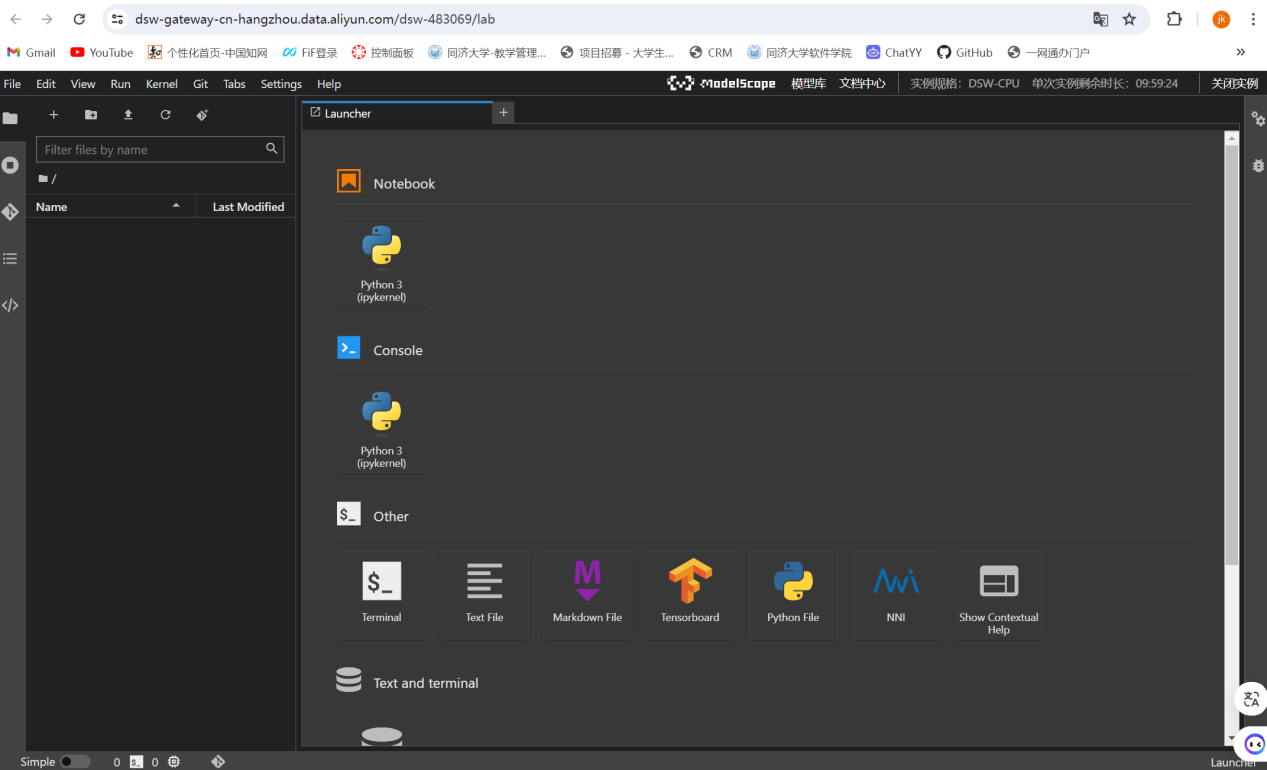
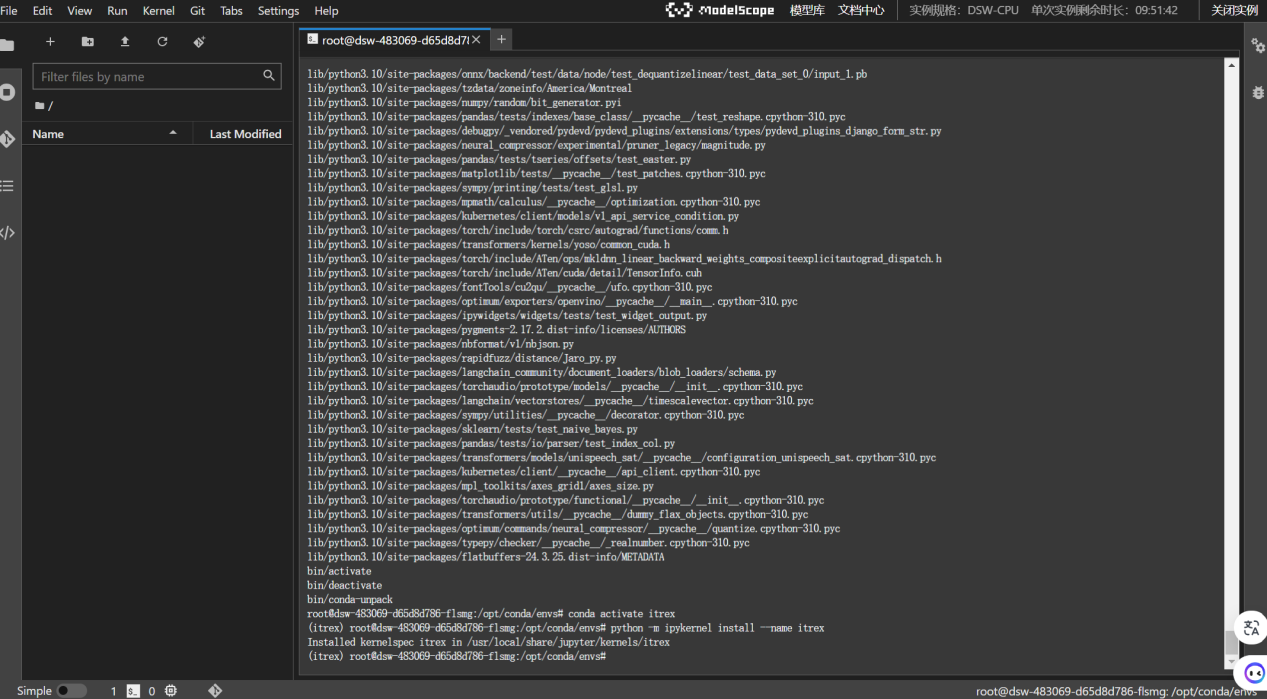
**大语言模型部署报告**

