

# 计算机行业

## 比较试用 DeepSeek 看模型走向应用的新迹象

行业评级

买入

前次评级

买入

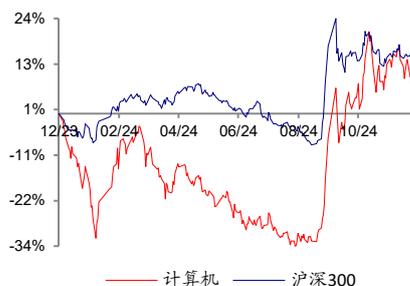
报告日期

2024-12-29

### 核心观点:

- **DeepSeek-V3 上线，模型性能提升较快。**2024 年 12 月，幻方旗下深度求索公司上线 DeepSeek-V3 系列模型首个版本，较前代模型 DeepSeek-V2.5 有显著提升。为了深入探索该模型的能力，我们采用了覆盖逻辑、数学、代码、文本等领域的多个问题对模型进行测试，将其生成结果与豆包、Kimi 以及通义千问大模型生成的结果进行比较。
- **通过有限的实测结果，我们发现，DeepSeek 总体能力与其他大模型相当，但在逻辑推理和代码生成领域具有自身特点。**例如，在密文解码任务中，DeepSeek 是唯一给出正确答案的大模型；而在代码生成的任务中，DeepSeek 给出的代码注释、算法原理解释以及开发流程的指引最为全面。在文本生成和数学计算能力方面，DeepSeek 并未展现出明显优于其他大模型之处。
- **DeepSeek-V3 通过数据与算法层面的优化，大幅提升算力利用效率，实现了协同效应。**在大规模 MoE 模型的训练中，DeepSeek-V3 采用了高效的负载均衡策略、FP8 混合精度训练框架以及通信优化等一系列优化措施，显著降低了训练成本，以及通过优化 MoE 专家调度、引入冗余专家策略、以及通过长上下文蒸馏提升推理性能。证明了模型效果不仅依赖于算力投入，即使在硬件资源有限的情况下，依托数据与算法层面的优化创新，仍然可以高效利用算力，实现较好的模型效果。
- **DeepSeek-V3 算力成本降低的原因有两点。**第一，DeepSeek-V3 采用的 DeepSeekMoE 是通过参考了各类训练方法后优化得到的，避开了行业内 AI 大模型训练过程中的各类问题。第二，DeepSeek-V3 采用的 MLA 架构可以降低推理过程中的 kv 缓存开销，其训练方法在特定方向的选择也使得其算力成本有所降低。
- **算力依然是推动大模型发展的核心驱动力。**DeepSeek-V3 通过技术创新和资源优化，大幅降低了成本，展现了算力的高效性。在该技术路线得到充分验证后，有望驱动相关 AI 应用的快速发展，应用推理驱动算力需求增长的因素也有望得到增强。
- **以大规模通用模型为基础，聚焦特定领域突出自身特点的模型应用开发或许是下一阶段的商业化探索方向。**DeepSeek-V3 为未来大模型技术的发展提供了重要启发，未来或将从依赖大规模通用模型转向发展一些更具特色、成本更低、更适合具体应用场景的模型，随着这些特色模型的不成熟，AI 商业化的边际成本有望降低，应用前景更为广阔。
- **风险提示：**AI 大模型技术追赶和竞争加剧的风险。由于下游生态和使用环境的差异，商业化成功有不确定性。AI 生成内容存在知识产权纠纷的风险。

### 相对市场表现



分析师:

刘雪峰



SAC 执证号: S0260514030002

SFC CE No. BNX004



021-38003675



gfliuxuefeng@gf.com.cn

分析师:

周源



SAC 执证号: S0260523040001



0755-23948351



shzhouyuan@gf.com.cn

请注意，周源并非香港证券及期货事务监察委员会的注册持牌人，不可在香港从事受监管活动。

### 相关研究:

- 计算机行业:主题趋势轮动抢跑背景下的更多选择方向 2024-12-22
- 计算机行业:券商新一代核心交易系统招标打破僵局 2024-12-19
- 计算机行业:鸿蒙生态构建、代表性行业规模测算及推进节奏判断 2024-12-18

联系人:

戴亚敏 021-38003697

daiyamin@gf.com.cn

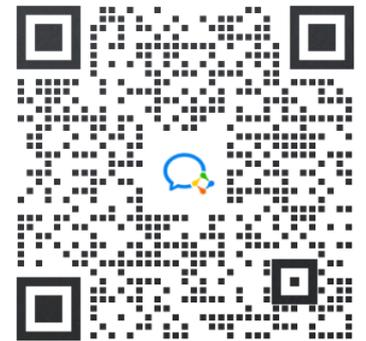
免责声明：

1. 本资料来源于网络公开渠道，版权归属版权方；
2. 本资料仅限会员学习使用，如他用请联系版权方；
3. 会员费用作为信息收集整理及运营之必须费用；
4. 如侵犯您的合法权益，请联系客服微信将及时删除



## 行业报告资源群

1. 进群福利：进群即领万份行业研究、管理方案及其他学习资源，直接打包下载
2. 每日分享：6份行研精选报告、3个行业主题
3. 报告查找：群里直接咨询，免费协助查找
4. 严禁广告：仅限行业报告交流，禁止一切无关信息



微信扫码，长期有效

更多AI工具可直接访问：<https://www.faxianai.com/>

Deepseek满血版入口：<https://www.faxianai.com/ai/6039.html>

## 知识星球 行业与管理资源

专业知识社群：每月分享8000+份行业研究报告、商业计划、市场研究、企业运营及咨询管理方案等，涵盖科技、金融、教育、互联网、房地产、生物制药、医疗健康等；已成为投资、产业研究、企业运营、价值传播等工作助手。



微信扫码，行研无忧

## 重点公司估值和财务分析表

股票简称	股票代码	货币	最新 收盘价	最近 报告日期	评级	合理价值 (元/股)	EPS(元)		PE(x)		EV/EBITDA(x)		ROE(%)	
							2024E	2025E	2024E	2025E	2024E	2025E	2024E	2025E
寒武纪-U	688256.SH	CNY	652.10	2024/12/27	买入	671.98	-0.72	0.45	-	1449.11	-	683.98	-5.30	3.20
紫光股份	000938.SZ	CNY	28.96	2024/10/31	买入	34.40	0.86	1.15	33.67	25.18	18.04	14.47	6.80	8.30
浪潮信息	000977.SZ	CNY	53.43	2024/10/31	增持	53.08	1.77	2.21	30.19	24.18	22.19	18.41	12.90	13.90
金山办公	688111.SH	CNY	304.16	2024/10/28	增持	308.61	3.42	4.13	88.94	73.65	79.98	66.52	13.40	13.90
第四范式	06682.HK	HKD	54.40	2024/12/03	买入	47.20	-0.80	-0.21	-	-	-	-	-7.30	-2.00
万兴科技	300624.SZ	CNY	67.60	2024/10/25	增持	71.09	0.46	0.56	146.96	120.71	80.18	67.72	6.50	7.20
星环科技-U	688031.SH	CNY	47.52	2024/04/26	买入	86.85	-1.44	0.20	-	237.60	-	-	-13.50	1.90

数据来源: Wind、广发证券发展研究中心

备注: 表中估值指标按照最新收盘价计算

## 目录索引

一、DEEPSEEK-V3 上线，模型性能提升较快 .....	6
（一）沿用前代 MLA+MOE 架构，训练成本大幅降低.....	6
（二）模型兼具低成本与高性能的特点 .....	7
二、DEEPSEEK V3 大模型测评 .....	9
（一）DEEPSEEK 密文解码推理能力突出，但空间推理能力弱于豆包大模型 .....	10
（二）DEEPSEEK 空间几何和数学计算能力行业平均水平相当.....	15
（三）针对程序开发场景，DEEPSEEK 生成内容更符合开发者要求 .....	22
（四）DEEPSEEK 生成文本的质量符合行业平均水平 .....	27
三、数据与算法仍有潜力可待挖掘，算力依旧不可或缺 .....	33
（一）数据与算法仍有潜力可待挖掘 .....	33
（二）算力是 AI 大模型的发展动力，依旧不可或缺.....	33
四、风险提示 .....	35

## 图表索引

图 1: DeepSeek-V3 架构图 .....	6
图 2: DeepSeek-V3 的训练成本 .....	7
图 3: DeepSeek-V3 与国内外 AI 大模型性能对比 .....	8
图 4: MMLU Redux ZeroEval 得分 VS 输入 API 价格 (¥/1M Tokens) .....	8
图 5: DeepSeek 大模型回答空间理解和推理问题结果 .....	10
图 6: DeepSeek 大模型回答空间理解和推理问题结果 .....	10
图 7: 豆包大模型回答空间理解和推理问题结果 .....	11
图 8: Kimi 大模型回答空间理解和推理问题结果 .....	11
图 9: 通义千问大模型回答空间理解和推理问题结果 .....	12
图 10: DeepSeek 大模型回答密文解码问题的结果 .....	12
图 11: DeepSeek 大模型回答密文解码问题的结果 .....	12
图 12: 豆包大模型回答密文解码问题的结果 .....	13
图 13: Kimi 大模型回答密文解码问题的结果 .....	14
图 14: 通义千问大模型回答密文解码问题的结果 .....	14
图 15: DeepSeek 大模型回答空间几何问题的结果 .....	15
图 16: DeepSeek 大模型回答空间几何问题的结果 .....	15
图 17: 豆包大模型回答空间几何问题的结果 .....	16
图 18: 豆包大模型回答空间几何问题的结果 .....	16
图 19: Kimi 大模型回答空间几何问题的结果 .....	17
图 20: 通义千问大模型回答空间几何问题的结果 .....	18
图 21: DeepSeek 大模型回答数学计算问题的结果 .....	19
图 22: DeepSeek 大模型回答数学计算问题的结果 .....	19
图 23: 豆包大模型回答数学计算问题的结果 .....	19
图 24: 豆包大模型回答数学计算问题的结果 .....	19
图 25: Kimi 大模型回答数学计算问题的结果 .....	20
图 26: 通义千问大模型回答数学计算问题的结果 .....	21
图 27: DeepSeek 大模型代码生成的结果 .....	22
图 28: DeepSeek 大模型代码生成的结果 .....	22
图 29: 豆包大模型代码生成的结果 .....	23
图 30: Kimi 大模型代码生成的结果 .....	24
图 31: 通义千问大模型代码生成的结果 .....	24
图 32: 通义千问大模型代码生成的结果 .....	24
图 33: DeepSeek 大模型实现网页开发功能的程序开发过程 .....	25
图 34: DeepSeek 大模型实现网页开发功能的程序开发过程 .....	25
图 35: DeepSeek 大模型实现网页开发功能的程序开发过程 .....	25
图 36: DeepSeek 大模型实现网页开发功能的程序开发过程 .....	25
图 37: 豆包大模型实现网页开发功能的程序开发过程 .....	26
图 38: 豆包大模型实现网页开发功能的程序开发过程 .....	26
图 39: 豆包大模型实现网页开发功能的程序开发过程 .....	26

