

# 实验四 基于大语言模型的复杂任务处理

## （一）目的及要求

- （1）熟悉大语言模型的工作原理
- （2）学会使用已有的大语言模型解决下游任务；
- （3）掌握不同的提示词策略

## （二）实验内容

（1）采用不同提示词策略，对比大模型结果的差异，并给出实验结果。对比的策略有：zero-shot、思维链、few-shot。

（2）测试数据集采用“数学推理数据集”（GSM8K）。该数据集的介绍、使用和已有大模型的实验结果都可以从这里查看：  
<https://paperswithcode.com/sota/arithmetic-reasoning-on-gsm8k>

（3）调用大模型的 API。什么大模型都可以，比如文心一言、通义千文等都可以。对比了一下，DeepSeek 应该是最便宜的，可以调用 DeepSeek 的 API（<https://www.deepseek.com/>），注册应该都有免费的额度，超过了字数价格也比较便宜，1 百万字需要 1 元左右。星火也有免费适用的 <https://xinghuo.xfyun.cn/sparkapi#price>。

参考论文： Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q.V. and Zhou, D., 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35, pp.24824-24837.