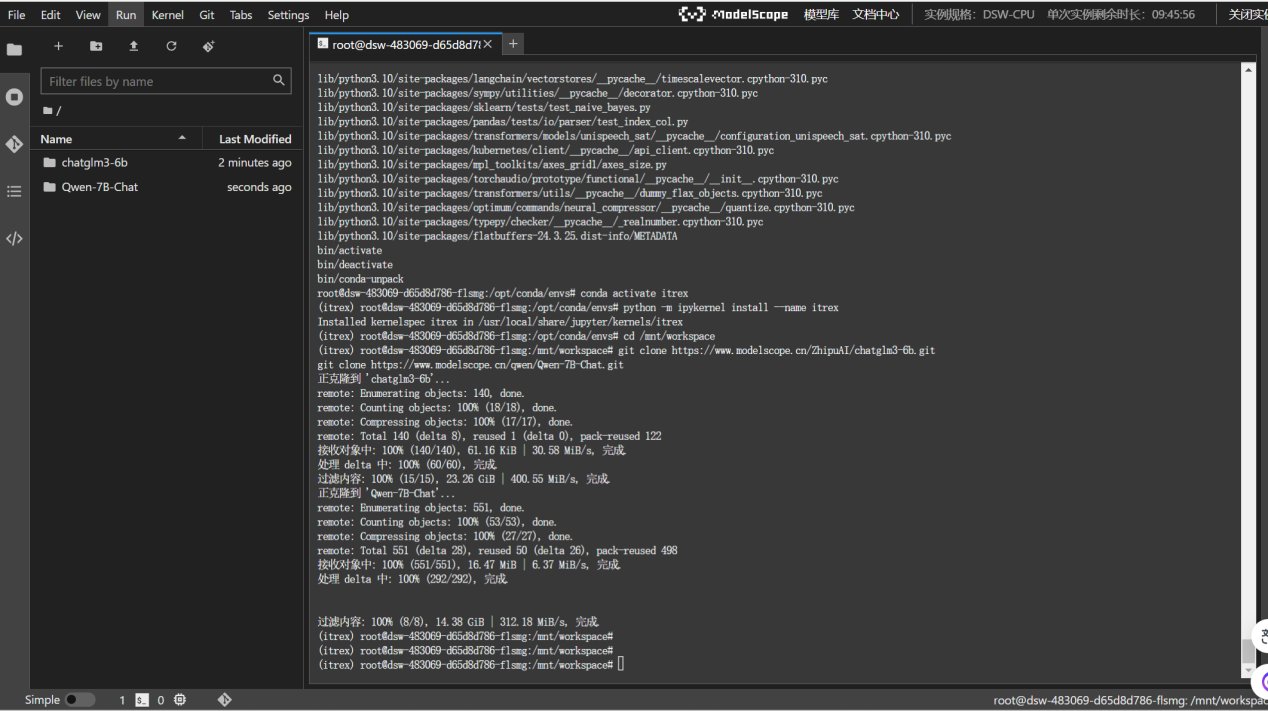
姓名：林觉凯 学号：2253744 指导老师：曾进

**1.配置和部署(此部分对应作业要求1：完成git clone相关git的截图或部署完成的相关截图)**