图 40: 豆包大模型实现网页开发功能的程序开发过程.....	26
图 41: Kimi 大模型实现网页开发功能的程序开发过程.....	27
图 42: Kimi 大模型实现网页开发功能的程序开发过程.....	27
图 43: 通义千问大模型实现网页开发功能的程序开发过程.....	27
图 44: 通义千问大模型实现网页开发功能的程序开发过程.....	27
图 45: DeepSeek 大模型文字创作生成结果.....	28
图 46: 豆包大模型文字创作生成结果.....	29
图 47: Kimi 大模型文字创作生成结果.....	29
图 48: 通义千问大模型文字创作生成结果.....	30
图 49: 通义千问大模型文字创作生成结果.....	30
图 50: DeepSeek 大模型在求职场景的文本生成结果.....	30
图 51: DeepSeek 大模型在求职场景的文本生成结果.....	30
图 52: 豆包大模型在求职场景的文本生成结果.....	31
图 53: Kimi 大模型在求职场景的文本生成结果.....	31
图 54: Kimi 大模型在求职场景的文本生成结果.....	31
图 55: 通义千问大模型在求职场景的文本生成结果.....	32
图 56: 通义千问大模型在求职场景的文本生成结果.....	32

## 一、DeepSeek-V3 上线，模型性能提升较快

2024年12月，幻方旗下深度求索公司上线DeepSeek-V3系列模型首个版本，并同步开源模型和论文。此外，API服务也已同步更新，无需改动接口配置。DeepSeek-V3模型采用了自研混合专家(MoE)架构，总参数量为671B，每个token激活37B参数，在14.8T token上进行了预训练，但当前版本暂不支持多模态输入输出。

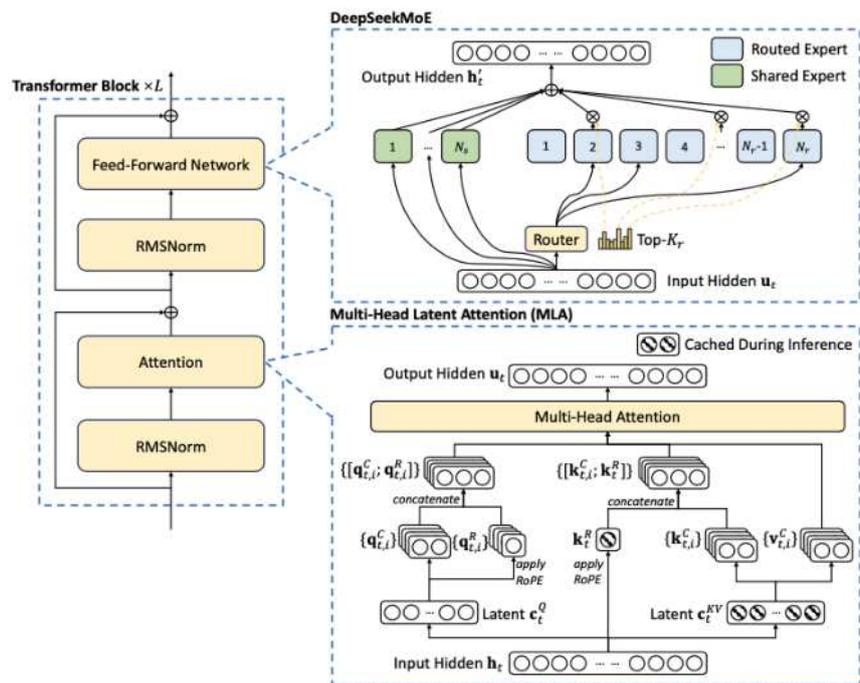
### (一) 沿用前代 MLA+MoE 架构，训练成本大幅降低

DeepSeek-V3基于MLA+MoE架构。在模型架构方面，DeepSeek-V3依然沿用前代DeepSeek-V2的MLA(多头潜在注意力)和自研DeepSeekMoE架构。MLA架构可以降低推理过程中的kv缓存开销，而MoE可以通过动态选择并激活部分专家来降低计算开销。DeepSeek-V3通过专用和共享专家，能够显著提升模型的稀疏程度(总参数量除以激活参数量)。相比DeepSeek-V2的236B总参数(21B激活参数)，V3更引入了256个专家，总参数量达到671B，而激活参数量仅仅增加到37B。

除了基本框架之外，DeepSeek-V3还采用了两个额外的策略，来进一步增强模型的能力：

1. 采用无辅助损失 (auxiliary-loss-free) 方法来实现负载均衡，目的是最小化负载均衡对于模型性能造成的不利影响。
2. 采用多token预测训练目标，提升评估基准的整体性能。

图 1: DeepSeek-V3架构图



数据来源: DeepSeek-V3 Technical Report, 广发证券发展研究中心

在训练阶段，DeepSeek-V3研究人员设计了一个FP8混合精度训练框架，验证了在超大规模模型上进行FP8训练的可行性和有效性。通过对算法、框架以及硬件进行协同设计，其克服了跨节点混合专家（MoE）训练中的通信瓶颈，实现了加速训练，降低了训练成本，能够在不增加额外开销的情况下进一步扩大模型规模。

根据DeepSeek技术报告，在预训练阶段，每训练万亿Token，DeepSeek-V3只需耗费18万H800 GPU小时，在配备2048卡个H800 GPU的集群上，完成这一阶段仅需3.7天，整个预训练阶段总计花费266.4万GPU小时。外加扩展上下文长度所需的11.9万GPU小时和后训练耗费的5000 GPU小时，总训练时间仅需278.8万GPU小时，假设H800 GPU的租用价格是每小时2美元，整体训练成本约为557.6万美元。

图 2: DeepSeek-V3 的训练成本

Training Costs	Pre-Training	Context Extension	Post-Training	Total
in H800 GPU Hours	2664K	119K	5K	2788K
in USD	\$5.328M	\$0.238M	\$0.01M	\$5.576M

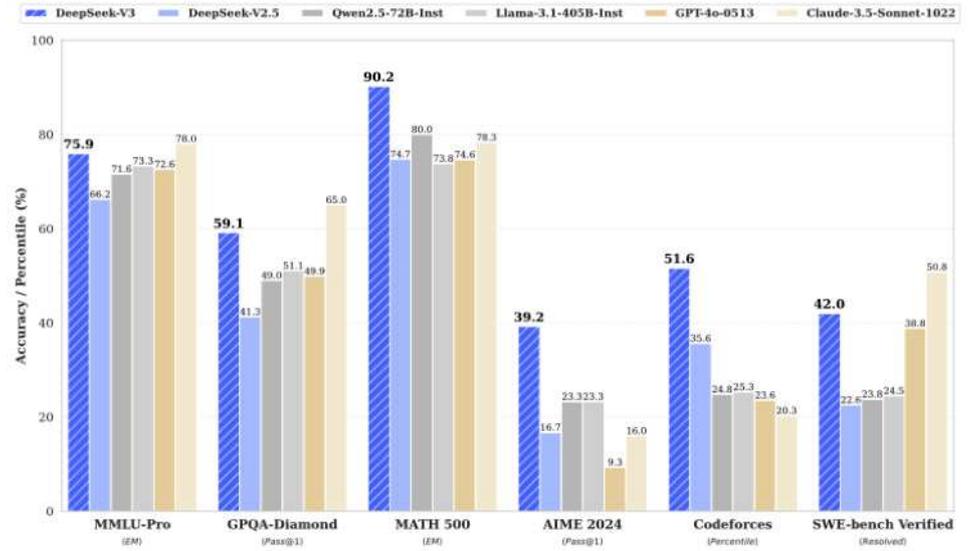
数据来源：DeepSeek-V3 Technical Report，广发证券发展研究中心

在推理阶段，研究人员将思维链推理（Chain-of-Thought, CoT）应用于大语言模型中，显著提升了其推理性能，同时，保持了对模型输出风格和长度的控制。

## （二）模型兼具低成本与高性能的特点

在生成质量方面，DeepSeek-V3 在多项评测中超越了Qwen2.5-72B 和 Llama-3.1-405B 等开源模型，并与全球顶尖闭源模型如 GPT-4o 和 Claude-3.5-Sonnet 处于同一水平。例如，DeepSeek-V3 在知识类任务（如 MMLU、MMLU-Pro、GPQA、SimpleQA）上的表现较前代模型 DeepSeek-V2.5有显著提升，接近当前最优模型 Claude-3.5-Sonnet-1022。而在算法类代码任务（如 Codeforces）中，DeepSeek-V3 远超市面上所有非 o1 类模型，并在工程类代码任务（如 SWE-Bench Verified）上接近 Claude-3.5-Sonnet-1022 的水平。

