

# 计算机行业跟踪周报

## Token 时代下算力租赁行业重构

增持（维持）

2026 年 04 月 13 日

证券分析师 王紫敬

执业证书：S0600521080005  
021-60199781

wangzj@dwzq.com.cn

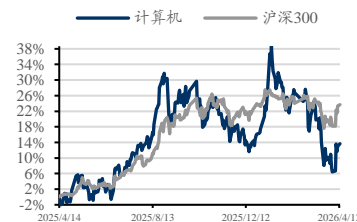
证券分析师 王世杰

执业证书：S0600523080004  
wangshijie@dwzq.com.cn

### 投资要点

- 算力租赁为客户提供以月、年为周期的 AI 计算资源租用服务。算力租赁公司在 AI 产业中具有重要地位，中国最前沿模型的训练与推理，均离不开算力租赁服务商所提供的先进算力。
- AI Agent、多模态、AI Coding、OpenClaw 等各类 AI 应用涌现，大幅拉动算力需求增长。根据 OpenRouter 的数据，2026 年元旦之后，全球头部模型的 Token 调用量大幅上升。
- 国内对算力租赁公司提供的高端算力需求更加旺盛。高端算力卡在训练与推理环节中的性价比更高。因此，能够提供稀缺高端算力服务的算力租赁厂商，将成为以 Token 为中心的新时代中，各家云厂商、大模型公司及应用企业竞相争夺的核心战略资源。
- 算力租赁呈现供不应求的态势。2026 年 Q1，头部云厂商及大模型公司纷纷加价采购算力租赁公司所持有的高端算力。仅有头部云厂商能够获得较为充足的高端算力，二线云厂商及大模型公司对高端算力的需求远未满足，缺口巨大。
- 算力租赁行业格局向头部集中。算力租赁行业长期存在三大壁垒，阻碍新供应商的进入。
- 算力租赁商业模式有望从“卖算力”转向“卖 Token”。当前算力紧缺程度较高，算力租赁厂商的行业议价权有望提升，其业务模式正从单纯的裸算力出租升级为模型服务或 Token 分成模式，即从“卖算力”转向“卖 Token”。这一转变有望大幅提升算力租赁公司的盈利能力，动其估值体系从 PE 向 PS 切换。
- 投资建议：云涨价周期刚刚开始，各大厂商陆续出现限购甚至涨价，随着 token 使用量提升，高端算力将愈发紧缺。但同时行业壁垒逐步提升，市场份额向头部供应商集中。
- 相关标的：宏景科技、协创数据、利通电子、盈峰环境等。
- 风险提示：技术发展不及预期；中美地缘政治风险；行业竞争加剧。

### 行业走势



### 相关研究

- 《如何理解十五五规划纲要中的数智化》  
2026-03-18
- 《政策加码算电协同，助力 token 出海》  
2026-03-17

## 内容目录

1. 算力租赁是中国尖端云的基础 .....	4
2. AI 应用加速落地拉动算力需求增加, 中国市场领先 .....	4
3. 算力租赁需求尤其旺盛 .....	5
4. 算力租赁商业模式升级 .....	6
5. 投资建议 .....	7
6. 风险提示 .....	7

## 图表目录

图 1: 2025-2026 全球头部 AI 模型调用量 .....	4
图 2: 2026 年 4 月 6 日 OpenRouter 大模型调用量排行 .....	5
图 3: 算力租赁厂商核心竞争力 .....	6

## 1. 算力租赁是中国尖端云的基础

算力租赁作为依托人工智能算力云服务构建的商业模式，为客户提供以月、年为周期的计算资源租用服务，服务范畴涵盖 CPU、GPU 及配套存储与内存等核心算力资源。在人工智能产业快速发展的背景下，算力租赁业务以 AI 芯片租赁为核心载体。

