

# 能源：AI的中国优势

证券分析师：袁理

执业证书编号：S0600511080001

联系邮箱：yuanl@dwzq.com.cn

二零二六年五月二十六日

## Tokens需求爆发，中国AI模型份额提升

- ✓ **Chatbot到AI Agent，Token调用量爆发。**Token（词元）是模型处理信息的最小计量单位。2026年以来Token调用量快速爆发，Token在AI运行中的角色发生系统性变化，Chatbot基本呈现为“一问一答”的形式，Tokens消耗基本跟用户数量线性相关；AI Agent时代，模型需要通过主动拆解任务、调用工具、搜索信息、循环推理等任务来完成工作，计算链条明显变长。**Token消耗开始呈现非线性规模扩张。**
- ✓ **中国模型调用占比提升。**在性能差距逐步收敛背景下，中国模型凭借更低推理成本、更高性价比与开源部署优势，市场份额开始领先。根据OpenRouter数据显示，2026年5月中国模型Token调用份额接近40%，持续提升。

## ✓算电协同，算力电力双向赋能

- ✓ **政府工作报告首提算电协同，AI产业加速发展，算力电力双向赋能产业趋势明确。**算电协同是以新型电力系统为支撑，深化智能调度、源网荷储、新型供备电等技术与机制创新，推动算力网与电力网深度融合、双向协同的新型运行模式。**电力支撑算力产业发展，算力反向赋能电力成长。**绿电直连政策频出，多方位支撑算电协同绿色发展。

## ✓中国的能源优势转化为AI的中国优势

- ✓ **芯片能效提升，度电生产转化Tokens量提升。**我们比较同样模型参数下，芯片能效逐步提升，单位Token耗电量逐步下降，每百万Token耗电量范围在0.36—1.1左右区间范围，均值在0.7度/百万Tokens水平，按照均值则1度电可生成输出1.67MTokens，技术进步电耗下降，**度电生成Tokens的量有50%以上的提升空间。**
- ✓ **电力—Tokens—应用，AI驱动电力能源价值全面重估。**1度电通过算力中心生成输出1.67MTokens（百万Tokens耗电0.6度），参考国内大模型企业定价，1M Tokens定价平均超过8元，**实现1度电成本（参考风电光伏）0.2元至输出Tokens价值13元+的价值跃升**，考虑大模型企业从输出Tokens到输出服务，Tokens价值提升的同时进一步带动电力价值提升。
- ✓ **单位Tokens成本拆分：**单位Tokens成本目前主要受到**服务器价格（单位算力服务器成本）、芯片效能（单位算力所需功率）、能源成本（电价）**的影响。经过我们测算，百万Tokens成本结构中设备折旧占比50%—75%，是最核心的成本构成；电力成本占比12%—14%，随着服务器价格下行、芯片效能提升，**电力成本的占比提升。**
- ✓ **能源优势有望成为AI竞争的胜负手。**AI下游需求爆发，上游设备制造技术逐年突破提升，芯片效能优化，服务器价格下行，单位Tokens成本中折旧成本逐年下降，能源等成本重要性逐年凸显。考虑能源供给的稳定（全国统一性电网，跨区调度能力全球领先），能源成本的竞争力（中国新增机组风光为主，边际成本低），能源优势有望成为AI竞争中中国的重要优势。

## ✓投资建议与风险提示

- ✓ **投资建议：**中国的能源优势转化为AI的中国优势，AI需求爆发，价值量膨胀驱动中国电力资产价值重估。**【电力板块】**AI驱动电力供需反转，电价企稳回升，电力公司积极拓展算电协同项目落地，建议关注**【绿电】【火电】**。垃圾焚烧分布式特点匹配算力中心需求，项目布局核心节点城市，综合考虑清洁性与稳定性，经济性有竞争力，建议关注**【垃圾焚烧】**。
- ✓ **风险提示：**电力需求增长不及预期、国内算力中心建设受阻、AI应用拓展不及预期等。