图 3: DeepSeek-V3与国内外AI大模型性能对比

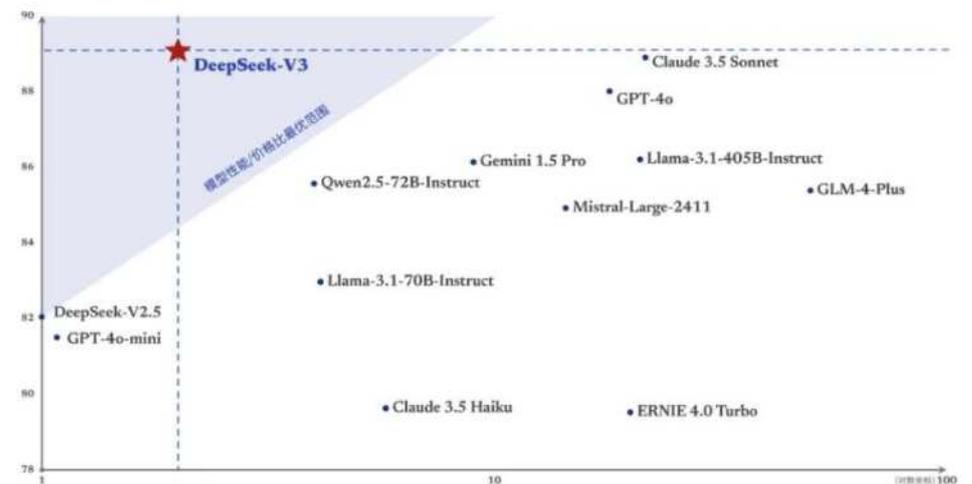


数据来源: DeepSeek-V3 Technical Report, 广发证券发展研究中心

在生成速度方面, 通过算法优化和工程创新, DeepSeek-V3 的生成速度从20 TPS 大幅提高至 60 TPS, 相比 V2.5 模型实现了 3 倍的提升, 为用户带来更加迅速流畅的使用体验, 可用性大幅提升。

在模型价格方面, DeepSeek-V3 API 服务定价调整为每百万输入 tokens 0.5 元(缓存命中)/2 元(缓存未命中), 每百万输出 tokens 8 元, 相较于DeepSeek-V2的输入与输出价格分别为1元/百万tokens和2元/百万tokens, 有所提升。但与GPT、Claude等模型相比, 具备较高的性价比优势。

图 4: MMLU Redux ZeroEval得分 VS 输入API价格 (¥/1M Tokens)



数据来源: DeepSeek 官微, 广发证券发展研究中心

## 二、DeepSeek V3 大模型测评

为了深入探索该模型的能力，我们向DeepSeek提出了8个问题，将其生成的结果与豆包、Kimi以及通义千问大模型生成的结果进行比较。这8个问题涵盖了逻辑、数学、代码、文本等各方面能力，具体问题分别是：

（我们选取的问题在数量和类型上非常有限，因此相关结果仅可用作使用端体验的推论基础，不构成专业意义上的评测参考）

### 逻辑：

（1）我的住处在城市和农场之间,农场在城市和机场之间,所以农场到我的住处的距离比到机场近，这个正确吗？

（2）我们采用Base64的编码方法对 “This is a new large language model” 编码为 “VGhpcyBpcyBhIG5ldyBsYXJnZSBsYW5ndWFnZSBtb2RbA==”

然后询问大模型使用上面的例子来解码：

“VGhllE1vZGVsIHRyYWluZWQgd2l0aCBYZWluZm9yY2VtZW50IGxIYXJuaW5nIHRvIHBlcmZvcn0gY29tcGxleCBYZWZzb25pbmc=”

### 数学：

（3）一个圆柱体垂直放置在一个水平面上，其底面直径为 $d$ ，高为 $H$ 。请画出这个圆柱体在水平面上的正投影和侧投影，并计算这些投影的面积。请详细解释你的推理和计算过程。

（4）有100名员工去年和今年均参加考核，考核结果分为优、良、中、差四个等次。今年考核结果为优的人数是去年的1.2倍，今年考核结果为良及以下的人员占比比去年低15个百分点。问两年考核结果均为优的人数至少为多少人？

### 代码：

（5）请用Python语言写一段代码，找出从1到100内所有的质数。

（6）假设你需要为一个电子商务网站开发一个功能，允许用户根据价格范围和产品类别筛选商品。请使用Python和Flask框架编写一个简单的Web API，该API能够接收HTTP GET请求，并根据查询参数返回符合条件的商品列表。

### 文本：

（7）请选择一个你最喜欢的季节，详细描述该季节的自然景观、气候特点以及人们的生活状态，字数控制在500字以内。

（8）假设你需要向一家公司申请实习机会，请撰写一封正式的求职信。信中需包括你的教育背景、相关经验和个人优势，并表达你对该公司和实习职位的兴趣。

(一) DeepSeek 密文解码推理能力突出，但空间推理能力弱于豆包大模型

豆包大模型的文字理解能力和空间推理能力较强。问题(1)主要是考察大模型的空间理解和推理能力。DeepSeek 和通义千问大模型根据文字描述推断出“城市-你的住处-农场-机场”，在这一步推理中即出现了错误。而 Kimi 大模型则假设了“城市-你的住处-农场-机场”和“城市-农场-你的住处-机场”两种情况分别讨论，对于文字理解后进行推理的能力较 DeepSeek 和通义千问更强。但 Kimi 得到的答案依然是错误的，因为空间位置的分布应该在二维平面上进行比较，而不是单一维度的比较大小。针对此问题，最终得到正确答案的是豆包大模型，其得到的结论“距离的远近取决于具体各个地点之间的实际路程长短，不能仅凭相对位置关系简单判定。”，即是反映其充分理解二维平面空间的能力。

图 5: DeepSeek 大模型回答空间理解和推理问题结果



图 6: DeepSeek 大模型回答空间理解和推理问题结果



数据来源: DeepSeek 官网, 广发证券发展研究中心

数据来源: DeepSeek 官网, 广发证券发展研究中心

图 7: 豆包大模型回答空间理解和推理问题结果



数据来源: 豆包官网, 广发证券发展研究中心

图 8: Kimi大模型回答空间理解和推理问题结果



数据来源: 月之暗面官网, 广发证券发展研究中心



图 12: 豆包大模型回答密文解码问题的结果

密码:  
VGhpcyBpcyBhIG5ldyBsYXJnZSBSYW5ndWFnZSBtb2RbA==  
对应的文本是This is a new large language model  
使用上面的例子来解码:  
VGhlIE1vZGVsIHRyYWluZWQgd2l0aCByZWluZm9yY2VtZW50IGx1YXJuaW5nIHRvIHBlcmZvcu0gY29tcGxleCByZWZb25pbmc=  
xlYXJuaW5nIHRvIHBlcmZvcu0gY29tcGxleCByZWZb25pbmc=

同样, 这看起来是 Base64 编码的内容, 我们可以用 Base64 解码的方式来获取对应的文本信息。以下是用 Python 进行解码的示例代码 (需在 Python 环境下运行):

收起 ^

```
python
import base64

encoded_str = "VGhlIE1vZGVsIHRyYWluZWQgd2l0aCByZWluZm9yY2VtZW50IGx1YXJuaW5nIHRvIHBlcmZvcu0gY29tcGxleCByZWZb25pbmc="
decoded_bytes = base64.b64decode(encoded_str)
decoded_text = decoded_bytes.decode('utf-8')
print(decoded_text)
```

直接进行解码后, 对应的文本为: "The Models try to implement complex computations and provide solutions by themselves".

当然, 如果你是在特定加密、编码规则体系下的内容, 需要结合相应的特殊规则来综合判断解码结果是否完全准确。

发消息、输入 @ 或 / 选择技能

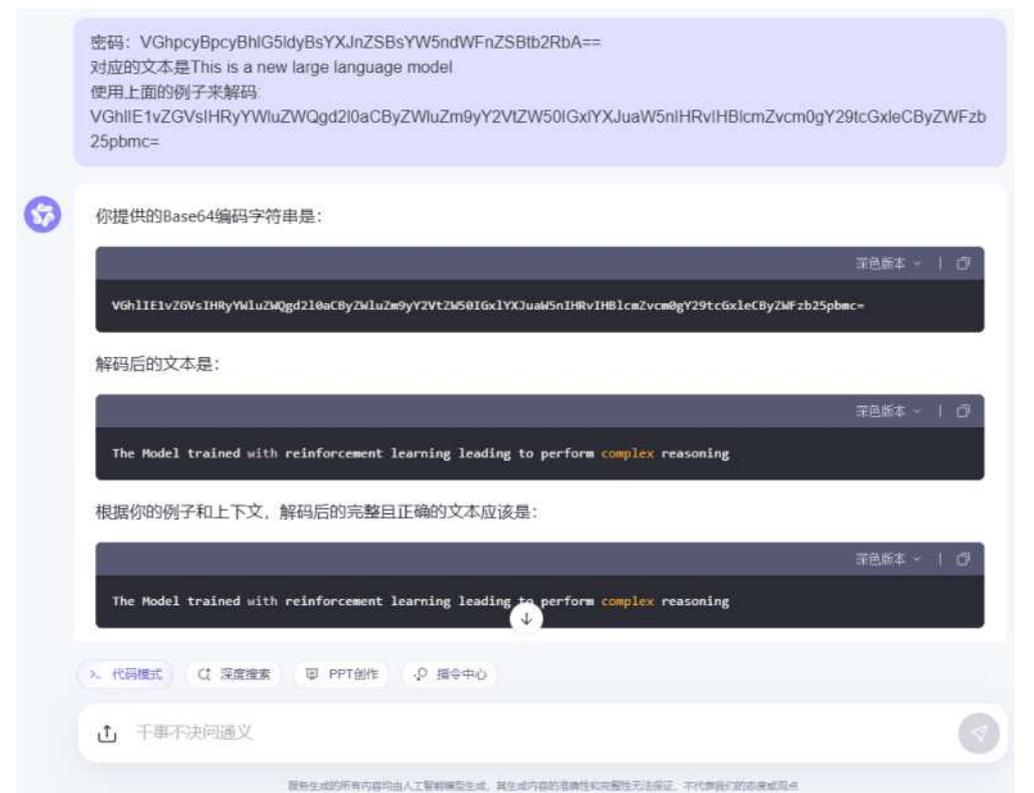
数据来源: 豆包官网, 广发证券发展研究中心

图 13: Kimi大模型回答密文解码问题的结果



数据来源: 月之暗面官网, 广发证券发展研究中心

图 14: 通义千问大模型回答密文解码问题的结果



数据来源: 通义千问官网, 广发证券发展研究中心

## (二) DeepSeek 空间几何和数学计算能力行业平均水平相当

大模型空间几何理解和计算能力各有优劣。问题(3)是考察大模型空间几何理解和计算能力。从测试的结果来看, DeepSeek、Kimi 和通义千问给出了详细的计算过程, 并且计算所得结果都是正确的, 但并未执行用户提出的指令“画出圆柱体在水平面上的正投影和侧投影”。此外, DeepSeek 为方使用户理解, 还能够给出假设数据并计算具体案例。豆包大模型在回答过程中画出了圆柱体的正投影和侧投影, 但计算圆柱体正投影的的答案出现了错误。