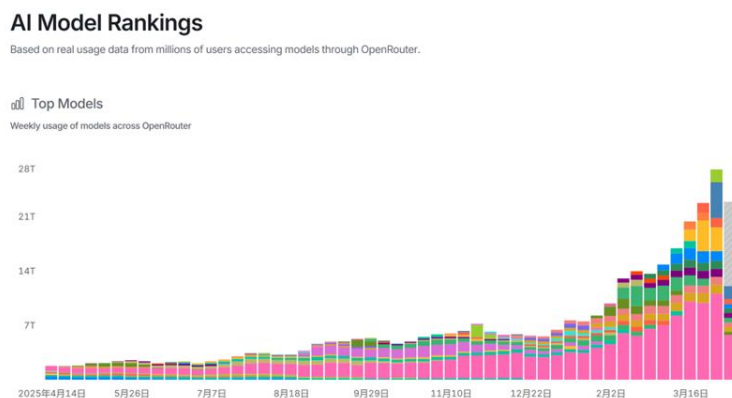
**算力租赁属于重资产商业模式。**算力租赁厂商对算力卡及服务器拥有所有权，并依据客户需求提供租赁服务，客户群体主要涵盖国内主流云服务提供商与头部大模型研发企业。相较于传统 IDC 厂商，算力租赁服务商通常不涉猎数据中心建设，其核心业务模式为算力服务器的租用。算力租赁公司在 AI 产业中具有重要地位，中国最前沿模型的训练与推理，均离不开算力租赁服务商所提供的先进算力。

## 2. AI 应用加速落地拉动算力需求增加，中国市场领先

AI Agent、多模态、AI Coding、OpenClaw 等各类 AI 应用涌现，大幅拉动算力需求增长。随着多模态等新型 AI 应用的落地与演进，Token 的消耗量将达到传统文本推理 AI 的数十倍甚至上百倍。以自动订票为例，完成一次操作即可产生 5,000 至 15,000 个 Token；而通过传统文本推理撰写一封邮件仅需约 300 个 Token，前者消耗量增长逾十倍。若加入多模态能力，Token 消耗将进一步扩大。例如，分析一张 4K 截图或一段 5 秒的视频，便需消耗 2,000 至 8,000 个 Token。可见，多模态生成式 AI 应用的普及对 Token 的消耗极为巨大。

根据 OpenRouter 的数据，2026 年元旦之后，全球头部模型的 Token 调用量大幅上升，这与新年以来 OpenClaw、Seedance、AI Coding 等 AI 应用的落地密切相关。

图1：2025-2026 全球头部 AI 模型调用量



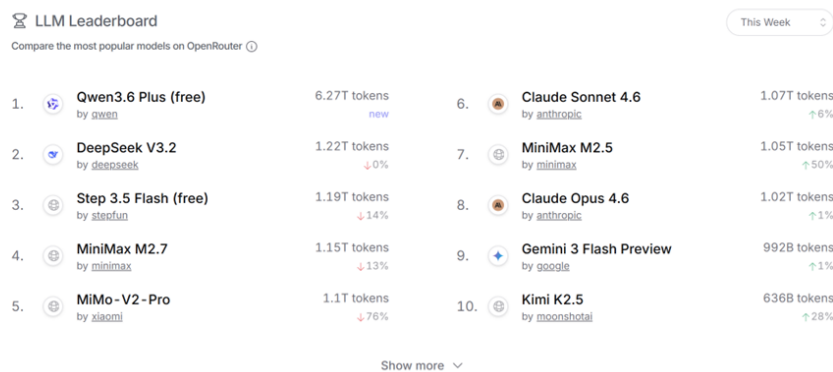
数据来源：OpenRouter，东吴证券研究所

**中国 Token 出海已成为明确产业趋势。**根据 OpenRouter 于 2026 年 2 月 16 日至 22 日发布的周度榜单，平台调用量排名前五的模型中，有四款来自中国厂商，分别为 MiniMax 的 M2.5、月之暗面的 Kimi K2.5、智谱的 GLM-5 以及 DeepSeek 的 V3.2。这四

款模型合计贡献了前五名模型总调用量的 85.7%。2026 年 2 月，中国 AI 模型调用量首次超越美国。这一局面的形成主要得益于两方面因素：一方面，中国 AI 模型经过多年发展，性能持续提升，已具备与海外模型竞争的實力；另一方面，中国 AI 模型依托国内相对低廉的电力及算力成本，Token 具备显著的成本优势。这使得中国在 OpenClaw 等各类 AI Agent 工具的底层调用中占据较大优势。性能红利驱动下，中国在本轮 AI 应用落地过程中的 Token 消耗量持续增长，甚至海外用户也开始调用中国模型的 Token。