■ Tokens需求爆发，中国AI模型份额提升

---

■ 算电协同，算力电力双向赋能

---

■ 中国的能源优势转化为AI的中国优势

---

■ 投资建议与风险提示

---

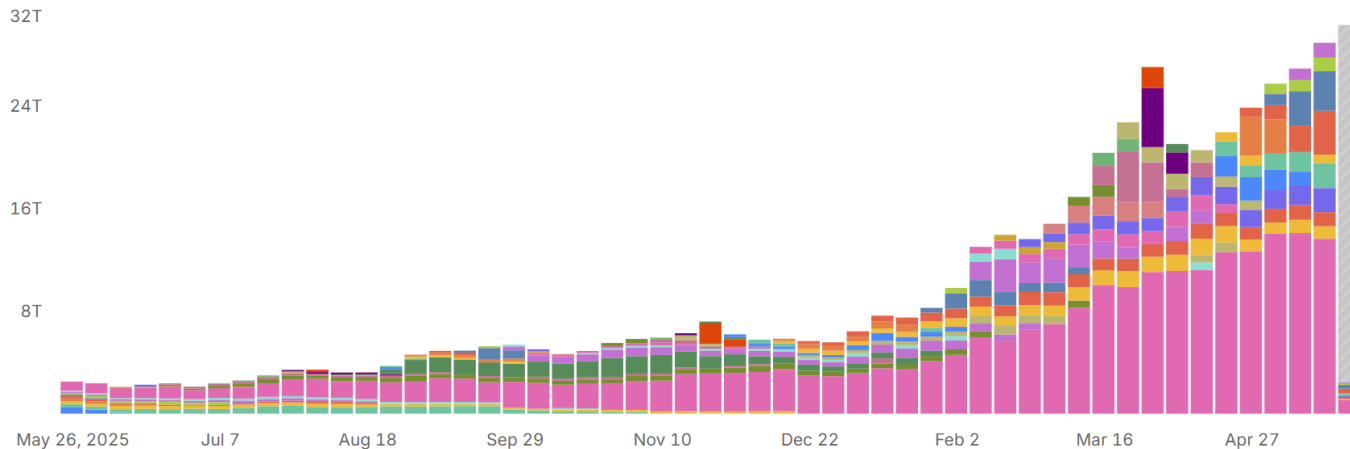
## Chatbot到AI Agent，Token调用量爆发

- ✓ **Token（词元）是模型处理信息的最小计量单位。**无论是用户的提问，还是AI生成的一段代码，最终都要被拆解成Token来完成运算。Token调用量成为衡量AI模型活跃度和产业价值的关键指标，Token调用量越高，意味模型被用得越多，创造的实际价值也就越大。
- ✓ **2026年以来Token调用量快速爆发。**2026年3月24日，国家数据局局长刘烈宏介绍，到2026年3月，我国日均Token（词元）调用量超过140万亿，相比2024年初的1000亿增长了1000多倍，相比2025年底的100万亿，三个月时间又增长了40%多。从2025年3月到2026年3月，OpenRouter平台的总Token使用量呈现爆发式增长，从约2T左右快速提升至接近30T，2025年底到2026年初增速明显加快。
- ✓ **Token增长逻辑正由“用户规模驱动”转向“任务复杂度驱动”。**随着交互模式从Chatbot到AI Agent转型，多Agent协同、长上下文等形式持续演进，Token消耗开始呈现非线性扩张特征。

Top Models

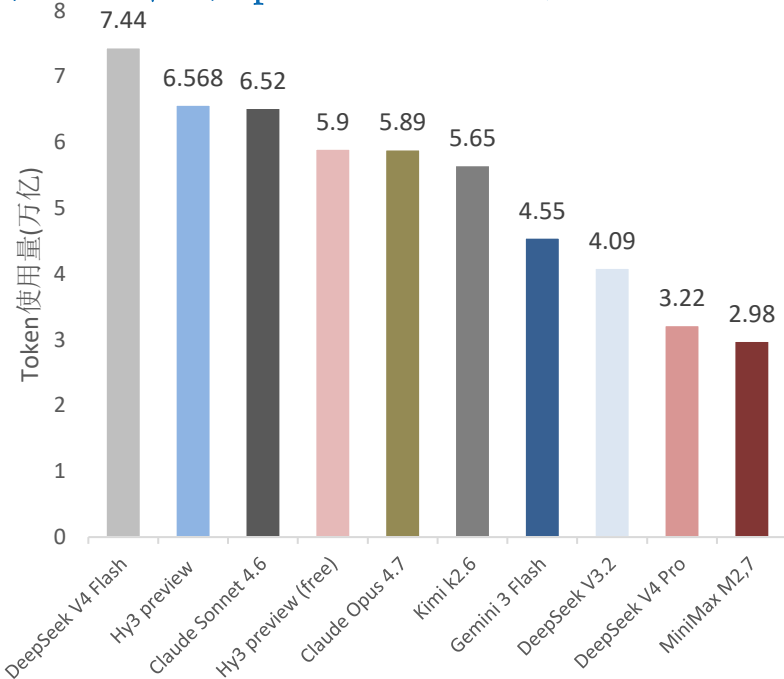
图：OpenRouter模型周使用量（单位：T Tokens）

Weekly usage of models across OpenRouter

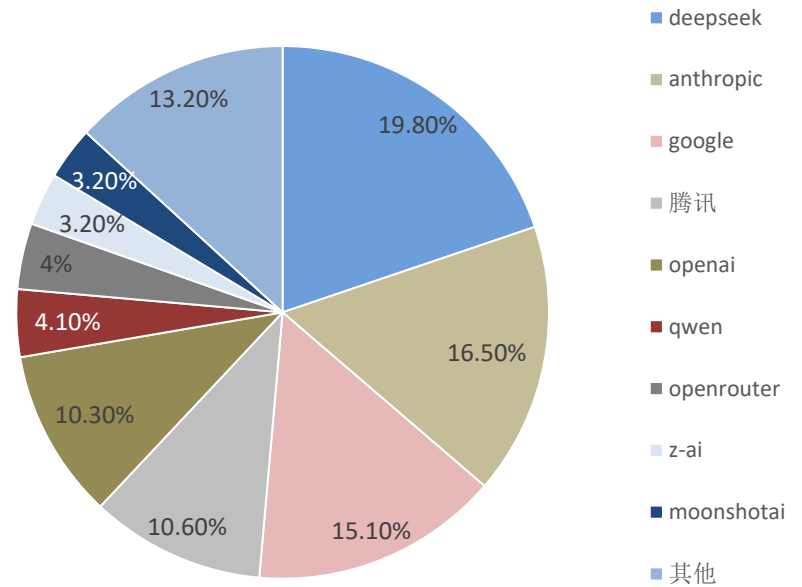


- ✓ 中国模型市场份额接近40%。根据OpenRouter数据显示，2026年5月中国模型Token调用量已进入全球第一梯队，中国模型整体市场份额接近40%。
- ✓ 2026年5月OpenRouter平台主要模型的Token调用量排名显示海外头部模型仍然占据主导地位，但中国模型的增长速度已经非常明显。其中，Deepseek V4 Flash以约7.44万亿Token的月调用量位居第一，超过Claude Sonnet 4.6和Claude Opus 4.7，从海外厂商来看，OpenAI (10.3%)、Anthropic(16.5%)与Google(15.1%)三者合计市场份额41.9%，仍然构成全球大模型市场的“海外御三家”。从国内厂商来看，中国模型整体市场份额已经接近40%。其中，DeepSeek单独占比达到19.8%，已经成为OpenRouter平台最大的单一模型来源之一；腾讯模型占比约10.6%，Moonshot(Kimi)约3.2%。

