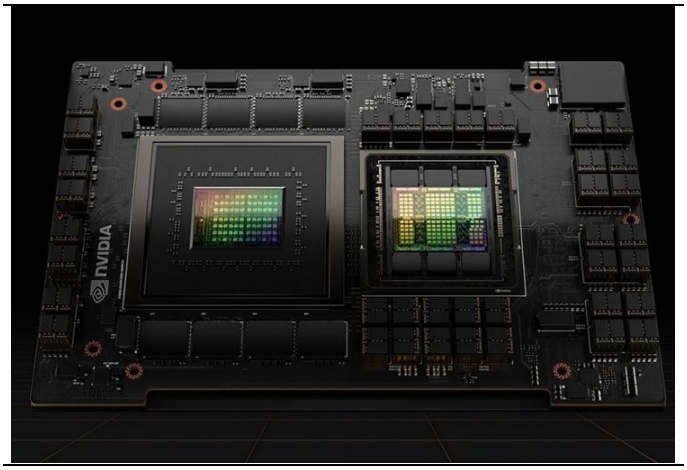
数据来源：Epoch AI 官网

英伟达高端 GPU 在硬件性能、集群协同能力及软件生态体系上相较国产主流

仅供内部参考，请勿外传

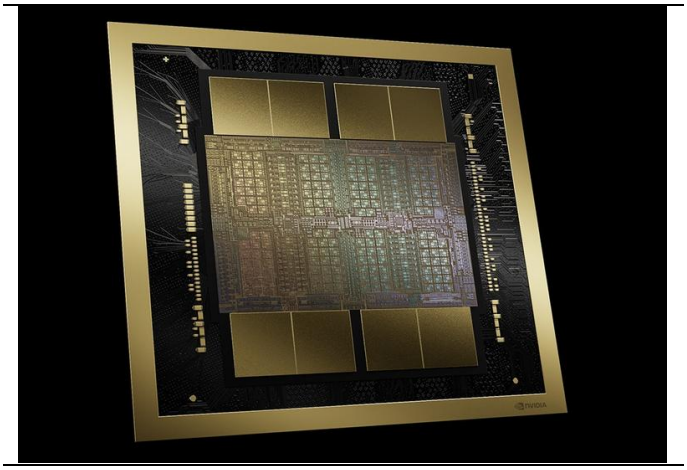
AI 芯片仍具备显著领先优势。根据 SemiAnalysis 与 Counterpoint Research 信息测算，国产旗舰芯片单卡实测性能约为英伟达 H100 的 60%，若对标 Blackwell 架构 B200 旗舰产品，综合性能存在明显代际差距。集群层面，英伟达凭借 NVLink 高速互联与成熟分布式训练框架，可支持万卡级集群高效协同；受高速互联与软件栈差距制约，国产多数集群方案的并行扩展效率普遍弱于英伟达。软件生态方面，CUDA 生态历经十余年积累，覆盖全栈开发工具与海量算法库，国产芯片虽通过兼容路线降低迁移门槛，但算子支持完整度、调优工具成熟度仍存明显差距，高端训练场景综合替代性价比不足。

图6：Grace CPU 与 H100 GPU 同板封装的 GH200



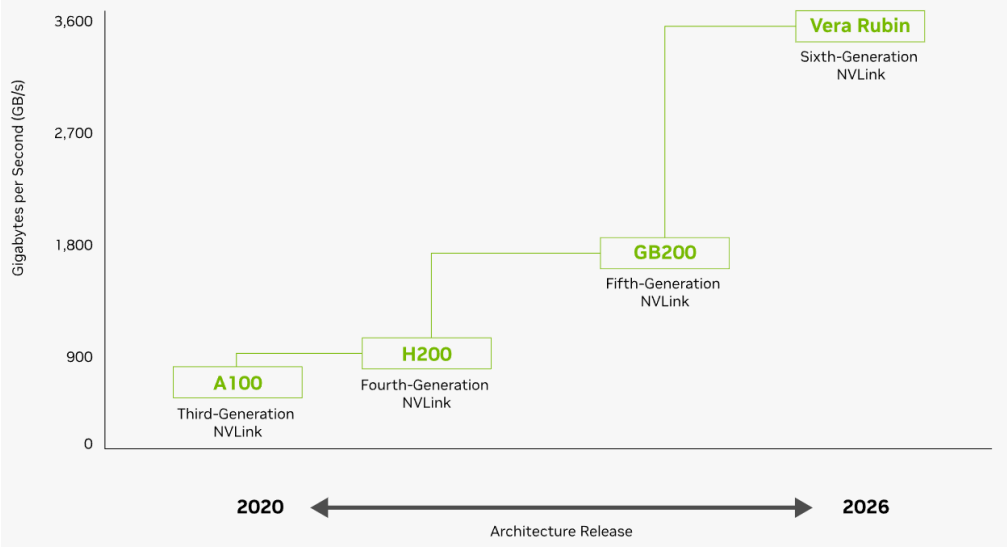
资料来源：英伟达官网

图7：Blackwell B200 GPU 裸片与 HBM 内存封装



资料来源：英伟达官网

图8：英伟达 Rubin 中的第六代 NVIDIA NVLink 与上一代相比，将 GPU 间的通信带宽提升了 2 倍，借助最新的 AI 模型架构实现更快的训练与推理