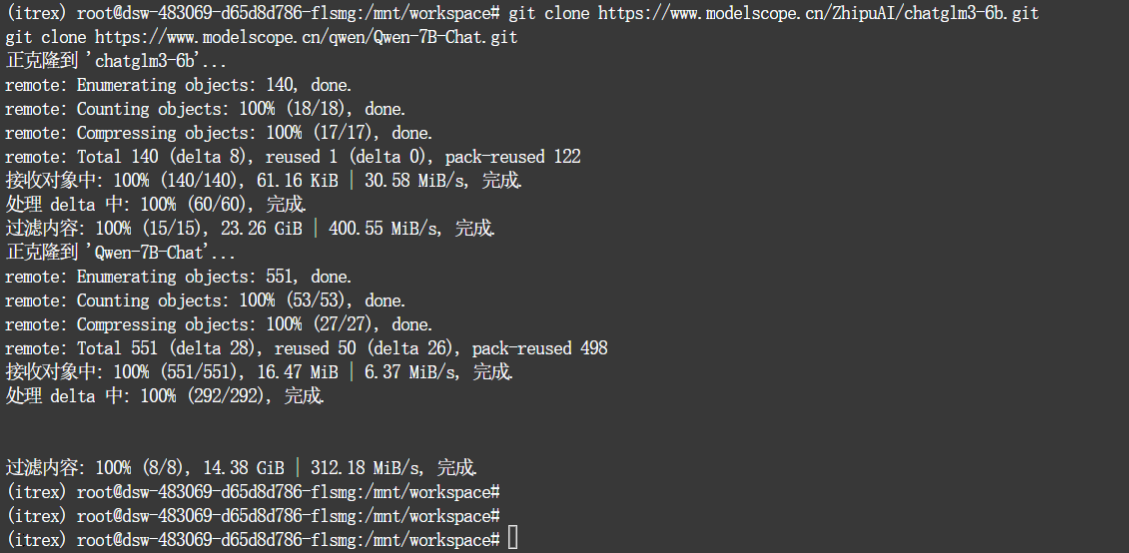
注册完成后按照指导手册的要求打开Notebook：

环境搭建的过程和完成配置成功：

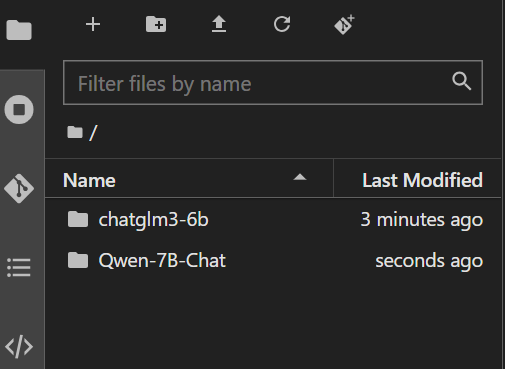