图：2026年5月OpenRouter模型使用量梳理



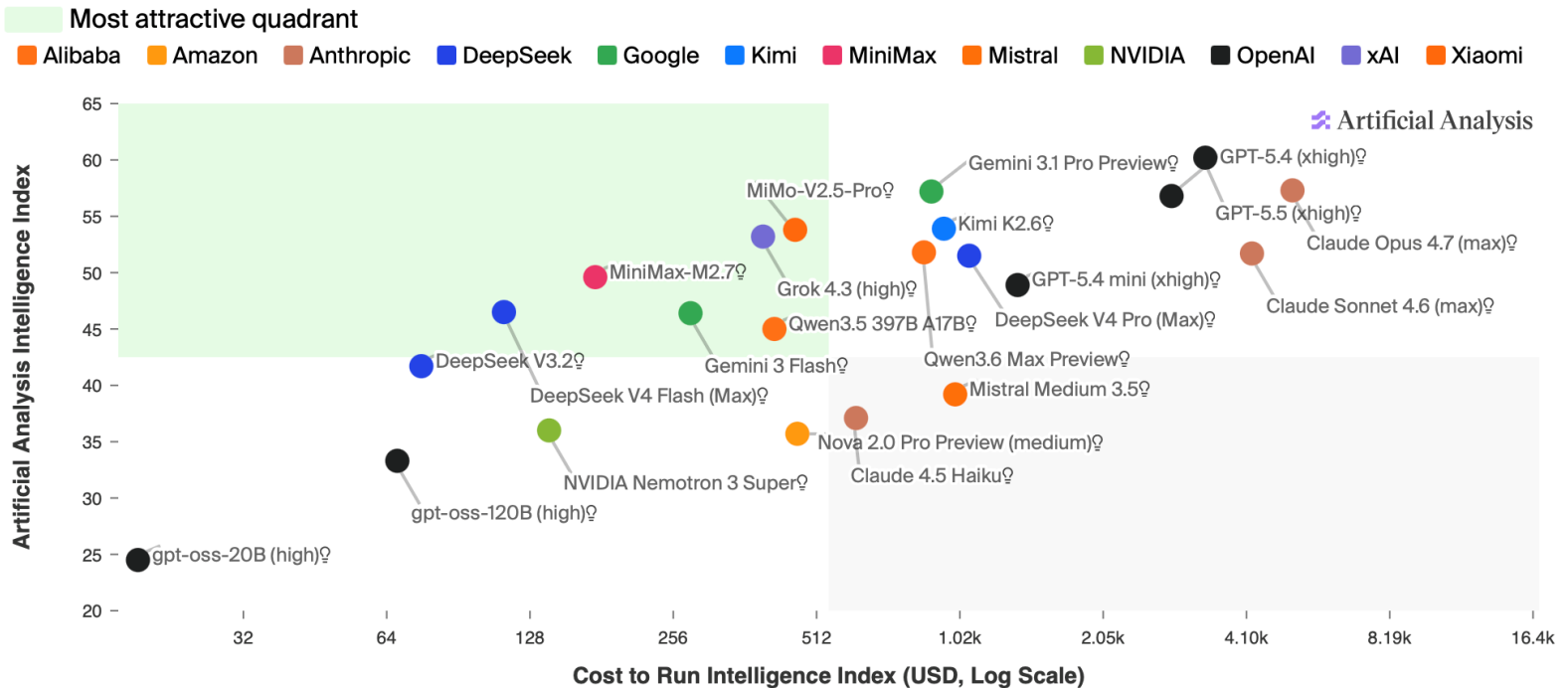
图：2026年5月OpenRouter市场份额占比



## 国产大模型优势加强，份额有望持续提升

- ✓ 在性能差距逐步收敛背景下，中国模型凭借更低推理成本、更高性价比与开源部署优势，市场份额开始领先。随着AI推理需求扩张，中国模型有望进一步提升全球Token调用中的占比。
- ✓ 中国模型大模型性价比突出，来自 1) **电力成本优势**：中国在工业电价、电网设施、数据中心配套等方面具备综合优势，“东数西算”与西部绿电资源推动算力向低成本能源区域集聚。2) **开源生态**：DeepSeek、Qwen等开源模型降低了企业部署与Agent化门槛，企业仅需微调模型，AI更容易嵌入真实业务流程。3) **注重商业化**：不追求极致性能领先，而是倾向于大规模可落地，不断追求性价比。

图：模型成本与智能程度分布，中国大模型为性价比之选（截至2026/5/26）





■ Tokens需求爆发，中国AI模型份额提升

---

■ 算电协同，算力电力双向赋能

---

■ 中国的能源优势转化为AI的中国优势

---

■ 投资建议与风险提示

---

## 中国算力用电占比接近2%，逐步成为用电需求的重要增量

- ✓ 中国算力用电增速持续高于全社会用电增速。2019 - 2025年，中国算力用电由931亿千瓦时提升至1933亿千瓦时，CAGR显著高于同期全社会用电CAGR，算力用电占比由1.3%提升至1.9%。2025年算力相关用电量同比增长17%，明显高于全社会用电5.2%的增速，反映出AI训练、推理及数据中心扩张正成为新增电力需求的重要来源。
- ✓ 中国算力用电逐步成为全社会用电需求的重要增量。根据中国信通院的预测，到2030年，在低/中/高情景下，中国算力相关用电量预计分别达到3000/4000/7000亿千瓦时，占全社会用电比例有望提升至2.3%/3.1%/5.3%。高情景下，算力用电增量对全社会用电增长贡献接近1个百分点，意味着AI产业链已从“边际负荷”逐步演变为影响电力需求结构的重要变量。
- ✓ 2026年3月，政府工作报告首提算电协同，AI产业加速发展，算力电力双向赋能产业趋势明确。

表: 算力用电贡献测算

年份	2019	2020	2021	2022	2023	2024	2025	2030E		
全社会用电量 (亿千瓦时)	72852	75110	83128	86372	92241	98521	103682	131072		
YOY	6.4%	3.1%	10.7%	3.9%	6.8%	6.8%	5.2%	4.80%		
算力相关用电量 (亿千瓦时)	931	1148	1314	1344	1473	1653	1933	低情景 3000	中情景 4000	高情景 7000
YOY	16.7%	23.3%	14.5%	2.3%	9.6%	12.2%	17.0%	9%	16%	29%
算力相关用电量占比	1.3%	1.5%	1.6%	1.6%	1.6%	1.7%	1.9%	2.3%	3.1%	5.3%
算力用电增长对于全社会用电增长的贡献	0.19%	0.30%	0.22%	0.04%	0.15%	0.19%	0.28%	0.20%	0.40%	0.96%

注：2030年假设中，全社会用电量YOY，低、中、高情景算力相关用电量YOY均为2026-2030年的CAGR

## 算电政策频出，基础设施高速发展

✓ 政策由点到面持续升位，加速落地节奏明显。2026年4月发改委、能源局、工信部、数据局四部门联合印发《关于促进人工智能与能源双向赋能的行动方案》，明确到2027年初步构建安全、绿色、经济的能源保障体系。算电协同已完成从地方试点到国家战略的整体跃升。