资料来源：英伟达官网

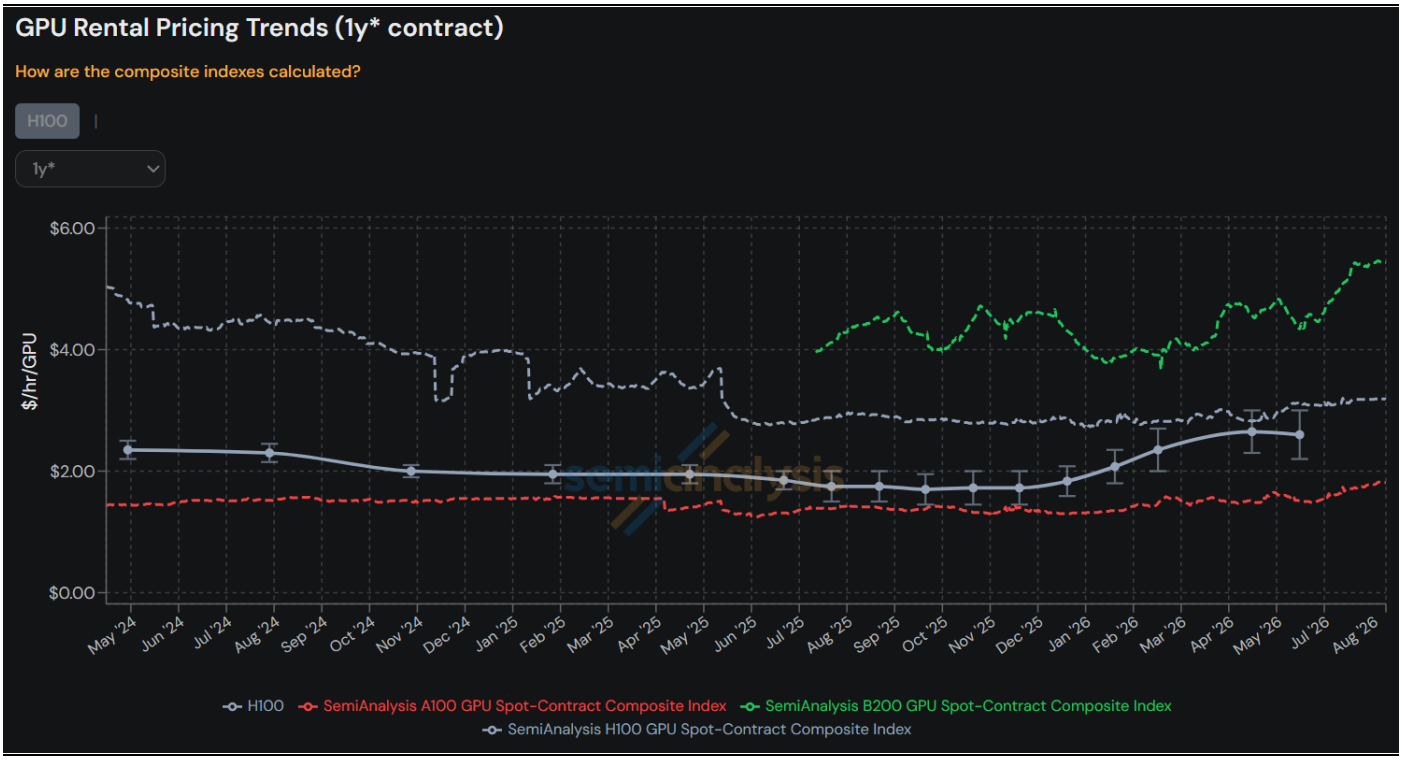
当前国内算力租赁市场呈现显著结构性分化，英伟达 GPU 仍是高端训练场景的核心需求载体，市场主导地位稳固。从需求结构看，训练端对芯片性能、生态兼容性及集群稳定性要求严苛，国产芯片在千亿参数以上大模型的训练效率、生态成熟度上仍存明显差距，主要承接政企通用推理与中小模型微调任务。训练市场被

仅供内部参考，请勿外传

NVIDIA（含其 CUDA 生态）占据超过 90%的份额，AMD、Google TPU 和华为昇腾虽在各自封闭生态中发起挑战但开放市场占有率较低，供需缺口持续扩大进一步强化英伟达 GPU 在租赁市场的稀缺性价值。

受全球供给约束与 AI 训练需求增长共振影响，英伟达高端 GPU 租赁价格持续上行。根据 SemiAnalysis 数据，H100 GPU 一年期租赁合约价格从 2025 年 10 月的 1.70 美元/小时低点上涨至 2026 年 3 月的 2.35 美元/小时，累计涨幅近 40%，且现货资源基本售罄；Blackwell 架构 B200 因产能爬坡缓慢，新订单交付周期延长至 12-15 个月，市场租赁溢价更为显著；国内 B300 顶配现货口头报价达 45 万元，大模型与智能体项目扩张下训练、微调刚需持续，B 系列新品尚不能完全填补高端训练算力缺口，优质算力进入抢货模式，现货上架即被溢价敲定。受台积电先进封装产能限制，高端 GPU 供需紧平衡状态或将贯穿 2026 年全年，租赁价格中枢有望维持上行趋势。