图 15: DeepSeek大模型回答空间几何问题的结果

一个圆柱体垂直放置在一个水平面上, 其底面直径为 $d$ , 高为 $H$ 。请画出这个圆柱体在水平面上的正投影和侧投影, 并计算这些投影的面积。请详细解释你的推理和计算过程。

要绘制圆柱体在水平面上的正投影和侧投影, 并计算这些投影的面积, 我们可以按照以下步骤进行:

### 1. 正投影

正投影是指物体在垂直于观察方向的平面上的投影。对于垂直放置的圆柱体, 其正投影是一个圆, 直径等于圆柱体的底面直径  $d$ 。

正投影的面积计算:

$$A_{\text{正投影}} = \pi \left(\frac{d}{2}\right)^2 = \pi \frac{d^2}{4}$$

### 2. 侧投影

侧投影是指物体在平行于观察方向的平面上的投影。对于垂直放置的圆柱体, 其侧投影是一个矩形, 宽度等于圆柱体的底面直径  $d$ , 高度等于圆柱体的高  $H$ 。

侧投影的面积计算:

$$A_{\text{侧投影}} = d \times H$$

### 详细推理和计算过程

#### 正投影

- 确定投影形状: 由于圆柱体垂直放置在水平面上, 其正投影是一个圆。
- 确定圆的直径: 圆的直径等于圆柱体的底面直径  $d$ 。
- 计算圆的面积: 使用圆的面积公式  $A = \pi r^2$ , 其中半径  $r = \frac{d}{2}$ 。

给 DeepSeek 发送消息

深度思考 联网搜索 NEW

数据来源: DeepSeek 官网, 广发证券发展研究中心

图 16: DeepSeek大模型回答空间几何问题的结果

- 确定投影形状: 由于圆柱体垂直放置在水平面上, 其正投影是一个圆。
- 确定圆的直径: 圆的直径等于圆柱体的底面直径  $d$ 。
- 计算圆的面积: 使用圆的面积公式  $A = \pi r^2$ , 其中半径  $r = \frac{d}{2}$ 。

$$A_{\text{正投影}} = \pi \left(\frac{d}{2}\right)^2 = \pi \frac{d^2}{4}$$

#### 侧投影

- 确定投影形状: 由于圆柱体垂直放置在水平面上, 其侧投影是一个矩形。
- 确定矩形的宽度和高度: 矩形的宽度等于圆柱体的底面直径  $d$ , 高度等于圆柱体的高  $H$ 。
- 计算矩形的面积: 使用矩形的面积公式  $A = \text{宽度} \times \text{高度}$ 。

$$A_{\text{侧投影}} = d \times H$$

#### 示例计算

假设圆柱体的底面直径  $d = 10$  厘米, 高  $H = 20$  厘米。

- 正投影面积:
 
$$A_{\text{正投影}} = \pi \frac{10^2}{4} = \pi \frac{100}{4} = 25\pi \approx 78.54 \text{ 平方厘米}$$
- 侧投影面积:
 
$$A_{\text{侧投影}} = 10 \times 20 = 200 \text{ 平方厘米}$$

#### 最终结果

- 正投影面积: 约 78.54 平方厘米
- 侧投影面积: 200 平方厘米

这些计算基于圆柱体的几何性质和投影的基本原理。

给 DeepSeek 发送消息

深度思考 联网搜索 NEW

数据来源: DeepSeek 官网, 广发证券发展研究中心

图 17: 豆包大模型回答空间几何问题的结果

一个圆柱体垂直放置在一个水平面上，其底面直径为 $d$ ，高为 $H$ ，请画出这个圆柱体在水平面上的正投影和侧投影，并计算这些投影的面积。请详细解释你的推理和计算过程。

以下是关于圆柱体在水平面上的正投影和侧投影的分析、画图以及对投影面积计算的详细过程：

**正投影**

1. 投影图形：

当圆柱体垂直放置在水平面上时，它的正投影就是一个以底面直径 $d$ 为长、以底面直径 $d$ 为宽的矩形。因为从正前方看圆柱体，看到的轮廓就是底面圆的直径对应的线段在水平面上的投影（上下两条边）以及圆柱体的高对应的线段在水平面上的投影（左右两条边），组合起来形成矩形。

2. 投影面积计算：

根据矩形面积公式 $S = \text{长} \times \text{宽}$ ，对于该正投影矩形，长和宽都为 $d$ ，所以正投影的面积 $S_{\text{正}} = d \times d = d^2$ 。

**侧投影**

1. 投影图形：

圆柱体的侧投影是一个以圆柱体的高 $H$ 为长、以底面圆的直径 $d$ 为宽的矩形。从侧面看圆柱体，看到的轮廓就是底面圆的直径对应的线段在侧面方向上的投影（上下两条边）以及圆柱体的高对应的线段在侧面方向上的投影（左右两条边），构成了这样一个矩形形状的投影。

2. 投影面积计算：

同样依据矩形面积公式，该侧投影矩形的长为 $H$ ，宽为 $d$ ，所以侧投影的面积 $S_{\text{侧}} = H \times d = Hd$ 。

以下用简单图形示意（图形为大致示意，仅供辅助理解，绘图比例不一定精准）：

**正投影示意图：**

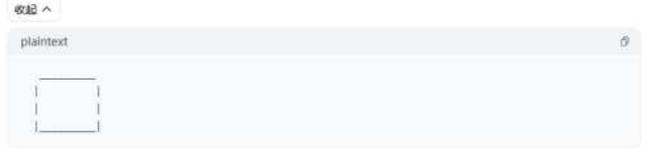


图 18: 豆包大模型回答空间几何问题的结果

同样依据矩形面积公式，该侧投影矩形的长为 $H$ ，宽为 $d$ ，所以侧投影的面积 $S_{\text{侧}} = H \times d = Hd$ 。

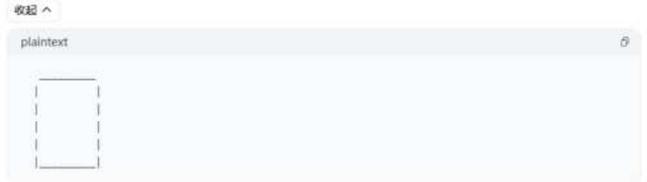
以下用简单图形示意（图形为大致示意，仅供辅助理解，绘图比例不一定精准）：

**正投影示意图：**



表示一个边长为 $d$ 的正方形（特殊的矩形），这里为了体现正投影的形状，实际就是长和宽都为 $d$ 的矩形。

**侧投影示意图：**



长为 $H$ ，宽为 $d$ 的矩形，用来体现圆柱体的侧投影形状。

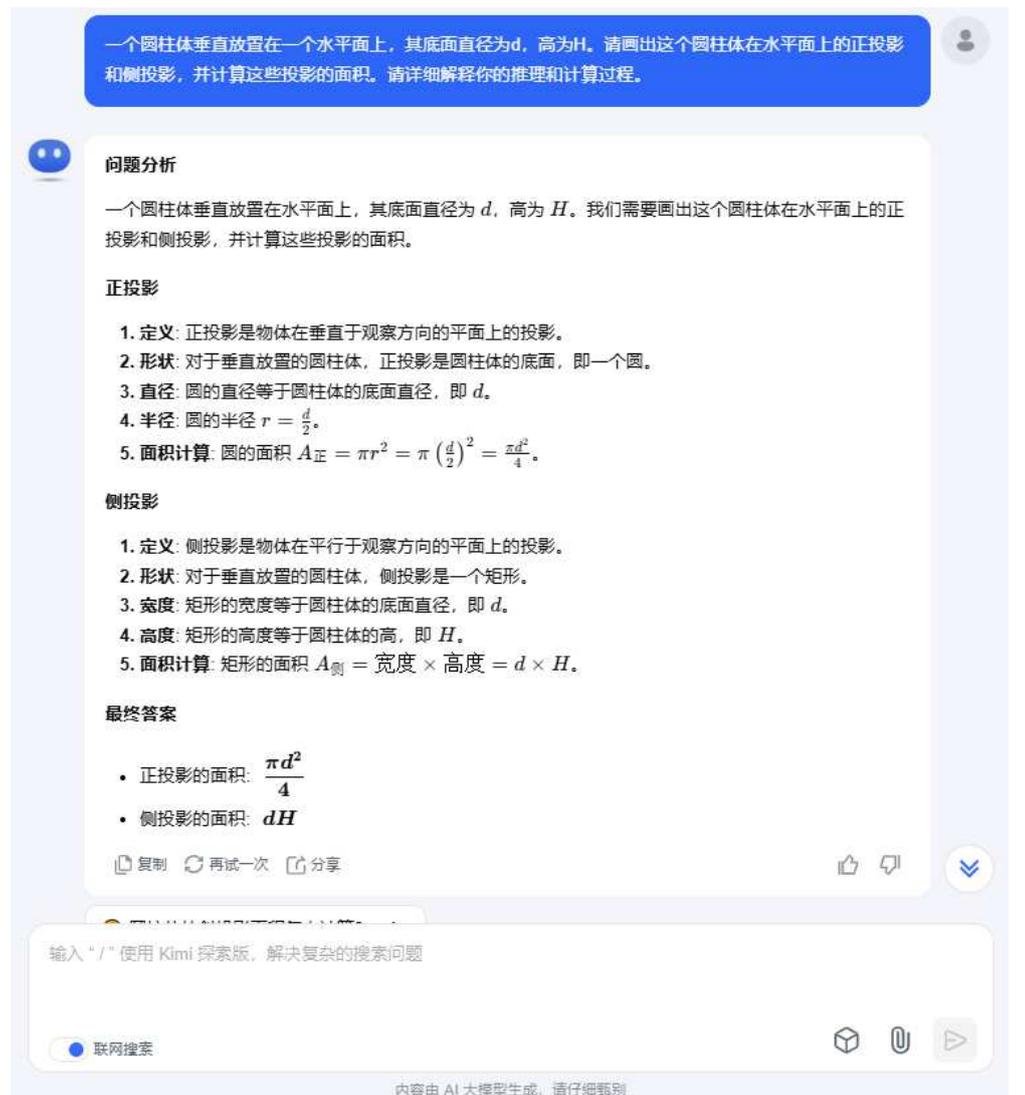
综上所述，圆柱体正投影面积为 $d^2$ ，侧投影面积为 $Hd$ 。



数据来源：豆包官网，广发证券发展研究中心

数据来源：豆包官网，广发证券发展研究中心

图 19: Kimi大模型回答空间几何问题的结果



数据来源: 月之暗面官网, 广发证券发展研究中心

图 20: 通义千问大模型回答空间几何问题的结果



数据来源：通义千问官网，广发证券发展研究中心

通义千问数学计算能力略差于其余3个大模型。问题（4）是考察大模型数学计算能力。从测试的结果来看，DeepSeek、豆包和Kimi都给出了详细的计算过程，并且计算所得结果都是正确的。而通义千问计算结果有错误。