根据实验需要下载相对应的中文大模型至本地(完成git clone相关git的截图或部署完成的相关截图)：

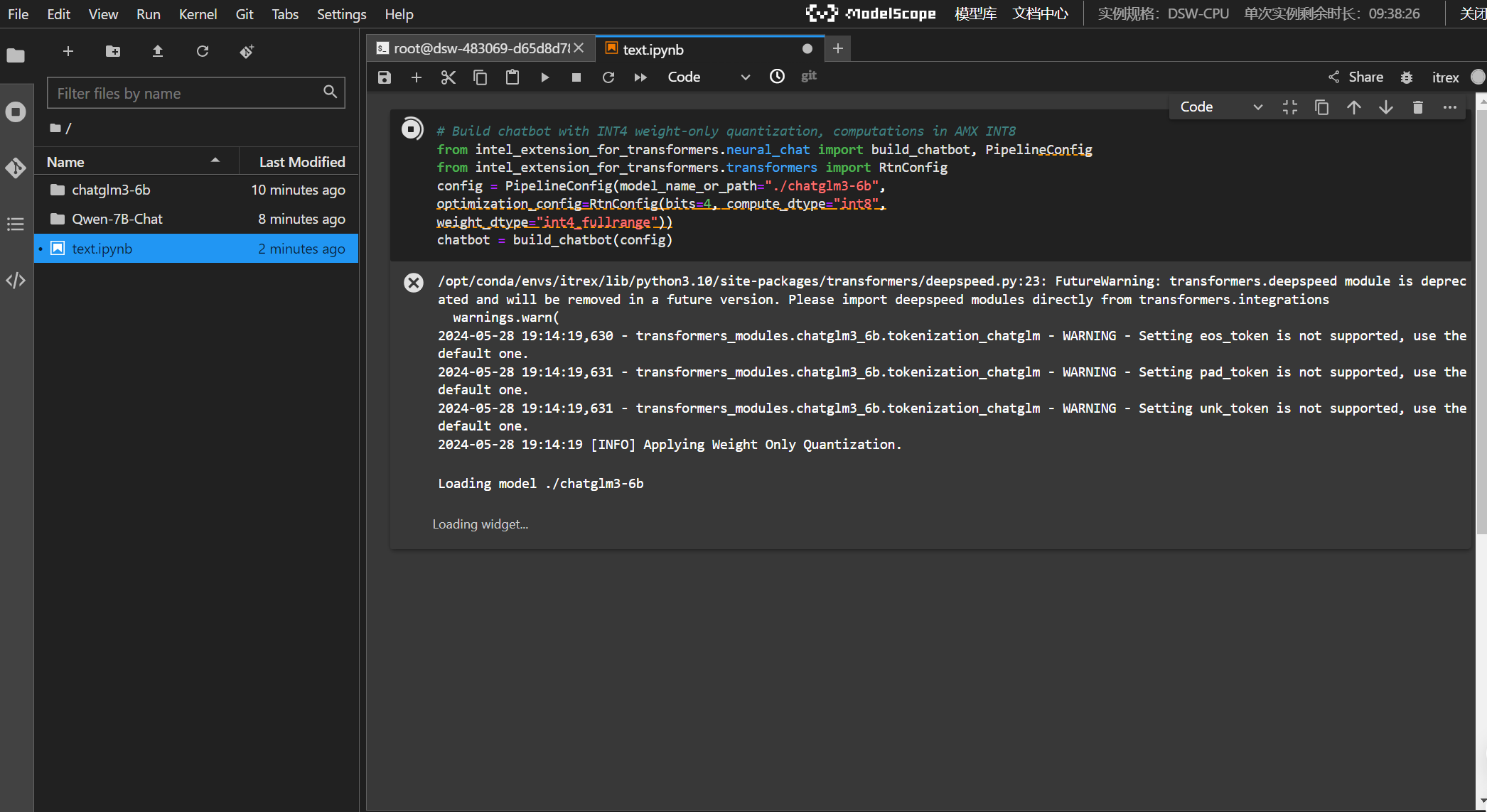
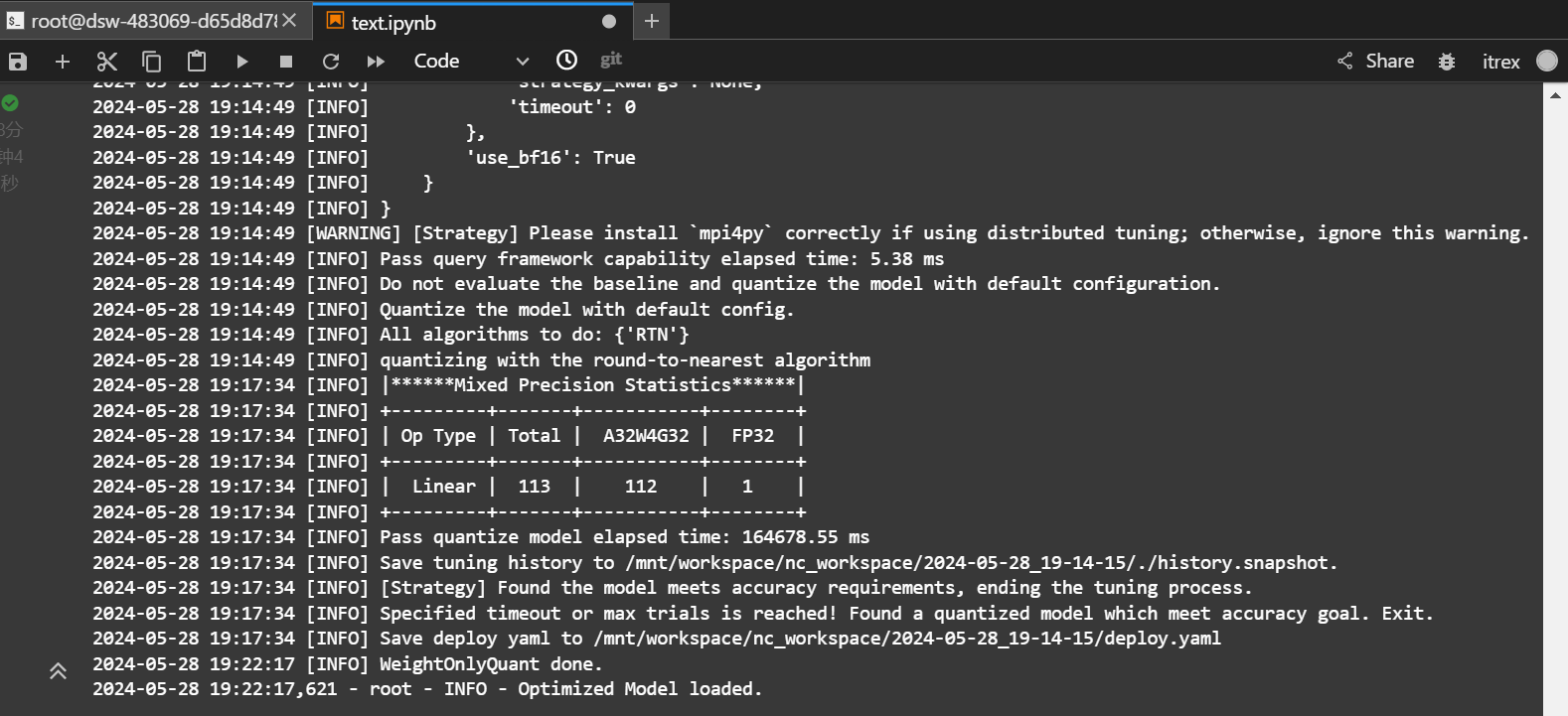