图9：英伟达多档位 GPU 1 年期租赁价格分化，高端旗舰型号持续走高



资料来源：semianalysis 官网

仅供内部参考，请勿外传

图10：H100 GPU 一年期租赁合约价格从 2025 年 10 月的 1.70 美元/小时低点上涨至 2026 年 3 月的 2.35 美元/小时

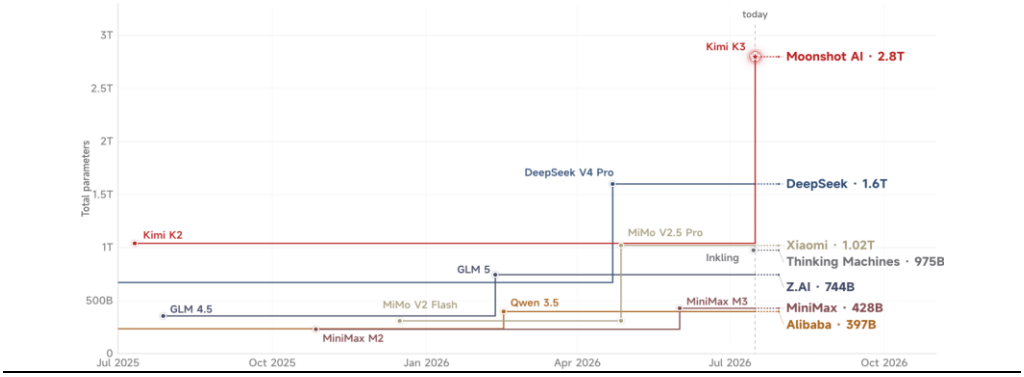


资料来源：semianalysis 官网

国内高端 AI 算力需求持续加大，大模型厂商、互联网科技企业及云服务商共同构成算力租赁市场的核心需求主体。生成式 AI 技术迭代加速推动大模型训练、微调及推理全链路算力需求增长，而受美国先进计算芯片出口管制持续加码影响，英伟达高端训练 GPU 对华交付周期拉长、供给规模受限，国内高端算力供给缺口持续扩大。分主体来看，大模型厂商是高端训练算力的核心采购方，主要用于基座模型训练与版本迭代；云服务商通过租赁算力补充自有算力池弹性，承接溢出的企业级需求；互联网科技企业则以模型微调、场景化落地需求为主。

大模型厂商是高端训练算力租赁的核心需求主体，基座模型参数军备竞赛与快速迭代驱动算力需求持续攀升。当前通用大模型赛道竞争日趋白热化，产品力直接决定厂商市场地位，参数规模扩张仍是通用基座模型能力提升的核心路径之一。月之暗面发布的 Kimi K3 模型总参数达 2.8 万亿，业内测算其预训练需数千至万张级高端 GPU 连续运转数月；字节跳动已启动超 5 万亿参数大模型的前期论证工作，若落地算力需求规模较现有水平将大幅提升。受美国芯片出口管制影响，大模型厂商自主采购高端 GPU 配额受限、交付周期长、采购条款严苛，通过算力租赁快速获取规模化集群资源成为核心路径，保障模型迭代节奏与市场竞争优势。

图11：国内大模型厂商参数规模持续抬升，Kimi K3 率先突破 2.8 万亿参数规模



资料来源：月之暗面 Kimi 公众号

仅供内部参考，请勿外传

互联网科技企业是算力租赁市场的中坚需求力量，AI 技术全场景落地与推理业务放量驱动算力需求持续扩容。头部互联网企业将大模型能力深度嵌入内容生成、智能推荐、搜索、智能客服等核心业务场景，需周期性开展行业垂类模型适配微调，并持续扩容推理算力支撑业务落地。以字节跳动、腾讯为代表的企业，一方面基于通用基座模型开展业务场景适配优化，另一方面应对用户调用量波动带来的推理算力峰谷差。相较于自建算力集群的高额固定资本开支，租赁模式可灵活匹配业务淡旺季的算力需求波动，显著降低峰值时段的资源闲置成本，适配互联网业务快速迭代、需求弹性大的产业特征。

云服务商是算力租赁市场重要的批量采购需求方，通过外部算力采购补充自有算力池的供给弹性，承接溢出的企业级客户需求。国内公有云厂商虽自有大规模算力基础设施，但高端训练 GPU 受出口管制影响供给缺口显著，面对头部客户的长周期大模型训练需求时，自有产能难以足额覆盖。阿里云、腾讯云等头部云厂商在算力紧张阶段，均会向第三方专业算力租赁商采购算力集群资源，用以保障客户订单交付、补足自有算力配额缺口；同时云厂商自身也对外提供标准化算力租赁服务，与专业租赁商形成既有竞争又有合作的产业生态格局。

