

# 计算机行业跟踪

## 从 Deepseek-V2 定价看大模型降本新风向

### 投资要点:

- 幻方发布第二代 MoE 模型 DeepSeek-V2，综合性能位列第一梯队。

2024年5月6日，幻方发布第二代 MoE 模型 DeepSeek-V2。在 AlignBench、MT-Bench、MMLU 等多个 benchmark 上表现出色，其中 AlignBench 在开源模型中居首位，与 GPT-4-Turbo，文心 4.0 比肩。MT-Bench 超过最强 MoE 开源模型 Mixtral 8x22B。

- DeepSeek-V2 定价降至冰点，大模型价格战拉开序幕。

DeepSeek-V2 (32k) 每百万 tokens 输入/输出价格分别为 1/2 元，而 GPT-4-Turbo-1106 分别为 72/217 元，DeepSeek-V2 性价比显著。相对于 Claude 3 Haiku，DeepSeek-V2 每百万 tokens 输入/输出价格也仅为其 50%/22.2%。除此之外，同为 32k 上下文版本的 moonshot-v1、SenseChat-32K、Qwen1.5 72B 每百万 tokens 输入/输出价格分别为 24/24、36/36、20/20 元。此次 DeepSeek-V2 定价发布有望掀起新一轮大模型价格战，api 定价有望持续走低。

- 大模型定价下降的背后离不开成本的优化。

目前，众多大模型已经通过多种方式降低成本。从模型压缩的方向看，可以通过量化的形式将浮点表征为低位宽模型来压缩模型存储空间，加速模型推理；从模型架构的方向看，MoE 架构由于其内部的专家模型能够分配到不同设备，并可以执行并行计算，其计算效率较稠密模型显著提升，进而带来更低的成本。从 tokens 量的方向来看，可以通过 prompt 压缩等方式直接降低输入端 tokens，进而降低成本。除此之外，多种新的方案已出现在相关论文中，未来多种成本优化方案的融合将进一步加速模型成本的下降。

- 我们认为，大模型成本优化与算力需求并不是直接的此长彼消，而是互相搭台、相互成就。定价的持续走低有望带来更快的商业化落地，进而会衍生出更多的微调及推理等需求，将逐步盘活国内 AI 应用及国产算力发展。

- 建议关注:

AI 算力：寒武纪、海光信息、浪潮信息、中科曙光等。

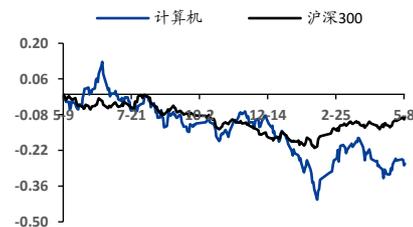
AI 应用：金山办公、科大讯飞、福昕软件、云从科技、商汤、出门问问等。

- 风险提示

大模型优化不及预期的风险；商业落地不及预期的风险等。

## 强于大市（维持评级）

### 一年内行业相对大盘走势



### 团队成员

分析师：钱劲宇(S0210524040006)  
QJY3773@hfzq.com.cn

### 相关报告

- 1、关注智能电表更新换代周期——2024.03.12
- 2、国产算力为基座，“人工智能+”蓄势待发——2024.03.05
- 3、Sora 技术深度解析——2024.02.25



## 分析师声明

本人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。

## 一般声明

华福证券有限责任公司（以下简称“本公司”）具有中国证监会许可的证券投资咨询业务资格。本报告仅供本公司的客户使用。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。

本报告的信息均来源于本公司认为可信的公开资料，该等公开资料的准确性及完整性由其发布者负责，本公司及其研究人员对该等信息不作任何保证。本报告中的资料、意见及预测仅反映本公司于发布本报告当日的判断，之后可能会随情况的变化而调整。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。本公司不保证本报告所含信息及资料保持在最新状态，对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

在任何情况下，本报告所载的信息或所做出的任何建议、意见及推测并不构成所述证券买卖的出价或询价，也不构成对所述金融产品、产品发行或管理人作出任何形式的保证。在任何情况下，本公司仅承诺以勤勉的职业态度，独立、客观地出具本报告以供投资者参考，但不就本报告中的任何内容对任何投资做出任何形式的承诺或担保。投资者应自行决策，自担投资风险。

本报告版权归“华福证券有限责任公司”所有。本公司对本报告保留一切权利。除非另有书面显示，否则本报告中的所有材料的版权均属本公司。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。未经授权的转载，本公司不承担任何转载责任。

## 特别声明

投资者应注意，在法律许可的情况下，本公司及其本公司的关联机构可能会持有本报告中涉及的公司所发行的证券并进行交易，也可能为这些公司正在提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。投资者请勿将本报告视为投资或其他决定的唯一参考依据。

## 投资评级声明

类别	评级	评级说明
公司评级	买入	未来 6 个月内，个股相对市场基准指数涨幅在 20%以上
	持有	未来 6 个月内，个股相对市场基准指数涨幅介于 10%与 20%之间
	中性	未来 6 个月内，个股相对市场基准指数涨幅介于-10%与 10%之间
	回避	未来 6 个月内，个股相对市场基准指数涨幅介于-20%与-10%之间
	卖出	未来 6 个月内，个股相对市场基准指数涨幅在-20%以下
行业评级	强于大市	未来 6 个月内，行业整体回报高于市场基准指数 5%以上
	跟随大市	未来 6 个月内，行业整体回报介于市场基准指数-5%与 5%之间
	弱于大市	未来 6 个月内，行业整体回报低于市场基准指数-5%以下

备注：评级标准为报告发布日后的 6~12 个月内公司股价（或行业指数）相对同期基准指数的相对市场表现。其中 A 股市场以沪深 300 指数为基准；香港市场以恒生指数为基准，美股市场以标普 500 指数或纳斯达克综合指数为基准（另有说明的除外）

## 联系方式

华福证券研究所 上海

公司地址：上海市浦东新区浦明路 1436 号陆家嘴滨江中心 MT 座 20 层

邮编：200120

邮箱：hfjys@hfzq.com.cn