**大语言模型部署体验**

**一、作业内容：**

参考视频教程（附件1&2），在魔搭平台上部署并测试大语言模型（提供项目公开可访问链接），并完成报告。具体包括：

1. 登录并使用魔搭平台，注册时需要关联阿里云账号来获得免费的CPU云计算资源；
2. 通过Jupyter Notebook或者相应环境镜像进入相应的项目部署环境，根据相应模型的部署文档完成模型的部署；
3. 针对2-3个不同的模型（参考补充材料1）进行一些应用场景（比如问答，可参考补充材料2，也可自行设计问题）的测试，并开展不同模型之间的横向对比；
4. **报告需要包含：**1）完成git clone相关git的截图或部署完成的相关截图 2）问答测试结果的相关截图 3）大语言模型之间的横向对比分析；

项目公开可访问链接参考（以下参考链接不作为评优标准）：

<https://modelscope.cn/models/Donald0316/Models_Compare/summary>

https://github.com/zzy0304Raymond/Comparing-Language-Models-Chinese-Semantic-Understanding

**二、作业要求：**

1. 作业总分20分
2. 作业文档的截止提交时间6月22日23:59（第十七周-周日），提交Canvas, 迟交1天扣2分，扣至10分为止
3. 作业文档命名格式：hw4\_学号\_姓名
4. 评分细则

1）项目公开可访问链接，8分，优8分，良6分，差4分

2）报告，12分

完成git clone相关git的截图或部署完成的相关截图，3分，优3分，良2分，差1分

问答测试结果的相关截图，3分，优3分，良2分，差1分

大语言模型之间的横向对比分析，6分，优6分，良4分，差2分

**补充材料1：推荐的开源大语言模型**

1. [通义千问Qwen-7B-Chat](https://www.modelscope.cn/models/qwen/Qwen-7B-Chat/summary)
2. [智谱ChatGLM3-6B](https://www.modelscope.cn/models/ZhipuAI/chatglm3-6b/summary)
3. [百川2-7B-对话模型](https://www.modelscope.cn/models/baichuan-inc/Baichuan2-7B-Chat/summary)
4. Intel 发布的开源模型如Neural Chat等（待更新）

**补充材料2：可供参考的部分问答测试**

1. 请说出以下两句话区别在哪里？ 1、冬天：能穿多少穿多少 2、夏天：能穿多少穿多少
2. 请说出以下两句话区别在哪里？单身狗产生的原因有两个，一是谁都看不上，二是谁都看不上
3. 他知道我知道你知道他不知道吗？ 这句话里，到底谁不知道
4. 明明明明明白白白喜欢他，可她就是不说。 这句话里，明明和白白谁喜欢谁？
5. 领导：你这是什么意思？ 小明：没什么意思。意思意思。 领导：你这就不够意思了。 小明：小意思，小意思。领导：你这人真有意思。 小明：其实也没有别的意思。 领导：那我就不好意思了。 小明：是我不好意思。请问：以上“意思”分别是什么意思。