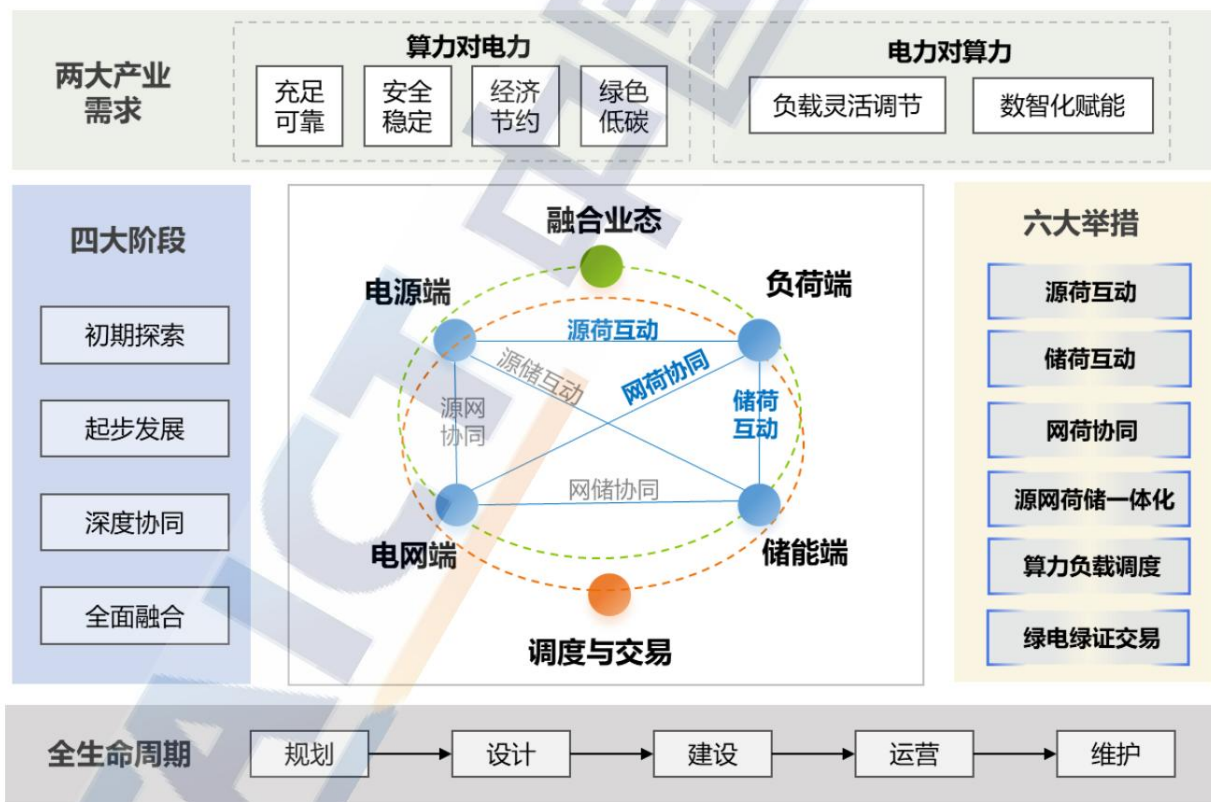
表: 算电协同相关政策梳理

日期	政策	典型案例
2026年3月	《中华人民共和国国民经济和社会发展第十五个五年规划纲要》	在国家最高层级规划中，提出推动绿色电力与算力协同布局，将"算电协同"上升为国家顶层设计
2026年3月	《政府工作报告》	首次将"算电协同"纳入新基建工程，确立其国家战略地位
2024年12月	《国家数据基础设施建设指引》	系统提出推进算力与绿色电力融合，鼓励数据中心探索绿电直供、聚合供应等新模式
2024年10月	《关于大力实施可再生能源替代行动的指导意见》	强调支持新型基础设施发展绿电直供等模式，提高"东数西算"工程中的新能源电力消费占比。
2024年7月	《加快构建新型电力系统行动方案（2024—2027年）》	响应并部署实施一批算力与电力协同项目，明确将算电协同作为重点任务持续推进。
2024年7月	《数据中心绿色低碳发展专项行动计划》	再次强化"到2025年底算力电力双向协同机制初步形成，国家枢纽节点新建数据中心绿电占比超过80%"的目标。
2023年12月	《关于深入实施“东数西算”工程加快构建全国一体化算力网的实施意见》	首次在国家层面明确"创新算力电力协同机制"，提出具体的试点路径与量化目标

## 算电协同，构建数字经济与绿色能源新范式

- ✓ 算电协同护航算力高质量发展。算力电力协同（简称“算电协同”）是以新型电力系统为支撑，深化智能调度、源网荷储、新型供备电等技术与机制创新，推动算力网与电力网深度融合、双向协同的新型运行模式。它成功打破传统数据中心单纯作为“用电大户”的被动属性，通过重塑，使其转化为电网的“柔性调节资源”，成为破解算力高速扩张与电力稳定供应矛盾的关键路径。

图：算力电力协同发展体系



## 调度机制以源荷互动为核心，三种商业模式打通数字与能源价值闭环

- ✓ 算电协同不仅是政策倡导的绿色转型的硬性要求，同时是重塑数字与能源行业价值链的必选项。1) 绿电物理直供是算电协同的基础商业模式，通过专线实现绿电的点对点供给，适用于资源富集区的低成本用能场景。2) 源网荷储一体化在直供基础上引入储能与智能调度，形成区域内发输储用的自平衡体系，提升了供电稳定性与市场参与能力。3) 虚拟电厂不依赖自建电源，依托算力负荷的柔性调节与平台聚合，以轻资产方式参与电力市场交易。