算力租赁业务合作模式灵活多元，可适配不同企业的资金实力与业务场景，头部厂商凭借早期产能锁定与集群布局进入商业化兑现阶段，行业盈利正向飞轮逐步启动。当前行业主流模式从资源隔离维度分为专属部署与整卡共享分时租赁两类，专属部署模式提供物理隔离的算力集群，兼具高稳定性与数据安全性，适配头部大模型厂商长周期预训练等核心场景；分时租赁模式通过整卡按时共享降低使用门槛，适配中小主体模型微调、推理测试等轻量需求。国内头部算力租赁企业提前锁定 GPU 产能订单，旗下旗舰万卡级算力集群已陆续完成交付验收并进入正式计费阶段，营收与盈利兑现节奏加快。

国内算力租赁市场处于高速扩容通道，供需共振驱动市场规模持续攀升，行业增长确定性与成长空间显著。据中国信通院监测数据，2026 年一季度国内算力租赁市场规模达 680 亿元，同比增长 62%，预计全年规模将突破 2600 亿元。驱动层面，大模型训练迭代与商业化推理需求持续释放，叠加高端芯片供给约束下自建算力门槛抬升，算力租赁模式渗透率有望稳步提升。中长期看，AI 技术向千行百业深度渗透将持续打开增量需求空间，行业高景气度具备持续性，市场规模中枢有望稳步上移。

3、 多重壁垒构筑综合竞争优势，核心资源能力分化行业梯队

GPU 获取能力、融资能力、订单结构与回款质量、运营成本管控、合规风控等共同构筑综合竞争力护城河。算力租赁属于典型重资产驱动赛道，单万卡级高端智算集群资本开支超数十亿元，低成本、稳定的中长期融资渠道是企业规模化扩张的基础支撑，直接决定产能扩张节奏；长周期大额订单与健康的回款结构决定企业现金流安全边际，是锁定长期盈利、对冲算力价格波动的核心抓手；液冷技术落地带来的 PUE 优化可直接摊薄长期运营电费成本，影响盈利弹性空间；出口管制持续加码背景下，供应链合规与数据安全合规则是行业存续的底线要求。行业马太效应持续凸显，头部厂商凭借全维度壁垒优势持续抢占高端市场份额，尾部中小厂商生存空间持续收窄。

GPU 获取能力是算力租赁厂商的核心生存壁垒，集中体现为供给稳定性、交付确定性、规模充足性与品质保障性四大维度。英伟达官方合作伙伴网络（NPN）分

仅供内部参考，请勿外传

为 Registered（注册级）、Preferred（优选级）、Elite（精英级）三级，同时配套 NCP 云合作伙伴认证，高阶认证厂商可获得原厂直供配额，交付周期较普通分销渠道缩短 6-12 个月，配额规模高出 30%-50%，且可保障芯片正品溯源与原厂技术支持。在美国对华芯片管制加码背景下，高阶认证稀缺性持续凸显，缺乏官方认证的中小厂商仅能通过二级市场采购，采购成本普遍偏高，交付节奏不可控且卡源质量存疑，供需紧张阶段将率先面临断供风险。

表3：NVIDIA Partner Network（NPN）是英伟达全球唯一官方认证的合作伙伴体系

维度	Registered（注册级/入门级）	Preferred（优选级/进阶级）	Elite（精英级/最高级）
身份认证	完成基础注册、少量技术认证、无强制销售额要求	通过多项专业技术认证、达成年度销售目标、具备基础服务团队	顶级技术认证（多领域）、高额年度营收、7×24 小时专业服务团队、行业标杆案例、全球 / 区域市场影响力
核心权益	基础产品资料、入门培训、标准市场素材、有限技术支持	优先产品供货、进阶培训、联合营销资源、专属客户经理、中端方案授权	**优先稀缺资源（H200/B200/DGX）、最高返利、原厂深度技术支持、新品首发测试、战略项目共研、全球品牌背书**
身份定位	入门级合作方，以基础产品分销、小型项目为主，无高端方案与服务权限	主流合作伙伴，覆盖区域市场与中型项目，可提供常规 AI 服务器、GPU 部署服务	英伟达生态**核心战略伙伴**，承载高端产品、超算方案、大型数据中心项目的交付与服务

资料来源：宽恒科技官网、开源证券研究所

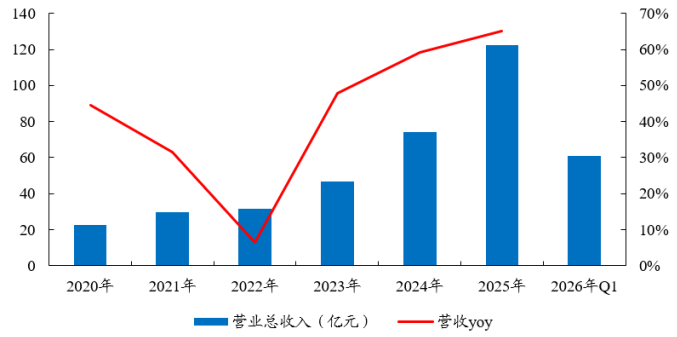
（1）协创数据

协创数据是国内深耕智能物联赛道的创业板上市企业，以“智能物联生态”为核心战略，构建“智能终端+算力基建+云端服务”三位一体的全球化产业体系。公司 2005 年成立、2020 年登陆深交所创业板，智能终端业务覆盖 AIoT 智能硬件、数据存储终端、边缘计算设备等品类，聚焦音视频感知、无线通信连接与场景化 AI Agent 研发，形成感知层与边缘层复合产品矩阵。算力基建端，公司依托控股子公司麦塔倍斯、奥佳软件布局全球算力节点，覆盖大模型训推、云视频等场景，具备异构算力纳管与服务器运维能力。云端服务侧打造 Fcloud 智能体训推平台，集成多模态大模型工具链，向生物制药、集成电路等垂直领域输出定制化 AI 方案，全价值链整合构筑全球竞争壁垒。