**中国 Token 的成本水平取决于三个关键要素：模型性能的持续提升、优质的能源成本，以及逐步优化的算力成本。**前两个方面中国相比于海外都有显著优势，因此，中国 Token 的需求量增长有望显著超过全球平均增长水平。OpenRouter 截至 2026 年 4 月 6 日的数据显示，全球 AI 产业重心已明确向中国转移。中国大模型不仅连续五周领先于美国，且领先差距持续扩大，单周调用量增速达到美国的 41 倍，标志着中国 AI 产业已彻底告别技术追赶阶段，迈入应用领先的全新时期。

图2：2026 年 4 月 6 日 OpenRouter 大模型调用量排行



数据来源：OpenRouter，东吴证券研究所

### 3. 算力租赁需求尤其旺盛

国内对算力租赁公司提供的高端算力需求更加旺盛。高端算力卡在训练与推理环节中的性价比更高。因此，能够提供稀缺高端算力服务的算力租赁厂商，将成为以 Token 为中心的新时代中，各家云厂商、大模型公司及应用企业竞相争夺的核心战略资源。

**算力租赁呈现供不应求的态势。**2026 年 Q1，头部云厂商及大模型公司纷纷加价采购算力租赁公司所持有的高端算力。仅有头部云厂商能够获得较为充足的高端算力，二线云厂商及大模型公司对高端算力的需求远未满足，缺口巨大。

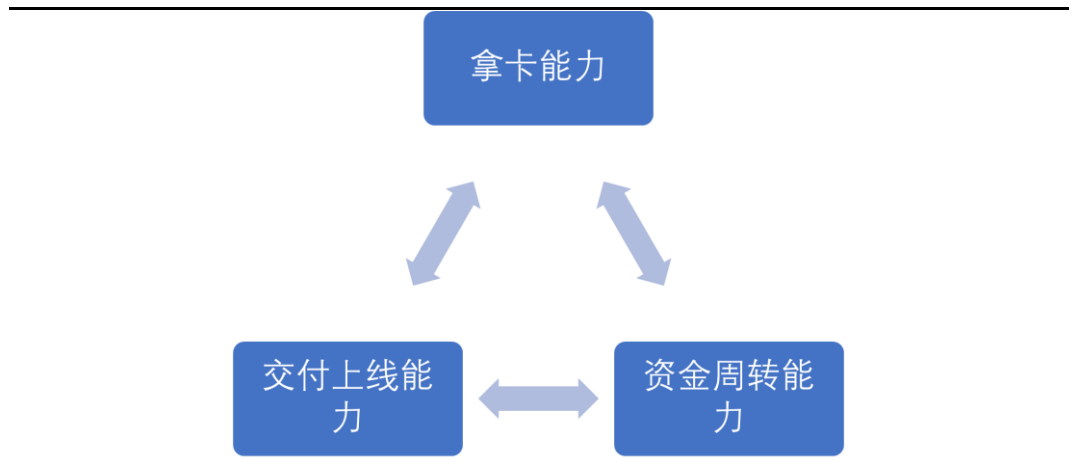
**算力租赁行业格局向头部集中：算力租赁行业长期存在三大壁垒，阻碍新供应商的进入。**

**稳定的拿卡能力。**互联网等客户对算力需求较大，单笔订单所需服务器数量较大，对供应商的供应能力有较强要求。如果不能按时交付，可能会有一定的处罚。

**强力的交付、上线和后续运营能力。**AI 算力服务器工作条件要求高，到货后需要调试、组网和调优，达到一定要求之后才能正常满足客户使用需求。同时，AI 算力服务器上线运营之后，需要专业团队维护和运营。较快的交付、上线时间，较少的损耗和故障情况，可以显著提升业务流畅性、降低运营成本，提升盈利能力。