表: 算电协同三种商业模式梳理

商业模式	模式特点	项目适用	典型案例
<p><b>绿电物理直供</b>                      绕开公共电网，专线风光直供算力中心</p>	<ul style="list-style-type: none"> <li>✓ 建设成本相对较低</li> <li>✓ 绿电占比高、碳排低</li> <li>✓ 时序波动（供电受天气影响，不稳定）</li> </ul>	西部“东数西算”枢纽中，对时延不敏感的冷数据存储、离线计算、AI大模型非实时训练等项目。	宁夏中卫源网荷储光伏项目
<p><b>源网荷储一体化</b>                      在物理直供基础上增加储能与智能调度，形成发-输-储-用自平衡体系</p>	<ul style="list-style-type: none"> <li>✓ 初期建设成本高，长期运营成本低</li> <li>✓ 供电稳定，可平滑新能源波动</li> <li>✓ 可参与电力市场（峰谷价差套利、辅助服务）</li> </ul>	西部或中部风光资源丰富的新建大型/超大型数据中心，追求100%绿电、长期稳定运营。	内蒙古乌兰察布·中金数据零碳算力基地 全国首个数据中心源网荷储一体化项目
<p><b>虚拟电厂直供</b>                      依托智算中心备用储能与柔性负荷，以平台聚合参与电力市场</p>	<ul style="list-style-type: none"> <li>✓ 低谷蓄能、高峰让路（削峰填谷）</li> <li>✓ 无需自建电厂，投入较低</li> <li>✓ 需要高精度负荷预测与快速响应算法</li> </ul>	东部沿海及高电价、高负荷地区的存量商用智算中心；跨区域算力调度项目。	上海临港“沪闽算力集群”项目 国内首次跨省百卡算力自动化迁移验证

## 绿电直连有效支撑算力协同推进

- ✓ 聚焦算力领域，国家枢纽节点新建数据中心“绿电占比确保超80%”已成为项目准入的硬性红线。国家层面密集出台绿电直连配套政策，通过放开多用户限制、规范价格结算、优化配电网架构进行体制与物理层面的层层松绑，为算力项目获取长期、低成本的源头绿电扫清了障碍。
- ✓ 绿电直连指算力设施与新能源电站专线直连、绕过公网输送环节，直连单一用户。其本质为专线供电、不经公网、碳排放核算清晰，契合国际规则、减少绿电来源争议、提升绿色竞争力。

图: 绿电直连 VS 集中式绿电

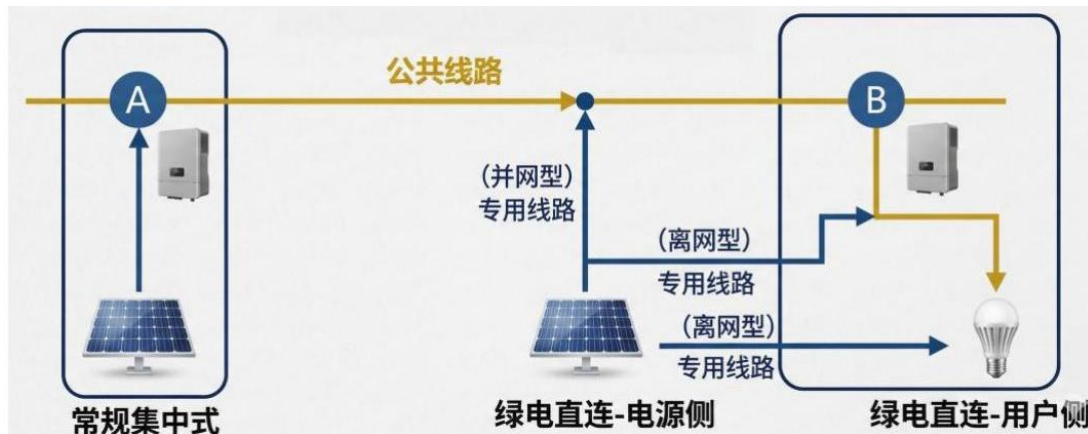


表: 绿电直连并网型 VS 离网型

直连模式	模式含义	适用项目	关键原则
离网型	项目作为整体接入公共电网，电源接入用户侧	<ul style="list-style-type: none"> <li>✓ 有降碳刚性需求的出口外向型企业</li> </ul>	<ul style="list-style-type: none"> <li>✓ 新能源年自发自用电量<math>\geq</math>总可用发电量的60%</li> </ul>
并网型	项目的电源、用户和线路均与公共电网没有电气连接，完全作为一个独立的微电网系统运营。	<ul style="list-style-type: none"> <li>✓ 新增负荷项目</li> <li>✓ 拥有燃煤燃气自备电厂的存量负荷企业</li> </ul>	<ul style="list-style-type: none"> <li>✓ 新能源年自发自用电量<math>\geq</math>总用电量的30%</li> <li>✓ 上网电量<math>\leq</math>总可用发电量的20%，限制“余电上网”的规模</li> </ul>

## 绿电直连走向规范化，构筑算电协同制度底座

- ✓ 中办、国办印发的《碳达峰碳中和综合评价考核办法》正式挂钩省级党委政府履职考核，碳排放总量、强度及非化石能源占比成为刚性指标，消纳绿色能源已成为地方平衡经济增长与能耗指标的战略核心。
- ✓ 聚焦算力领域，国家枢纽节点新建数据中心“绿电占比确保超80%”已成为项目准入的硬性红线。国家层面密集出台绿电直连配套政策，通过放开多用户限制、规范价格结算、优化配电网架构进行体制与物理层面的层层松绑，为算力项目获取长期、低成本的源头绿电扫清了障碍。

表: 绿电直连政策梳理

政策名称	日期	内容
《关于有序推动多用户绿电直连发展有关事项的通知》(发改能源〔2026〕688号)	2026.05	绿电直连供电对象首次由单用户扩展至多个不同法人实体（不含居民和农业）。支持增量配电网、工业/零碳园区、外向型上下游企业及算力设施、绿氢氨醇等新产业开展。
《关于促进电网高质量发展的指导意见》(发改能源〔2025〕1710号)	2025.12	目标到2030年初步建成“大电网为主、配电网互动、智能微电网补充”的三层协同平台，接纳分布式新能源能力达到9亿千瓦，支撑充电基础设施超4000万台。
《关于完善价格机制促进新能源发电就近消纳的通知》(发改价格〔2025〕1192号)	2025.09	完善就近消纳项目稳定供应保障费用计算。
《关于有序推动绿电直连发展有关事项的通知》(发改能源〔2025〕650号)	2025.05	首次明确绿电直连含义和类型，提出绿电自发自用比例要求（占总发电量 $\geq 60\%$ ，占总用电量 $\geq 30\%$ ）