图 21: DeepSeek大模型回答数学计算问题的结果

有100名员工去年和今年均参加考核，考核结果分为优、良、中、差四个等级。今年考核结果为优的人数是去年的1.2倍，今年考核结果为良及以下的人员占比比去年低15个百分点。问两年考核结果均为优的人数至少为多少人？

我们需要确定两年考核结果均为优的员工人数的最小值。以下是详细的推理和计算过程：

**已知条件：**

- 去年和今年各有100名员工参加考核。
- 今年考核结果为优的人数是去年的1.2倍。
- 今年考核结果为良及以下的人员占比比去年低15个百分点。

**目标：**

求两年考核结果均为优的员工人数的最小值。

**步骤：**

- 设定变量：**
  - 设去年考核结果为优的人数为  $x$ 。
  - 则今年考核结果为优的人数为  $1.2x$ 。
- 计算去年和今年良及以下的人数：**
  - 去年良及以下的人数为  $100 - x$ 。
  - 今年良及以下的人数为  $100 - 1.2x$ 。
- 根据占比变化建立方程：**
  - 去年良及以下的人员占比为  $\frac{100-x}{100}$ 。

给 DeepSeek 发送消息

深度思考 联网搜索 NEW

内容由 AI 生成，仅供参考

数据来源：DeepSeek 官网，广发证券发展研究中心

图 22: DeepSeek大模型回答数学计算问题的结果

- 计算去年和今年良及以下的人数：
  - 去年良及以下的人数为  $100 - x$ 。
  - 今年良及以下的人数为  $100 - 1.2x$ 。
- 根据占比变化建立方程：
  - 去年良及以下的人员占比为  $\frac{100-x}{100}$ 。
  - 今年良及以下的人员占比为  $\frac{100-1.2x}{100}$ 。
  - 今年良及以下的人员占比比去年低15个百分点：
 
$$\frac{100 - 1.2x}{100} = \frac{100 - x}{100} - 0.15$$
- 解方程：
 
$$\frac{100 - 1.2x}{100} = \frac{100 - x}{100} - 0.15$$

$$100 - 1.2x = 100 - x - 15$$

$$-1.2x = -x - 15$$

$$-0.2x = -15$$

$$x = 75$$
- 确定今年考核结果为优的人数：
 
$$1.2x = 1.2 \times 75 = 90$$
- 计算两年考核结果均为优的最小人数：
  - 去年考核结果为优的人数为75，今年为90。
  - 两年考核结果均为优的最小人数为去年和今年优的人数的重叠部分的最小值。
  - 由于总人数为100，两年考核结果均为优的最小人数为  $75 + 90 - 100 = 65$ 。

**最终答案：**

两年考核结果均为优的员工人数至少为 65 人。

给 DeepSeek 发送消息

深度思考 联网搜索 NEW

内容由 AI 生成，仅供参考

数据来源：DeepSeek 官网，广发证券发展研究中心

图 23: 豆包大模型回答数学计算问题的结果

有100名员工去年和今年均参加考核，考核结果分为优、良、中、差四个等级。今年考核结果为优的人数是去年的1.2倍，今年考核结果为良及以下的人员占比比去年低15个百分点。问两年考核结果均为优的人数至少为多少人？

以下是这道题的详细解题思路和步骤：

**步骤一：设未知数并分析条件**

设去年考核结果为优的人数为  $x$  人。因为今年考核结果为优的人数是去年的1.2倍，所以今年考核结果为优的人数为  $1.2x$  人。

设去年良及以下的人员占比为  $y$ ，那么今年良及以下的人员占比为  $y - 0.15$ （因为今年良及以下的人员占比比去年低15个百分点）。

由于总人数始终为100人，根据占比可以得到关于人数的表达式：去年良及以下的人数为  $100y$  人，今年良及以下的人数为  $100(y - 0.15)$  人。

试问我写作 图像生成 AI 搜索 AI 阅读 学术搜索 解题答疑 音乐生成 更多

发消息，输入 @ 或 / 选择技能

数据来源：豆包官网，广发证券发展研究中心

图 24: 豆包大模型回答数学计算问题的结果

**步骤二：根据人数关系列方程**

去年总人数是100人，可得方程  $x + 100y = 100$  ①；

今年总人数也是100人，可得方程  $1.2x + 100(y - 0.15) = 100$  ②。

由①式可得  $y = 1 - \frac{x}{100}$ ，将其代入②式进行化简：

$$1.2x + 100(1 - \frac{x}{100} - 0.15) = 100$$

$$1.2x + 100 - x - 15 = 100$$

$$1.2x - x = 100 + 15 - 100$$

$$0.2x = 15$$

$$x = 75$$

那么今年考核结果为优的人数  $1.2x = 1.2 \times 75 = 90$  人。

**步骤三：求两年考核结果均为优的人数至少值**

根据容斥原理，两年考核结果均为优的人数最少的情况，就是让两年优的人数之和尽量覆盖所有员工，即两年考核结果均为优的人数至少为  $90 + 75 - 100 = 65$  人。

所以，两年考核结果均为优的人数至少为65人。

试问我写作 图像生成 AI 搜索 AI 阅读 学术搜索 解题答疑 音乐生成 更多

发消息，输入 @ 或 / 选择技能

数据来源：豆包官网，广发证券发展研究中心

图 25: Kimi大模型回答数学计算问题的结果

有100名员工去年和今年均参加考核，考核结果分为优、良、中、差四个等次。今年考核结果为优的人数是去年的1.2倍，今年考核结果为良及以下的人员占比比去年低15个百分点。问两年考核结果均为优的人数至少为多少人？

设去年考核结果为优的人数为  $x$ ，则今年考核结果为优的人数为  $1.2x$ 。去年考核结果为良及以下的人数为  $100 - x$ ，今年考核结果为良及以下的人数为  $100 - 1.2x$ 。

根据题意，今年考核结果为良及以下的人员占比比去年低15个百分点，可以得到以下方程：

$$\frac{100 - 1.2x}{100} = \frac{100 - x}{100} - 0.15$$

化简方程，得到：

$$100 - 1.2x = 100 - x - 15$$

$$100 - 1.2x = 85 - x$$

$$15 = 0.2x$$

$$x = 75$$

所以，去年考核结果为优的人数为75人，今年考核结果为优的人数为  $1.2 \times 75 = 90$  人。

设两年考核结果均为优的人数为  $y$ ，则去年考核结果为优但今年考核结果不为优的人数为  $75 - y$ ，今年考核结果为优但去年考核结果不为优的人数为  $90 - y$ 。因此，去年考核结果为良及以下的人数为  $25 + (75 - y) = 100 - y$ ，今年考核结果为良及以下的人数为  $10 + (90 - y) = 100 - y$ 。

为了使两年考核结果均为优的人数最少，我们需要使去年考核结果为优但今年考核结果不为优的人数和今年考核结果为优但去年考核结果不为优的人数尽可能多。由于去年考核结果为优的人数为75人，今年考核结果为优的人数为90人，所以两年考核结果均为优的人数至少为：

$$y = 75 + 90 - 100 = 65$$

因此，两年考核结果均为优的人数至少为 **65** 人。

输入“/”使用 Kimi 探索版，解决复杂的搜索问题

联网搜索

内容由 AI 大模型生成，请仔细甄别

数据来源：月之暗面官网，广发证券发展研究中心

图 26: 通义千问大模型回答数学计算问题的结果

有100名员工去年和今年均参加考核，考核结果分为优、良、中、差四个等级。今年考核结果为优的人数是去年的1.2倍，今年考核结果为良及以下的人员占比比去年低15个百分点。问两年考核结果均为优的人数至少为多少人？

为了求解这个问题，我们可以设去年考核结果为优的人数为  $x$ ，那么今年考核结果为优的人数就是  $1.2x$ 。因为员工总数是固定的100人，所以我们可以根据良及以下的人员占比变化来建立方程。

假设去年良及以下的人员占比为  $y\%$ ，那么今年良及以下的人员占比为  $y\% - 15\%$ 。

去年良及以下的人数为  $100 \times y\% = y$ （这里将百分比直接表示成数值，例如45%就表示为45），因此去年优的人数为  $100 - y$ 。

今年良及以下的人数为  $100 \times (y\% - 15\%) = y - 15$ ，因此今年优的人数为  $100 - (y - 15) = 115 - y$ 。

由题意知今年优的人数是去年的1.2倍，所以我们有等式：  
 $1.2(100 - y) = 115 - y$

解这个等式，我们得到：  
 $120 - 1.2y = 115 - y$   
 $120 - 115 = 1.2y - y$   
 $5 = 0.2y$   
 $y = 25$

这意味着去年良及以下的人员占比为25%，去年考核结果为优的人数占75%，即75人。今年优的人数是去年的1.2倍，即  $1.2 \times 75 = 90$  人。