公司把握 AI 与算力基础设施建设机遇，以智能算力产品及服务为核心驱动力，带动整体营收规模实现大幅增长。公司智能算力集群陆续交付计费，收入大幅跃升，为核心增长引擎；服务器再制造依托回收体系对冲成本波动，盈利强劲；数据存储凭长协模式保持稳健，充当业绩压舱石；物联网终端主动调结构，聚焦高附加值产品并拓展海外市场。2025 年实现营收 122.36 亿元，同比增长 65.13%，归母净利润 11.64 亿元，同比增长 68.32%；2026 年 Q1 实现营收 60.85 亿元，同比增长 192.90%，归母净利润 7.50 亿元，同比增长 343.45%。

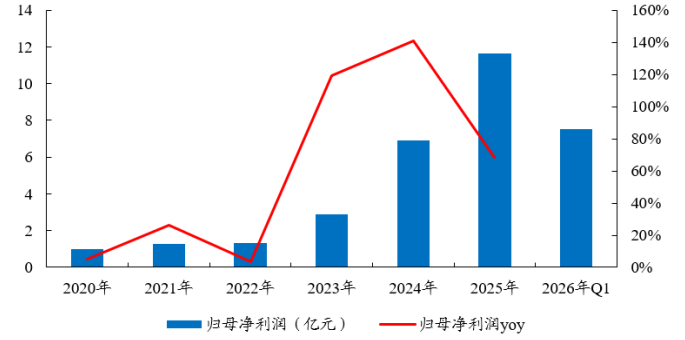
仅供内部参考，请勿外传

图12：2026 年 Q1 协创数据营收同比增长 192.90%



数据来源：Wind、开源证券研究所

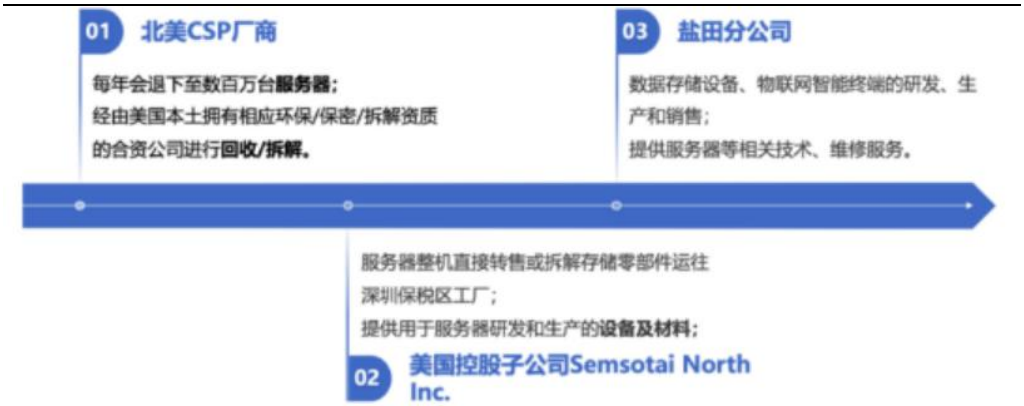
图13：2026 年 Q1 协创数据归母净利润同比增长 343.45%



数据来源：Wind、开源证券研究所

公司经营模式主要包括云服务、硬件制造和服务器再制造业务。云服务主要包括 VSaaS 视频云服务、AIGC视频生成领域和云算力服务；硬件制造主要是以合作研发制造（JDM）和自主设计制造（ODM）模式为客户提供支持和服务；服务器再制造业务方面，公司控股子公司在北美开展服务器的回收、拆解，并经由协创数据深圳盐田分公司和协创数据保税区分公司对购买的设备及材料经过严格测试，性能达标的服务器重新销售业务，针对整机不再适用的设备，拆解出高价值通用部件，进入二级市场流通。

图14：服务器再制造依托回收体系对冲成本波动，盈利强劲



资料来源：协创数据公司公告

协创数据依托奥佳软件与英伟达的深度合作，深耕算力租赁与定制化运营服务。奥佳软件是协创数据间接控股的核心算力业务主体，已正式取得 NVIDIA CLOUD PARTNER 资质。该资质是英伟达合作伙伴的高级别认证，表明奥佳软件在 AI 算力租赁和云计算服务方面具备了与英伟达合作的资质和能力。公司把握行业算力基础设施建设需求增长机遇，依靠与英伟达的合作，持续推进算力服务器及集群解决方案落地，在北京、上海、乌兰察布、廊坊、中卫、广州、深圳、宁波等多地部署并运营算力中心，为客户提供以高端算力为基础的云算力租赁业务。重点服务于头部互联网公司模型训练推理、自动驾驶、生物医药等领域客户。