■ Tokens需求爆发，中国AI模型份额提升

---

■ 算电协同，算力电力双向赋能

---

■ 中国的能源优势转化为AI的中国优势

---

■ 投资建议与风险提示

---

## Tokens调用量 VS 用电量，AI激发用电潜力，电力支撑AI发展

- ✓ 我们构建Tokens调用量与用电量模型。由算力需求驱动的用电量需求，由IT负载用电与能量使用效率（PUE）共同决定。IT负载用电来自计算过程的能耗消纳，可以进一步拆解为总计算量（FLOPs）与单位计算能耗的乘积。

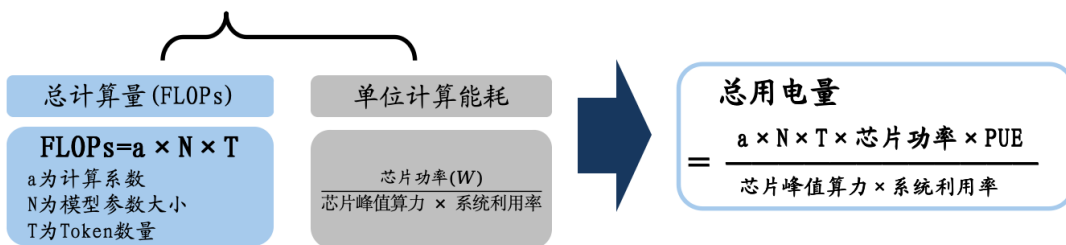
**总计算量：**采用以Transformer为基础的标准计算框架，总计算量近似为Tokens调用量、计算系数、模型参数量的乘积。**计算系数随推理复杂度提升而提升。**推理阶段，计算系数近似为2，在训练阶段由于需要叠加反向传播与参数更新，计算系数近似为6，实际应用中模型并未仅仅包含推理、训练两种简单划分，是多种计算模式的叠加，且呈现出随任务复杂度非线性增长的特征。**模型参数量增长但有效收益边际递减。**模型性能在一定范围内随参数规模提升而持续改善，推动参数规模不断增长，随着MoE架构、蒸馏、小模型优化以及推理成本约束导致行业正从单纯扩大参数规模向效率优化与结构创新转变。

**单位计算能耗：**不同芯片型号单位功耗下提供的计算能力存在差异，每瓦算力（FLOPS/W）参数随芯片性能提升与系统优化逐步提升，推动完成计算任务所需要时间以及总能耗的下降。

**PUE：**数据中心总能耗与IT设备能耗之比，国家发改委对于新建大型数据中心PUE普遍要求不高于1.3。

- ✓ 随着AI走向高频生产工具，推理调用量与Agent workflow快速增长，Tokens调用量带动总计算量持续提升，芯片能效与系统效率持续提升，有望推动单位Token的成本（包括用电量等）下降，进一步激发算力需求与能源需求增长。

$$\text{总用电量} = \text{IT用电} \times \text{能量使用效率 (PUE)}$$



## 芯片能效提升，度电生产转化Tokens量提升

- ✓ 芯片能效提升，度电生产转化Tokens量提升。我们比较同样模型参数下，芯片能效逐步提升，单位Token耗电量逐步下降，每百万Token耗电量范围在0.36—1.1左右区间范围，均值在0.7度/百万Tokens水平，按照均值则1度电生成输出1.67MTokens，技术进步电耗下降，度电生成Tokens的量有50%以上的提升空间。

图：不同类型芯片单位Token耗电量对比

芯片品牌	英伟达	英伟达	AMD	华为	海光	寒武纪	昆仑芯
芯片型号	H100SXM	A100PCle	MI325X	昇腾310P	BW100	MLU590	P800
TFLOPS	989	312	1300	70	400	345	345
功率(W)	700	300	1000	72	350	350	300
假设模型参数为500B	1.780	0.562	2.340	0.126	0.720	0.621	0.621
<b>单位token电耗(度/百万Tokens)</b>	<b>0.590</b>	<b>0.801</b>	<b>0.641</b>	<b>0.857</b>	<b>0.729</b>	<b>0.845</b>	<b>0.966</b>
假设模型参数为8B	111.263	35.100	146.250	7.875	45.000	38.813	38.813
<b>单位token电耗(度/百万Tokens)</b>	<b>0.009</b>	<b>0.013</b>	<b>0.010</b>	<b>0.014</b>	<b>0.012</b>	<b>0.014</b>	<b>0.015</b>

## 电力—Tokens—应用，AI驱动电力能源价值全面重估

- ✓ 除了转化Tokens量的提升之外，大模型、应用赋能Tokens价值量再提升。我们选取单位Token电耗均值，即度电通过算力中心生成输出1.67MTokens（百万Tokens耗电0.6度），参考国内大模型企业定价，1M Tokens定价平均超过8元，实现1度电成本（参考风电光伏）0.2元至输出Tokens价值13元+的价值跃升，考虑大模型企业从输出Tokens到输出服务，Tokens价值提升的同时进一步带动电力价值提升。