现在我们要找的是两年均为优的人数至少有多少。考虑到最极端的情况，即尽可能多的去年不是优的人在今天转成了优，那么去年75个优的人中，如果有尽可能多的人保持了优，那这部分人就是两年均为优的人数。

由于今年总共有90人被评为优，而去年只有75人被评为优，这意味着去年所有评为优的人都必须在今年继续保持优的状态，才能满足今年有90人被评为优这一条件。因此，两年考核结果均为优的人数至少为75人。这是在假定去年所有非优等级的人在今天都提升到了优等级的情况下得出的最小值。

代码模式 深度搜索 PPT制作 指令中心

千事不决问通义

服务生成的所有内容均由人工智能模型生成，其生成内容的准确性和完整性无法保证，不代表我们的态度或观点

数据来源：通义千问官网，广发证券发展研究中心

### (三) 针对程序开发场景，DeepSeek 生成内容更符合开发者要求

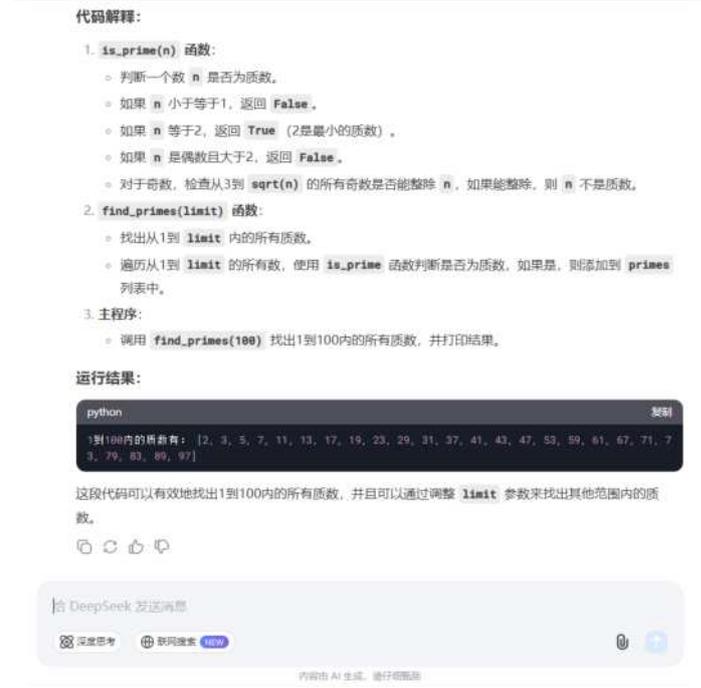
4 个大模型对于简单算法的代码生成能力接近。问题 (5) 是考察大模型代码生成能力。从测试的结果来看，4 个大模型代码生成的结果较为接近。在结果呈现方面，DeepSeek 和通义千问大模型会对每一段代码进行注释，并且在最后还会解释算法实现的原理。豆包大模型和 Kimi 大模型在生成代码的最后也对算法原理进行了解释，其中豆包大模型的算法解释较为详细。

图 27: DeepSeek 大模型代码生成的结果



数据来源：DeepSeek 官网，广发证券发展研究中心

图 28: DeepSeek 大模型代码生成的结果



数据来源：DeepSeek 官网，广发证券发展研究中心

图 29: 豆包大模型代码生成的结果

请用Python语言写一段代码，找出从1到100内所有的质数。

以下是使用 Python 语言来找出 1 到 100 内所有质数的代码示例：

收起 ^

```
python
if num < 2:
    return False
for i in range(2, int(num ** 0.5) + 1):
    if num % i == 0:
        return False
    return True

# 遍历1到100的数字，找出其中的质数
prime_numbers = []
for num in range(1, 101):
    if is_prime(num):
        prime_numbers.append(num)

print(prime_numbers)
```

在上述代码中：

首先定义了一个名为 `is_prime` 的函数，用于判断一个数是否为质数。函数接收一个参数 `num`，如果 `num` 小于 2，按照质数的定义，直接返回 `False`，因为质数是大于 1 的自然数。然后通过一个 `for` 循环，从 2 开始到该数的平方根取整再加 1（这是因为如果一个数不是质数，那么它一定可以分解成两个因数，其中至少一个因数小于等于它的平方根，通过判断到平方根位置就能确定是否为质数了），依次判断 `num` 能否被这些数整除，如果能被整除，说明不是质数，返回 `False`，如果遍历完都没有能整除的情况，就返回 `True`，表示该数是质数。

接着使用一个 `for` 循环遍历从 1 到 100 的所有数字，对于每个数字调用 `is_prime` 函数进行判断，如果是质数，就将其添加到 `prime_numbers` 列表中。

最后打印出 `prime_numbers` 列表，里面存储的就是 1 到 100 内的所有质数。

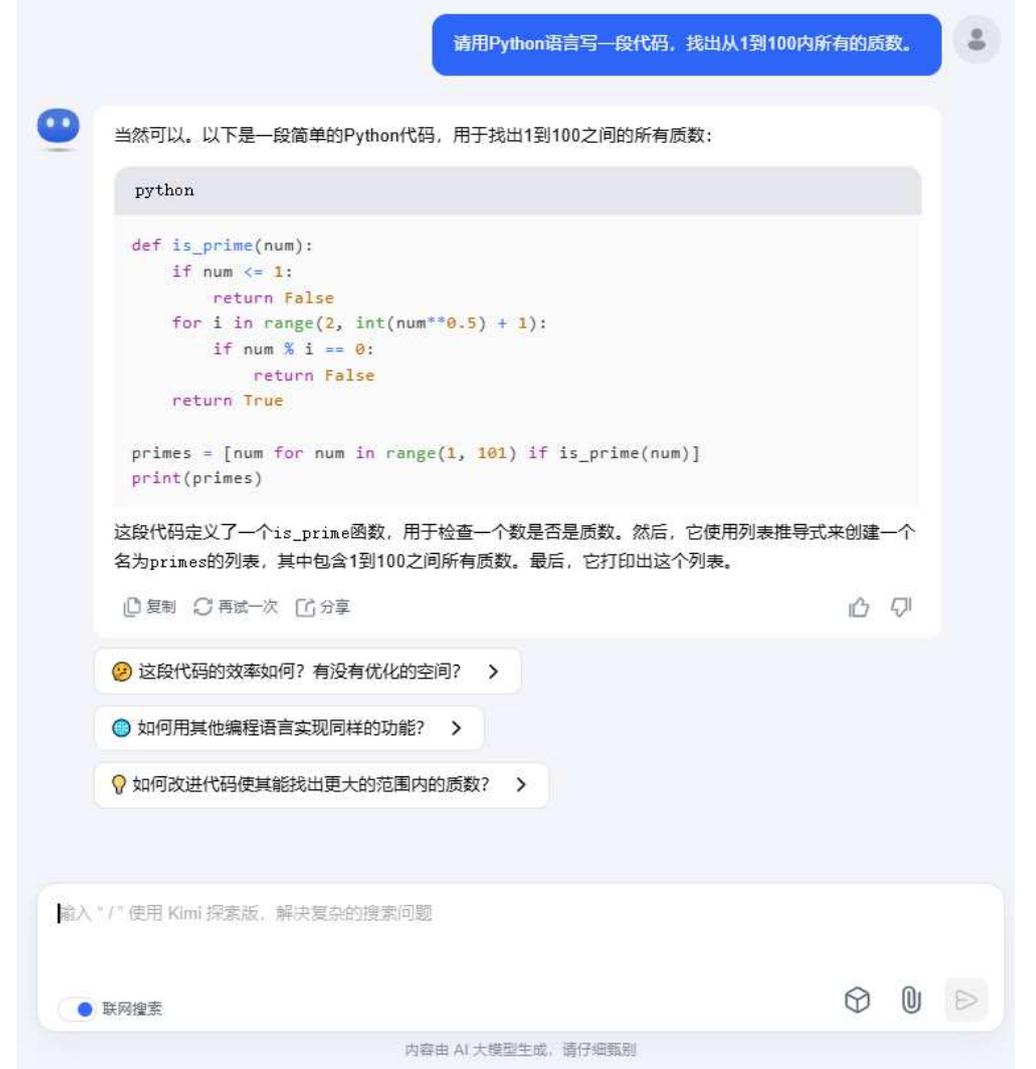
📝 帮我写作   🖼️ 图像生成   🔍 AI 搜索   📖 AI 阅读   🔍 学术搜索   🧠 解题答疑   🎵 音乐生成   🍷 更多

发消息，输入 @ 或 / 选择技能

🗨️ 📄 ✂️ 🗑️ ↩️

数据来源：豆包官网，广发证券发展研究中心

图 30: Kimi大模型代码生成的结果



数据来源：月之暗面官网，广发证券发展研究中心

图 31: 通义千问大模型代码生成的结果



数据来源：通义千问官网，广发证券发展研究中心

图 32: 通义千问大模型代码生成的结果



数据来源：通义千问官网，广发证券发展研究中心

针对特定应用场景的程序开发，DeepSeek和豆包大模型详细解释和引导每一步程序开发的原理和细节。问题（6）是考察大模型针对实际业务场景的编程能力。在实现最终功能的过程中，DeepSeek和豆包大模型会对实现程序开发每一个步骤进行详细的引导和解释，更加便于用户的理解。而Kimi和通义千问仅给出了实现程序的代码，对于每一步实现的过程以及原理解释较少。