表4：协创数据持续加大算力租赁业务投入

公告时间	合同类型	交易标的	标的用途	标的价格
2026-02-13	购买资产	服务器	提供云算力服务	110.00 亿元
2025-12-17	购买资产	服务器	提供云算力服务	90.00 亿元

仅供内部参考，请勿外传

公告时间	合同类型	交易标的	标的用途	标的价格
2025-10-22	购买资产	服务器	提供云算力服务	40.00 亿元
2025-08-16	购买资产	服务器	提供云算力服务	12 亿元
2025-05-28	购买资产	服务器	提供算力租赁服务	40 亿元
2025-03-08	购买资产	服务器	提供算力租赁服务	30 亿元

数据来源：协创数据公司公告、开源证券研究所

图15：协创数据在北京、上海、乌兰察布、廊坊、中卫等多地部署并运营算力中心



资料来源：协创数据公司公告

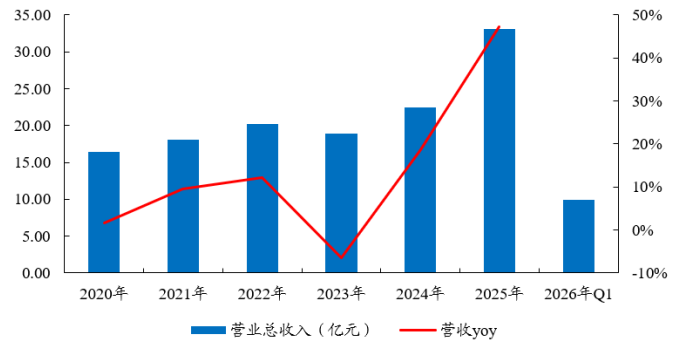
（2）利通电子

利通电子是国内液晶显示结构件龙头企业，已形成“精密制造+AI 算力”双主业格局，正加速向 AI 算力综合服务商转型。公司 1991 年成立、2018 年登陆上交所主板，传统主业为液晶电视精密金属结构件与电子元器件研发制造，2023 年公司战略性切入 AI 算力赛道，通过全资子公司世纪利通开展 GPU 算力租赁、设备经销与集群运维服务，算力规模快速扩张，业务收入与盈利水平大幅跃升，已成为公司核心利润来源与增长引擎，是传统制造企业向 AI 算力赛道转型的代表性标的。

公司目前已经进入成熟和稳定的双主业发展时期，其中 AI 算力业务收入与利润大幅增加，已完全成为推动公司未来增长的重要引擎。受益于 AI 算力需求的持续增加以及前期算力租赁项目的持续执行，公司的算力业务已形成规模化、可持续发展态势，收入、利润增长显著，2025 年实现营收 33.07 亿元，同比增长 47.16%，归母净利润 2.93 亿元，同比增长 1088.59%；2026 年 Q1 实现营收 9.97 亿元，同比增长 41.61%，归母净利润 2.71 亿元，同比增长 821.08%。

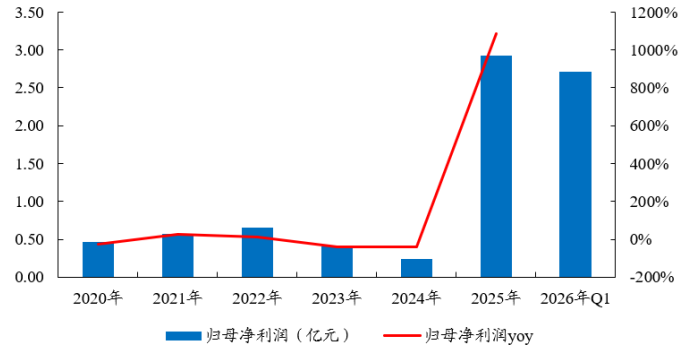
仅供内部参考，请勿外传

图16：2026 年 Q1 利通电子营收同比增长 41.61%



数据来源：Wind、开源证券研究所

图17：2026 年 Q1 利通电子归母净利润同比增长 821.08%



数据来源：Wind、开源证券研究所

公司与英伟达、超聚变等 GPU 服务器制造商、机房、需求端的诸多头部企业建立起战略合作关系，供需两端的资源优势为 AI 算力业务开展建立了牢固的商业基础。全资子公司世纪利通与全资子公司利通控股积极布局海内外算力业务市场，已建立起专业的算力业务团队，贯通了业务链的上下游资源。其中，世纪利通是英伟达认定的在 DGX AI 服务器（DGX AI Compute Systems）方面优先级（perferred）的云业务伙伴（Cloud Partner)。2025 年，公司用于经营的 AI 算力资源已达 38,000P 以上，AI 算力规模在业内居于前列。