**资金周转能力。**AI 算力服务订单金额较大，上市公司需要各种融资方式来完成订单要求。较强的资金周转能力决定公司能否按时完成客户的订单交付已经上线运营。

图3: 算力租赁厂商核心竞争力



数据来源：东吴证券研究所绘制

相较于 2025 年第一季度参与的部分厂商，到 2026 年第一季度，行业已出现明显的缩圈现象，参与者数量大幅减少。行业呈现出需求大幅扩张而供给向头部厂商收缩的态势，这种供紧需增的局面使得涨价成为必然趋势。

#### 4. 算力租赁商业模式升级

在传统商业模式下，算力租赁厂商获得订单后，通常签订三至五年的长期合同，验收合格后开始上线计费，按月向客户收取租金。自 2026 年第一季度起，算力租赁公司将逐季度兑现业绩，且每个季度均有望实现显著的同环比增长，订单不断累积并持续转化为收入。

**算力租赁商业模式有望从“卖算力”转向“卖 Token”。**当前算力紧缺程度较高，算力租赁商的行业议价权有望提升，其业务模式正从单纯的裸算力出租升级为模型服务或 Token 分成模式，即从“卖算力”转向“卖 Token”。这一转变有望大幅提升算力租赁公司的盈利能力，动其估值体系从 PE 向 PS 切换。

云厂商及模型公司未来的收入水平在很大程度上取决于其掌握的算力规模。模型能力再强，缺乏算力支撑便无法实现商业化变现；而算力租赁公司提供算力即帮助客户实现收入增长，未来有望形成 Token 分成商业模式，即客户每售出一定量的 Token，算力租赁公司即可按比例分得收益。

## 5. 投资建议

云涨价周期刚刚开始，各大厂商陆续出现限购甚至涨价，随着 token 使用量提升，高端算力将愈发紧缺。但同时行业壁垒逐步提升，市场份额向头部供应商集中。

相关标的：

宏景科技、协创数据、利通电子、盈峰环境等。

## 6. 风险提示

**技术发展不及预期。**AI 算力基建能否持续主要取决于 AI 大模型等技术的发展，如果 AI 技术发展不及预期，可能会影响算力基建投资速度。

**中美地缘政治风险。**美国可能加大对中国的科技制裁，导致中国无法获取最先进的 AI 芯片。

**行业竞争加剧。**算力基建市场发展迅速，新进参与者众多，可能会导致行业竞争加剧。

## 免责声明

东吴证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本研究报告仅供东吴证券股份有限公司（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。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，本公司及作者不对任何人因使用本报告中的内容所导致的任何后果负任何责任。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。

在法律许可的情况下，东吴证券及其所属关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供投资银行服务或其他服务。

市场有风险，投资需谨慎。本报告是基于本公司分析师认为可靠且已公开的信息，本公司力求但不保证这些信息的准确性和完整性，也不保证文中观点或陈述不会发生任何变更，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。

本报告的版权归本公司所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。经授权刊载、转发本报告或者摘要的，应当注明出处为东吴证券研究所，并注明本报告发布人和发布日期，提示使用本报告的风险，且不得对本报告进行有悖原意的引用、删节和修改。未经授权或未按要求刊载、转发本报告的，应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

## 东吴证券投资评级标准

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力相对基准表现的预期（A 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 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），北交所基准指数为北证 50 指数），具体如下：

公司投资评级：

- 买入：预期未来 6 个月个股涨跌幅相对基准在 15% 以上；
- 增持：预期未来 6 个月个股涨跌幅相对基准介于 5% 与 15% 之间；
- 中性：预期未来 6 个月个股涨跌幅相对基准介于 -5% 与 5% 之间；
- 减持：预期未来 6 个月个股涨跌幅相对基准介于 -15% 与 -5% 之间；
- 卖出：预期未来 6 个月个股涨跌幅相对基准在 -15% 以下。

行业投资评级：

- 增持：预期未来 6 个月内，行业指数相对强于基准 5% 以上；
- 中性：预期未来 6 个月内，行业指数相对基准 -5% 与 5%；
- 减持：预期未来 6 个月内，行业指数相对弱于基准 5% 以上。

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应当充分考虑自身特定状况，如具体投资目的、财务状况以及特定需求等，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。

东吴证券研究所  
苏州工业园区星阳街 5 号  
邮政编码：215021

传真：（0512）62938527

公司网址：<http://www.dwzq.com.cn>