图 33: DeepSeek大模型实现网页开发功能的程序开发过程

图 34: DeepSeek大模型实现网页开发功能的程序开发过程



数据来源：DeepSeek 官网，广发证券发展研究中心

数据来源：DeepSeek 官网，广发证券发展研究中心

图 35: DeepSeek大模型实现网页开发功能的程序开发过程

图 36: DeepSeek大模型实现网页开发功能的程序开发过程



数据来源：DeepSeek 官网，广发证券发展研究中心

数据来源：DeepSeek 官网，广发证券发展研究中心

图 37: 豆包大模型实现网页开发功能的程序开发过程



数据来源：豆包官网，广发证券发展研究中心

图 39: 豆包大模型实现网页开发功能的程序开发过程



数据来源：豆包官网，广发证券发展研究中心

图 38: 豆包大模型实现网页开发功能的程序开发过程



数据来源：豆包官网，广发证券发展研究中心

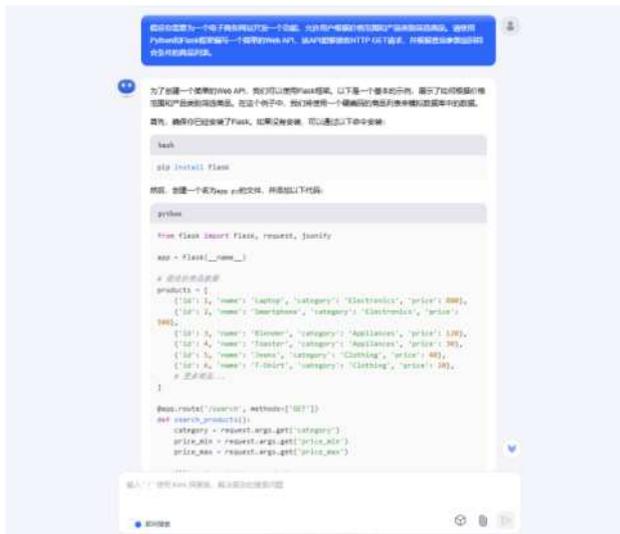
图 40: 豆包大模型实现网页开发功能的程序开发过程



数据来源：豆包官网，广发证券发展研究中心

图 41: Kimi大模型实现网页开发功能的程序开发过程

图 42: Kimi大模型实现网页开发功能的程序开发过程



数据来源: 月之暗面官网, 广发证券发展研究中心

数据来源: 月之暗面官网, 广发证券发展研究中心

图 43: 通义千问大模型实现网页开发功能的程序开发过程

图 44: 通义千问大模型实现网页开发功能的程序开发过程



数据来源: 通义千问官网, 广发证券发展研究中心

数据来源: 通义千问官网, 广发证券发展研究中心

(四) DeepSeek 生成文本的质量符合行业平均水平

在文字创作功能上, 通义千问生成的文本长度较长。问题(7)是考察大模型文字生成和描述能力。针对季节描述的文本内容方面, 4个大模型都能按照要求描述自然景观、气候特点以及人们的生活状态等。在生成文本长度方面, Kimi大模型生成文字的内容长度不超过600字; DeepSeek和豆包大模型的内容长度接近700字; 而通义千问大模型生成内容长度超过1000字。

图 45: DeepSeek大模型文字创作生成结果

请选择一个你最喜欢的季节，详细描述该季节的自然景观、气候特点以及人们的生活状态。

我最喜欢的季节是春天。春天是一个充满生机与希望的季节，它以其独特的自然景观、温和的气候和人们的生活状态深深吸引着我。

### 自然景观

春天的自然景观如同一幅生机勃勃的画卷。大地从冬日的沉睡中苏醒，万物开始复苏。树木抽出嫩绿的新芽，草地上点缀着各色野花，仿佛大自然在为新的一年铺上五彩斑斓的地毯。樱花、桃花、梨花竞相开放，花瓣随风飘落，宛如一场浪漫的花雨。河流解冻，溪水潺潺，鸟儿在枝头欢快地歌唱，仿佛在庆祝春天的到来。田野里，农民开始播种，一片片新绿的麦苗在阳光下茁壮成长，预示着丰收的希望。

### 气候特点

春天的气候温和宜人，既没有冬天的寒冷，也没有夏天的炎热。阳光温暖而不刺眼，空气中弥漫着泥土和花草的芬芳。春风轻拂，带来阵阵暖意，让人感到舒适和愉悦。春天的雨水滋润着大地，偶尔的春雨如丝如缕，给万物带来生机和活力。雨后的空气中充满了清新的气息，令人心旷神怡。

### 人们的生活状态

春天是一个充满活力和希望的季节，人们的生活状态也因此充满了积极向上的气息。农民们忙着播种，脸上洋溢着对丰收的期待。城市里的人们也开始脱下厚重的冬装，换上轻便的春装，走出家门，享受春天的美好。公园里，人们散步、跑步、骑自行车，孩子们在草地上玩耍，老人们在长椅上晒太阳，整个社会都沉浸在一种和谐与愉悦的氛围中。春天也是一个适合旅行的季节，人们纷纷计划春游，去郊外踏青、赏花、野餐，感受大自然的魔力。

### 总结

春天是一个充满生机与希望的季节，它以其绚丽的自然景观、温和的气候和人们的生活状态，给人们带来了无尽的享受和感动。无论是新绿的嫩芽，还是盛开的花朵，都让人感受到生命的力量和自然的恩赐。春天，是我心中最美的季节。

🔄 🔄 🔄 🔄

给 DeepSeek 发送消息

深度思考

联网搜索 NEW

🗣️ 📤

内容由 AI 生成，请仔细甄别

数据来源: DeepSeek 官网, 广发证券发展研究中心

图 46: 豆包大模型文字创作生成结果



数据来源：豆包官网，广发证券发展研究中心

图 47: Kimi大模型文字创作生成结果



数据来源：月之暗面官网，广发证券发展研究中心

图 48: 通义千问大模型文字创作生成结果



数据来源：通义千问官网，广发证券发展研究中心

图 49: 通义千问大模型文字创作生成结果



数据来源：通义千问官网，广发证券发展研究中心

在文字创作功能上，通义千问生成的文本长度较长。问题（8）是考察大模型在具体应用场景中的文字生成能力。针对求职场景的文本内容方面，4 个大模型都能按照要求描述教育背景、相关经验和个人优势等，并且针对未提供的信息留有空白给用户自行修改。虽然，通义千问生成的文本内容长度明显高于其他 3 个大模型，但在求职场景中，过于冗长的描述不太符合职场书信往来内容简洁的要求。

图 50: DeepSeek 大模型在求职场景的文本生成结果



数据来源：DeepSeek 官网，广发证券发展研究中心

图 51: DeepSeek 大模型在求职场景的文本生成结果



数据来源：DeepSeek 官网，广发证券发展研究中心

图 52: 豆包大模型在求职场景的文本生成结果



数据来源: 豆包官网, 广发证券发展研究中心

图 53: Kimi大模型在求职场景的文本生成结果



数据来源: 月之暗面官网, 广发证券发展研究中心

图 54: Kimi大模型在求职场景的文本生成结果



数据来源: 月之暗面官网, 广发证券发展研究中心

图 55: 通义千问大模型在求职场景的文本生成结果



图 56: 通义千问大模型在求职场景的文本生成结果



数据来源: 通义千问官网, 广发证券发展研究中心

数据来源: 通义千问官网, 广发证券发展研究中心

通过对比DeepSeek、豆包、Kimi和通义千问这四个大模型回答逻辑、数学、代码和文本等问题的答案。我们发现，DeepSeek总体能力与其他大模型相当，但在逻辑推理和代码生成领域具有自身特点。例如，在密文解密任务中，DeepSeek是唯一给出正确答案的大模型；而在代码生成的任务中，DeepSeek给出的代码注释、算法原理解释以及开发流程的指引是最为全面的。在文本生成和数学计算能力方面，DeepSeek并未展现出明显优于其他大模型之处。

### 三、数据与算法仍有潜力可待挖掘，算力依旧不可或缺

#### （一）数据与算法仍有潜力可待挖掘

在算法方面，DeepSeek-V3的主要亮点包含：

- (1) 多专家混合架构 (MoE) 优化: 采用DeepSeekMoE 架构, 使用更细粒度的专家和共享专家, 通过辅助无损失负载均衡策略, 提高计算效率。
- (2) 多头潜在注意力机制 (MLA): 运用MLA架构, 对注意力键值进行低秩联合压缩, 减少推理时的键值缓存, 同时对查询也进行低秩压缩, 降低训练时的激活内存, 在保持性能的同时提高了推理和训练效率。
- (3) 多令牌预测 (MTP) 目标: 设置 MTP目标, 扩展预测范围到多个未来令牌, 增强模型的预测能力。
- (4) 高效的训练框架设计: 通过DualPipe算法实现高效流水线并行, 通过重叠计算和通信阶段提高训练效率。同时充分利用 InfiniBand (IB) 和 NVLink 带宽, 优化内存占用。
- (5) FP8混合精度训练框架: 提出基于FP8数据格式的细粒度混合精度训练框架, 通过分组量化、提高累积精度等策略, 在保证训练稳定性的同时提高训练效率, 首次在超大规模模型上验证了FP8训练的有效性。

在数据方面，DeepSeek-V3使用了14.8T高质量和多样化的训练数据。根据DeepSeek-V3技术报告，相比DeepSeek-V2，DeepSeek-V3提升了数学和编程样本的比例，并扩大了中英语言之外的多语言覆盖范围，优化数据处理流程以减少冗余并确保语料库的多样性和完整性，同时，采用了文档打包方法以维护数据完整性。

**DeepSeek-V3通过数据与算法层面的优化，大幅提升算力利用效率，实现了协同效应。**在大规模MoE模型的训练中，DeepSeek-V3采用了高效的负载均衡策略、FP8混合精度训练框架以及通信优化等一系列优化措施，显著降低了训练成本，以及通过优化MoE专家调度、引入冗余专家策略、以及通过长上下文蒸馏提升推理性能。证明了模型效果不仅依赖于算力投入，即使在硬件资源有限的情况下，依托数据与算法层面的优化创新，仍然可以高效利用算力，实现较好的模型效果。

#### （二）算力是 AI 大模型的发展动力，依旧不可或缺

DeepSeek-V3 通过高质量数据以及算法优化取得了较好的模型效果,但算力才是让这些数据和算法发挥作用的资源前提:

1. 从训练成本来看, 根据DeepSeek-V3技术报告, 整个预训练阶段总计花费 266.4万GPU小时。外加扩展上下文长度所需的11.9万GPU小时和后训练耗费的5000

GPU小时，总训练时间仅需278.8万GPU小时，假设H800 GPU的租用价格是每小时2美元，整体训练成本约为557.6万美元。然而，这一数字仅涵盖正式训练阶段的成本，未包含模型架构设计、算法优化或数据处理等前期研究及消融实验的费用。