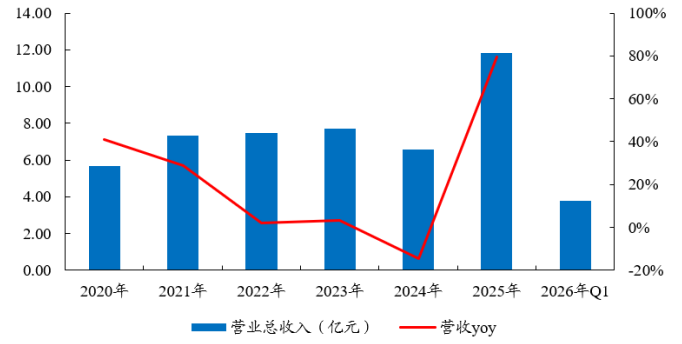
（3）宏景科技

宏景科技是国内深耕智能算力服务与智慧城市解决方案企业，依托二十余年技术积淀构建一体化算力服务体系。公司 1997 年成立以来持续推进人工智能、大数据、智算技术的行业应用与成果转化，具备 HPC 集群、异构计算集群及数据中心全栈建设能力，依托专家技术团队与供应链整合优势，面向 AI 产业提供全链条智能算力服务，覆盖大模型训推、科学计算等核心场景。智慧城市领域为多行业客户提供全流程系统解决方案，实现计算、网络、存储与应用协同创新，推动算力基础设施高质量发展，充分释放算力对产业升级与数字经济的核心驱动价值。

政策红利与 AI 产业需求高景气驱动公司算力业务大幅增长。国家出台算力基础设施相关顶层政策，明确算力新型生产力定位，打开行业发展空间。AI 大模型从测试走向产业落地，Token 调用量快速增长，企业推理需求持续释放，叠加 MoE 等模型技术迭代，带动算力需求高增。公司依托人才、技术、供应链及项目交付积淀，充分把握产业红利，为业务扩张提供坚实支撑，2025 年实现营收 11.82 亿元，同比增长 79.62%，归母净利润 0.26 亿元，同比增长 134.94%；2026 年 Q1 实现营收 3.77 亿元，同比增长 7.01%，归母净利润 0.30 亿元，同比增长 39.75%。

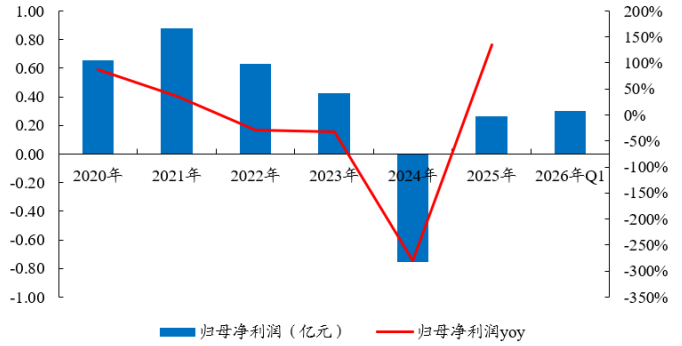
仅供内部参考，请勿外传

图18：2026 年 Q1 宏景科技营收同比增长 7.01%



数据来源：Wind、开源证券研究所

图19：2026 年 Q1 宏景科技归母净利润同比增长 39.75%



数据来源：Wind、开源证券研究所

公司已打造多元协同的算力服务业务模式矩阵。公司的算力服务可覆盖不同客户场景需求，构建“项目制收入+经常性收入+高附加值开发收入”的多层次业绩支撑体系。算力设备集成服务采用项目制交付，验收后一次性确认收入，直接受益于下游智算中心建设扩容浪潮，可快速兑现收入增量；算力设备运营服务以直租、转租两种模式开展，运营期内保留资产控制权并按月确认收入，形成稳定经常性现金流，筑牢业绩基本盘；算力平台定制开发服务依托自研集群调度技术，切入高附加值平台化场景，持续提升盈利水平与客户粘性。

表5：多元协同的算力服务业务模式矩阵，是公司业绩持续增长的核心驱动力

公告时间	合同类型	合同总金额	提供的项目及服务内容
2025-05-20	智算项目服务合同	59699.20 万元	服务器、组网配套服务以及对服务器进行必要的改配服务，并提供算力服务
2025-05-14	智算项目服务合同	56320.00 万元	服务器、组网配套服务以及对服务器进行必要的改配服务，并提供算力服务
2025-04-16	项目集成服务合同	23458.00 万元	算力组网集成服务
2025-04-01	中国移动宁夏公司 2025 年算力服务引入采购框架协议	31152.04 万元	算力服务
2025-03-19	项目集成服务合同	16128.00 万元	算力服务器采购、组网调试、服务器改配调优服务
2025-03-12	智算项目服务合同	72089.60 万元	服务器、组网配套服务以及对服务器进行必要的改配服务，并提供算力服务
2025-02-25	项目集成服务合同	16128.00 万元	算力服务器采购、组网调试、服务器改配调优服务
2024-11-19	智算项目服务合同	40908 万元	硬件资源设备、组网配套服务及对服务器进行必要的改配服务，并提供算力服务
2024-10-30	算力服务合同	48569.32 万元	算力服务
2024-10-22	智算项目服务合同	40908 万元	硬件资源设备、组网配套服务及对服务器进行必要的改配服务，并提供算力服务

资料来源：宏景科技公司公告、开源证券研究所

深博信息是宏景科技的重要关联方与算力硬件上游合作渠道，具备高端 GPU 芯片稳定供给能力，可有效支撑公司算力业务的硬件采购需求。深博信息由宏景科技董事、副总经理的配偶实际控制。深博信息成立于 2009 年，深耕 IT 硬件分销与