表：主流模型Tokens价格梳理

公司	模型名称	输入价格(元/M)	输出价格(元/M)	Chatbot综合价格(元/M)	Agent综合价格(元/M)
OpenAI	gpt-5.5(短文本)	35.0	210.0	78.8	52.5
	gpt-5.5-pro(短文本)	210.0	1260.0	472.5	315.0
	gpt-5.5-pro(长文本)	420.0	1890.0	787.5	567.0
Anthropic	Haiku4.5	7.0	35.0	14.0	9.8
	Sonnet4.6	21.0	105.0	42.0	29.4
	Opus4.7	35.0	175.0	70.0	49.0
Google	Gemini 3.5 Flash	10.5	63.0	23.6	15.8
	Gemini 3.1 Flash-Lite	1.8	10.5	3.9	2.6
	Gemini 3 Pro预览版	21.0	105.0	42.0	29.4
DeepSeek	DeepSeek-V4-Flash(缓存命中)	0.02	2.0	0.5	0.2
	DeepSeek-V4-Flash(缓存未命中)	1.0	2.0	1.3	1.1
	DeepSeek-V4-Pro(缓存未命中)	3.0	6.0	3.8	3.3
Qwen	qwen3.7-max	12.0	36.0	18.0	14.4
	qwen3.6-max-preview (0<Token≤128K)	9.0	54.0	20.3	13.5
	qwen3.6-max-preview (128K<Token≤256K)	15.0	90.0	33.8	22.5
GLM	GLM-5.1 输入长度 (0<Token≤32K)	6.0	24.0	10.5	7.8
	GLM-5.1 输入长度 (Token≤32K)	8.0	28.0	13.0	10.0
	GLM-5-Turbo 输入长度 (0<Token≤32K)	5.0	22.0	9.3	6.7
Kimi	k2.6(缓存命中)	1.1	27.0	7.6	3.7
	k2.6(缓存未命中)	6.5	27.0	11.6	8.6
	moonshot-v1-8k	2.0	10.0	4.0	2.8
	moonshot-v1-32k	5.0	20.0	8.8	6.5

## 设备折旧是算力中心最核心成本构成，能源成本重要性提升

- ✓ 我们搭建模型对Tokens成本进行测算，测算以DeepSeek V3 (671B 参数)推理场景为基准，横向对比 NVIDIA H100 SXM、H100中国特供版 H20、华为昇腾910B三款主流芯片：经过我们测算，百万Tokens对应成本目前主要受到服务器价格（单位算力服务器成本）、芯片效能（单位算力所需功率）、能源成本（电价）的影响。成本结构来看：百万Tokens成本中设备折旧占比50% - 75%，是最核心的成本构成；电力成本占比12% - 14%，随着服务器价格下行、芯片效能提升，电力成本的占比提升。

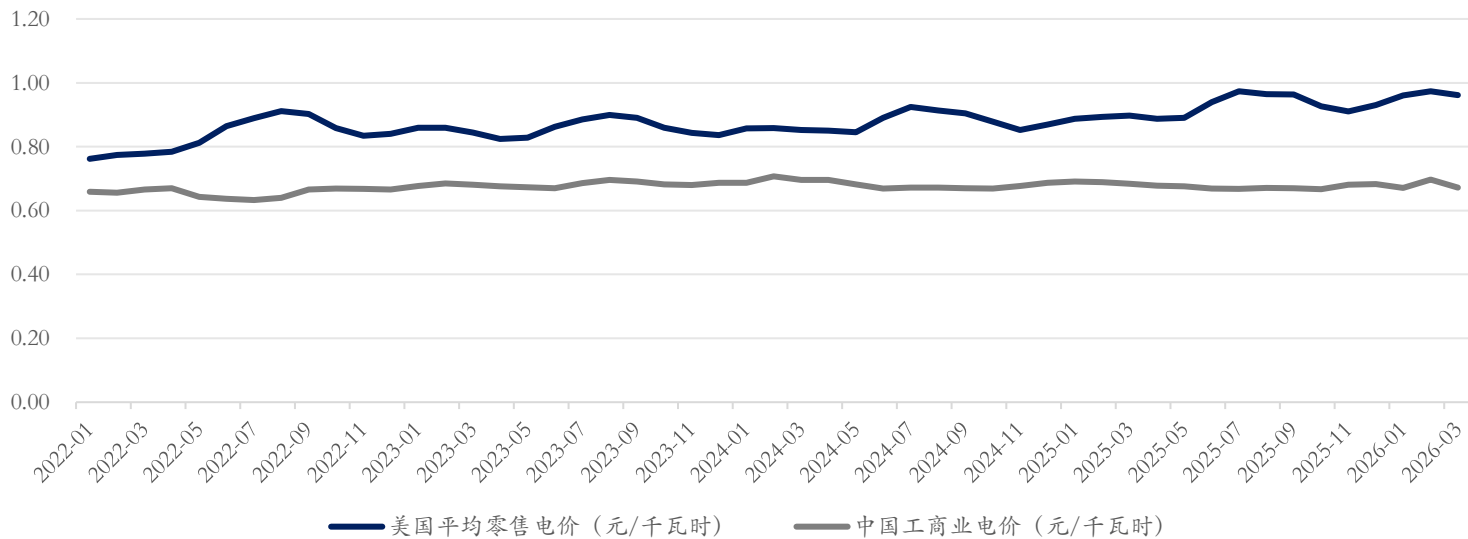
表: 数据中心百万Tokens成本测算

	H100SXM	H20	华为910B		H100SXM	H20	华为910B
<b>芯片基本规格参数</b>				<b>模型与推理效率假设</b>			
理论峰值算力(FP16/BF16) (TFLOPS)	1,979	148	320	目标模型参数量(十亿)	671.00	671.00	671.00
显存容量(GB)	80	96	64	MFU (模型算力利用率)	0.40	0.35	0.25
芯片热设计功耗 (TDP) (W)	700	400	400	每Token计算量系数	2.00	2.00	2.00
单卡采购价格 (估算)	300,000	100,000	120,000	<b>单卡日持有成本 (元/卡/天)</b>			
整机服务器卡数(卡/台)	8	8	8	整机单卡采购成本 (元)	60,000	20,000	25,500
整机成本/芯片价格倍数(倍)	1.60	1.60	1.70	日折旧费用 (芯片+基建) (元)	62.47	20.82	25.77
<b>性价比 (元/TFLOPS)</b>	<b>151.59</b>	<b>675.68</b>	<b>375.00</b>	日电费 (元)	14.11	8.06	8.06
<b>能效 (W/TFLOPS)</b>	<b>0.35</b>	<b>2.70</b>	<b>1.25</b>	日运维/人力摊销 (元)	21.92	21.92	32.88
<b>运营与部署假设</b>				单卡日总成本 (元)	98.50	50.80	66.71
芯片折旧年限(年)	5.00	5.00	5.00	<b>单卡日产出Token量</b>			
残值率	0.05	0.05	0.05	有效算力(TFLOPS)	791.60	51.80	80.00
数据中心PUE	1.40	1.40	1.40	每Token消耗算力(TFLOPs)	1.34	1.34	1.34
<b>工业电价(元/度)</b>	<b>0.60</b>	<b>0.60</b>	<b>0.60</b>	理论每秒生成Token	589.87	38.60	59.61
运维及人力摊销(元/卡/年)	8,000.00	8,000.00	12,000.00	日产出Token (百万Token/卡/天)	40.77	2.67	4.12
其他基建投入占芯片裸卡比例	60.0%	60.0%	60.0%	<b>单位Token成本 (元/百万Token)</b>	<b>2.42</b>	<b>19.04</b>	<b>16.19</b>
<b>其他基建折旧年限(年)</b>	<b>15.00</b>	<b>15.00</b>	<b>15.00</b>	折旧成本	1.53	7.80	6.25
GPU有效利用率	0.80	0.80	0.80	电力成本	0.35	3.02	1.96
				运维成本	0.54	8.22	7.98