2. 从部署门槛来看，一个完整的DeepSeek-V3部署单元便需要消耗数百块H800，对于硬件设施的要求较高。

(1) Prefilling阶段：最小部署单元需要4个节点，每个节点配备8个GPU，共需32个GPU。使用了多种并行技术，同时通过冗余专家策略实现负载均衡，进一步提高计算效率。在每块GPU上，除了原有的8个专家，还额外增加了1个冗余专家。

(2) Decoding阶段：最小部署单元需要40个节点，每个节点配备8个GPU，共需320个GPU。每个GPU仅承担一个专家的任务，同时有64个GPU承担冗余专家的部署任务。

我们认为，DeepSeek-V3算力成本降低的原因有两点。第一，DeepSeek-V3采用的DeepSeekMoE是通过参考了各类训练方法后优化得到的，避开了行业内AI大模型训练过程中的各类问题。第二，DeepSeek-V3采用的MLA架构可以降低推理过程中的kv缓存开销，其训练方法在特定方向的选择也使得其算力成本有所降低。

算力依然是推动大模型发展的核心驱动力。与其他同规模模型（如Llama 3 405B 30.84M）的训练成本相比，DeepSeek-V3通过技术创新和资源优化，大幅降低了成本，展现了算力的高效性。在该技术路线得到充分验证后，有望驱动相关AI应用的快速发展，应用推理驱动算力需求增长的因素也有望得到增强。尤其在应用中，推理过程涉及到对大量实时数据的快速处理和决策，仍然需要强大的算力支持。

以大规模通用模型为基础，聚焦特定领域突出自身特点的模型应用开发或许是下一阶段商业化探索方向。DeepSeek-V3为未来大模型技术的发展提供了重要启发，未来或将从依赖大规模通用模型转向发展一些更具特色、成本更低的模型，这些模型可能更适合具体应用场景，随着这些特色模型不断成熟，AI商业化的边际成本有望降低，将迎来更广阔的应用前景。

## 四、风险提示

### （一）AI大模型技术追赶和竞争加剧的风险

DeepSeek采用的MLA和MoE架构的训练方法为其他大模型开发提供了技术参考。国内AI大模型或集合自身特点，加强对数据利用和模型优化的效率。技术的追赶或加剧行业竞争态势。

### （二）由于下游生态和使用环境的差异，商业化成功有不确定性

AI大模型技术发展时间较短，目前仍处于商业化模式的探索阶段。由于下游生态和使用环境的差异，商业化成功有不确定性。

### （三）AI生成内容存在知识产权纠纷的风险

AI大模型生成的文字内容版权归属，目前我国并未出台相关法律法规进行明确清晰的界定。AI大模型生成内容若涉及版权纠纷，或给研发团队带来一定法律风险和不确定性，从而影响AI应用的开发和推广。

## 广发计算机行业研究小组

- 刘雪峰：首席分析师，东南大学工学士，中国人民大学经济学硕士，1997年起先后在数家IT行业跨国公司从事技术、运营与全球项目管理。2010年就职于招商证券研究发展中心负责计算机组行业研究工作，2014年加入广发证券发展研究中心。
- 吴祖鹏：资深分析师，中南大学材料工程学士，复旦大学经济学硕士，曾先后任职于华泰证券、华西证券，2021年加入广发证券发展研究中心。
- 李婉云：资深分析师，西南财经大学金融学硕士，2022年加入广发证券发展研究中心。
- 周源：资深分析师，慕尼黑工业大学硕士，2021年加入广发证券，曾任职于TUMCREATE自动驾驶科技公司，负责大数据相关工作。
- 王钰翔：研究员，哥伦比亚大学运筹学硕士，2024年加入广发证券发展研究中心。
- 戴亚敏：研究员，北京大学金融硕士，2024年加入广发证券发展研究中心。

## 广发证券—行业投资评级说明

- 买入：预期未来12个月内，股价表现强于大盘10%以上。
- 持有：预期未来12个月内，股价相对大盘的变动幅度介于-10%~+10%。
- 卖出：预期未来12个月内，股价表现弱于大盘10%以上。

## 广发证券—公司投资评级说明

- 买入：预期未来12个月内，股价表现强于大盘15%以上。
- 增持：预期未来12个月内，股价表现强于大盘5%-15%。
- 持有：预期未来12个月内，股价相对大盘的变动幅度介于-5%~+5%。
- 卖出：预期未来12个月内，股价表现弱于大盘5%以上。

## 联系我们

	广州市	深圳市	北京市	上海市	香港
地址	广州市天河区马场路 26号广发证券大厦 47楼	深圳市福田区益田路 6001号太平金融大厦 31层	北京市西城区月坛北 街2号月坛大厦18 层	上海市浦东新区南泉 北路429号泰康保险 大厦37楼	香港湾仔骆克道81 号广发大厦27楼
邮政编码	510627	518026	100045	200120	-
客服邮箱	gfzqyf@gf.com.cn				

## 法律主体声明

本报告由广发证券股份有限公司或其关联机构制作，广发证券股份有限公司及其关联机构以下统称为“广发证券”。本报告的分销依据不同国家、地区的法律、法规和监管要求由广发证券于该国家或地区的具有相关合法合规经营资质的子公司/经营机构完成。

广发证券股份有限公司具备中国证监会批复的证券投资咨询业务资格，接受中国证监会监管，负责本报告于中国（港澳台地区除外）的分销。

广发证券（香港）经纪有限公司具备香港证监会批复的就证券提供意见（4号牌照）的牌照，接受香港证监会监管，负责本报告于中国香港地区的分销。

本报告署名研究人员所持中国证券业协会注册分析师资质信息和香港证监会批复的牌照信息已于署名研究人员姓名处披露。

## 重要声明

免责声明：

1. 本资料来源于网络公开渠道，版权归属版权方；
2. 本资料仅限会员学习使用，如他用请联系版权方；
3. 会员费用作为信息收集整理及运营之必须费用；
4. 如侵犯您的合法权益，请联系客服微信将及时删除



## 行业报告资源群

1. 进群福利：进群即领万份行业研究、管理方案及其他学习资源，直接打包下载
2. 每日分享：6份行研精选报告、3个行业主题
3. 报告查找：群里直接咨询，免费协助查找
4. 严禁广告：仅限行业报告交流，禁止一切无关信息



微信扫码，长期有效

## 知识星球 行业与管理资源

专业知识社群：每月分享8000+份行业研究报告、商业计划、市场研究、企业运营及咨询管理方案等，涵盖科技、金融、教育、互联网、房地产、生物制药、医疗健康等；已成为投资、产业研究、企业运营、价值传播等工作助手。



微信扫码，行研无忧

广发证券股份有限公司及其关联机构可能与本报告中提及的公司寻求或正在建立业务关系，因此，投资者应当考虑广发证券股份有限公司及其关联机构因可能存在的潜在利益冲突而对本报告的独立性产生影响。投资者不应仅依据本报告内容作出任何投资决策。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者口头承诺均为无效。

本报告署名研究人员、联系人（以下均简称“研究人员”）针对本报告中相关公司或证券的研究分析内容，在此声明：（1）本报告的全部分析结论、研究观点均精确反映研究人员于本报告发出当日的关于相关公司或证券的所有个人观点，并不代表广发证券的立场；（2）研究人员的部分或全部的报酬无论在过去、现在还是将来均不会与本报告所述特定分析结论、研究观点具有直接或间接的联系。

研究人员制作本报告的报酬标准依据研究质量、客户评价、工作量等多种因素确定，其影响因素亦包括广发证券的整体经营收入，该等经营收入部分来源于广发证券的投资银行类业务。

本报告仅面向经广发证券授权使用的客户/特定合作机构发送，不对外公开发布，只有接收人才可以使用，且对于接收人而言具有保密义务。广发证券并不因相关人员通过其他途径收到或阅读本报告而视其为广发证券的客户。在特定国家或地区传播或者发布本报告可能违反当地法律，广发证券并未采取任何行动以允许于该等国家或地区传播或者分销本报告。

本报告所提及证券可能不被允许在某些国家或地区内出售。请注意，投资涉及风险，证券价格可能会波动，因此投资回报可能会有所变化，过去的业绩并不保证未来的表现。本报告的内容、观点或建议并未考虑任何个别客户的具体投资目标、财务状况和特殊需求，不应被视为对特定客户关于特定证券或金融工具的投资建议。本报告发送给某客户是基于该客户被认为有能力独立评估投资风险、独立行使投资决策并独立承担相应风险。

本报告所载资料的来源及观点的出处皆被广发证券认为可靠，但广发证券不对其准确性、完整性做出任何保证。报告内容仅供参考，报告中的信息或所表达观点不构成所涉证券买卖的出价或询价。广发证券不对因使用本报告的内容而引致的损失承担任何责任，除非法律法规有明确规定。客户不应以本报告取代其独立判断或仅根据本报告做出决策，如有需要，应先咨询专业意见。

广发证券可发出其它与本报告所载信息不一致及有不同结论的报告。本报告反映研究人员的不同观点、见解及分析方法，并不代表广发证券的立场。广发证券的销售人员、交易员或其他专业人士可能以书面或口头形式，向其客户或自营交易部门提供与本报告观点相反的市场评论或交易策略，广发证券的自营交易部门亦可能会有与本报告观点不一致，甚至相反的投资策略。报告所载资料、意见及推测仅反映研究人员于发出本报告当日的判断，可随时更改且无需另行通告。广发证券或其证券研究报告业务的相关董事、高级职员、分析师和员工可能拥有本报告所提及证券的权益。在阅读本报告时，收件人应了解相关的权益披露（若有）。

本研究报告可能包括和/或描述/呈列期货合约价格的事实历史信息（“信息”）。请注意此信息仅供用作组成我们的研究方法/分析中的部分论点/依据/证据，以支持我们对所述相关行业/公司的观点的结论。在任何情况下，它并不（明示或暗示）与香港证监会第5类受规管活动（就期货合约提供意见）有关联或构成此活动。

## 权益披露

(1)广发证券（香港）跟本研究报告所述公司在过去12个月内并没有任何投资银行业务的关系。

## 版权声明

未经广发证券事先书面许可，任何机构或个人不得以任何形式翻版、复制、刊登、转载和引用，否则由此造成的一切不良后果及法律责任由私自翻版、复制、刊登、转载和引用者承担。