仅供内部参考，请勿外传

算力基础设施领域，厂商授权体系层级完备。成立初期即成为 IBM、思科、微软等国际厂商核心合作伙伴，2014 年取得浪潮广东区总代理资质，2021 年晋升英伟达 Elite NPN 精英级合作伙伴，同时持有超聚变、新华三、锐捷、焱融等厂商高级别代理权限，全品类高等级授权构筑了稳定的高端算力硬件供货保障。

（4）行云科技

行云科技以数字贸易为场景入口、以算力基础设施为底层支撑，聚焦跨境数字贸易与 AI 算力基础设施融合发展。公司 2000 年成立，核心布局全球跨境数字贸易、Token 工厂、液冷技术、ASIC 芯片四大业务板块：数字贸易端打造全链路数字化基础设施，夯实跨境服务能力；算力端依托 Token 工厂搭建弹性高效的 Token 生成与交付体系；同时深耕液冷与 ASIC 芯片等异构计算领域，推动 GPU、液冷、ASIC 协同演进，持续降低 AI 应用落地门槛，强化全球资源配置能力，构筑贸算融合的差异化竞争壁垒。

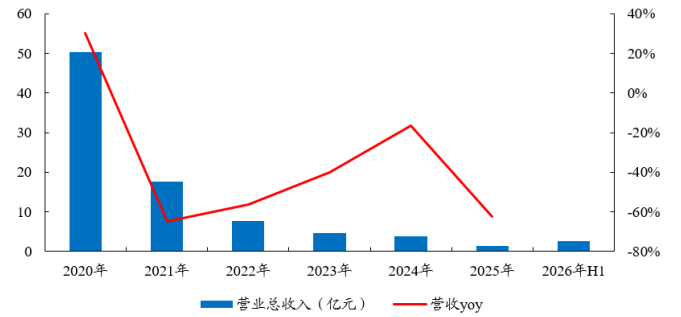
图20：行云科技布局全球跨境数字贸易、Token 工厂、液冷技术解决方案及行业 ASIC 芯片等核心业务



资料来源：行云科技官网

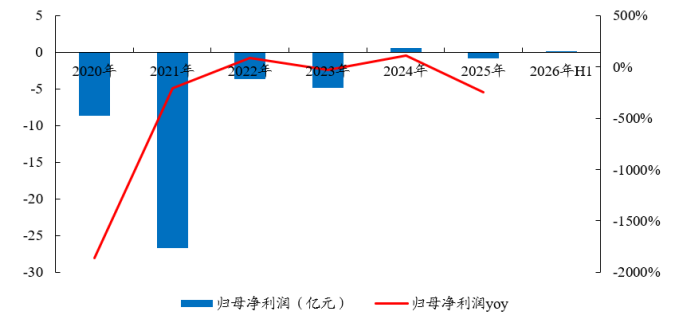
公司拓展 AI 算力服务业务，相关业务收入实现增长，对公司整体营收形成较大贡献。公司规模化落地 AI 算力运营服务，整合多型号异构 GPU 算力集群，提供算力租赁、弹性推理供给及全栈基础设施解决方案，下游覆盖大模型厂商、AI 科创企业、通信运营商等多元客户。采用 TaaS 计费模式，以算力 Token 为单位按需弹性计费，支持短期租赁、专属集群、私有化部署等合作形式，适配模型训练、多模态生成、行业智能推理等核心场景。2026 年上半年 AI 算力业务营收 6715.71 万元，占总营收 26.42%；公司总营收 2.54 亿元，归母净利润 0.12 亿元，分别同比增长 497.11%、540.15%，新旧业务协同增效，产业转型成效显著。

图21：2026 年 H1 行云科技营收同比增长 497.11%



数据来源：Wind、开源证券研究所

图22：2026 年 H1 行云科技归母净利润同比增长 540.15%



数据来源：Wind、开源证券研究所

仅供内部参考，请勿外传

4、 风险提示

互联网大厂资本开支不及预期风险。互联网头部厂商为国内算力租赁市场的核心需求主体，其 AI 领域资本开支节奏是驱动行业景气度的核心变量。若宏观环境波动或企业盈利预期发生调整，导致 AI 算力相关资本开支投放进度滞后，新增算力采购需求的释放节奏或有所放缓，行业需求增速存在阶段性承压的可能性。

行业竞争加剧风险。算力租赁行业景气度上行背景下，云厂商、IDC 运营商及新进入者持续加大供给布局，市场参与者数量稳步扩容。若后续行业算力供给集中落地，同质化竞争程度或有所提升，可能对市场价格体系及行业整体盈利水平形成阶段性扰动，市场份额分化进程或相应加快。

AI 模型发展不及预期风险。大模型技术迭代与参数规模扩张是训练侧算力需求增长的核心驱动力。若后续大模型技术突破节奏、落地进度低于市场预期，训练侧算力需求的增长弹性或存在一定不确定性，行业整体需求扩容的节奏可能相应放缓。

AI 推理需求不及预期风险。AI 推理算力是当前算力租赁行业重要的增量需求来源，其释放节奏高度依赖下游 AI 应用的商业化渗透进程。若下游应用落地速度、客户付费转化进度不及市场预期，推理侧算力需求放量节奏或有所延后，行业长期增长动能的释放进程存在一定波动可能。

仅供内部参考，请勿外传

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。		

仅供内部参考，请勿外传

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