## 中国电力与IDC成本优势，有望成为国产推理集群的重要竞争壁垒

- ✓ AI推理本质属于高功耗、高负载、持续运行型业务，电力成本在推理集群运营中的重要性持续提升。相比美国，中国在工业电价、IDC建设成本、电网稳定性以及新能源配套方面具备一定优势。
- ✓ 美国高端AI数据中心主要集中于德州、弗吉尼亚等区域，工业电价整体较高，叠加AI负载快速增长，局部区域已出现电力紧张与电价上涨趋势。而中国西部地区由于绿电资源丰富、水电及风光装机占比较高，具备更低的长期电力成本，为大规模推理集群部署提供基础。
- ✓ 随着推理需求持续扩张，AI产业竞争将逐步从“芯片竞争”延伸至“能源竞争”，低成本电力资源的重要性将持续提升。

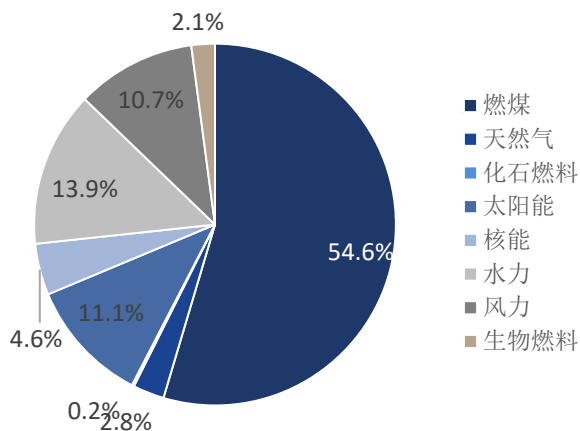
图：中国工商业电价长期稳定低位运行



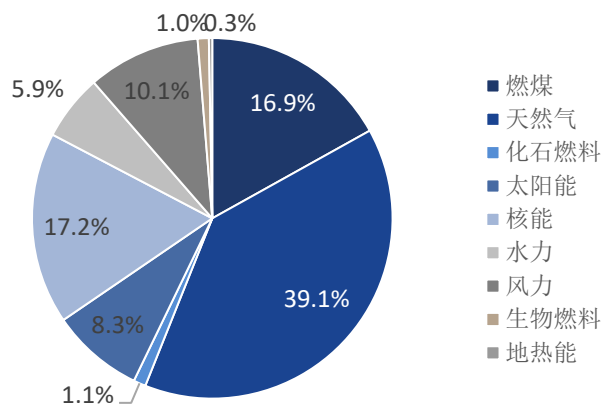
## 中国电力与IDC成本优势，有望成为国产推理集群的重要竞争壁垒

- ✓ **电源：**中美电力结构差异，本质上源于资源禀赋不同：美国依托页岩气革命形成以天然气为主的发电体系，中国则在“双碳”目标推动下持续提升风光水核等新能源占比。展望未来，美国新增电源仍以气电为主，电价受天然气价格波动影响较大；中国新能源装机进入快速扩张阶段，“东数西算”持续推进，西部低成本绿电资源优势逐步强化。
- ✓ **电网：**AI对电网稳定性、跨区域调度与供电冗余能力要求显著提升。近期美国频繁讨论“AI缺电”，本质上反映的是美国电网老化、区域割裂与输配能力不足的问题。美国电网以区域化运营为主，缺乏全国统一调度体系，州际输电与跨区域协同能力较弱。叠加新能源建设与电网投资长期错配，AI数据中心快速扩张后，局部地区出现供电瓶颈、电价波动与排队接入现象。相比之下，中国形成了以国家电网、南方电网为核心的全国性统一电网体系，特高压、“西电东送”“东数西算”等工程持续推进，跨区域调配能力全球领先。中国在统一电网体系、特高压建设及能源调度能力方面具备长期优势，有望支撑国产AI集群持续扩张。

图：中国电力结构（2025年）



图：美国电力结构（2025年）





■ Tokens需求爆发，中国AI模型份额提升

---

■ 算电协同，算力电力双向赋能

---

■ 中国的能源优势转化为AI的中国优势

---

■ 投资建议与风险提示

---

## 投资建议与风险提示

### ✓ 投资建议:

中国的能源优势转化为AI的中国优势，AI需求爆发，价值量膨胀驱动中国电力资产价值重估。【**电力板块**】AI驱动电力供需反转，电价企稳回升，电力公司积极拓展算电协同项目落地，建议关注【**绿电**】【**火电**】。垃圾焚烧分布式特点匹配算力中心需求，项目布局核心节点城市，综合考虑清洁性与稳定性，经济性有竞争力，建议关注【**垃圾焚烧**】。

### ✓ 风险提示:

**电力需求增长不及预期:** AI带来新增电力需求，但目前占比相对较小，如果电力需求整体偏弱，会影响到电力行业相关企业的电量与电价表现；

**国内算力中心建设受阻:** 算力中心建设受到政策、资源匹配等多种因素影响，如果算力中心建设节奏低于预期，会影响到AI产业发展，进一步影响AI需求爆发对能源电力需求的拉动；

**AI应用拓展不及预期:** AI各下游领域应用拓展如果不及预期，会影响到AI产业整体发展，也会影响到AI对能源电力需求拉动的作用

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

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>

# 东吴证券 财富家园