此时便可以在左侧目录下看到下载的两个大语言模型了。

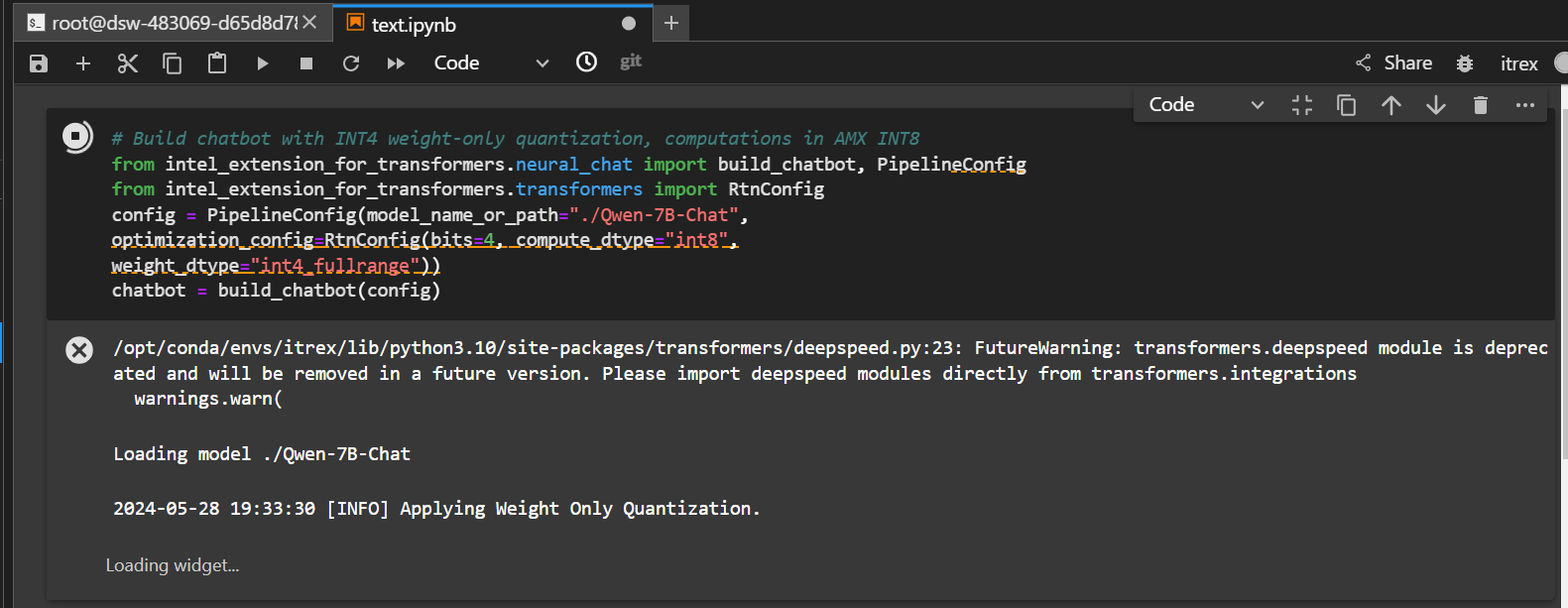
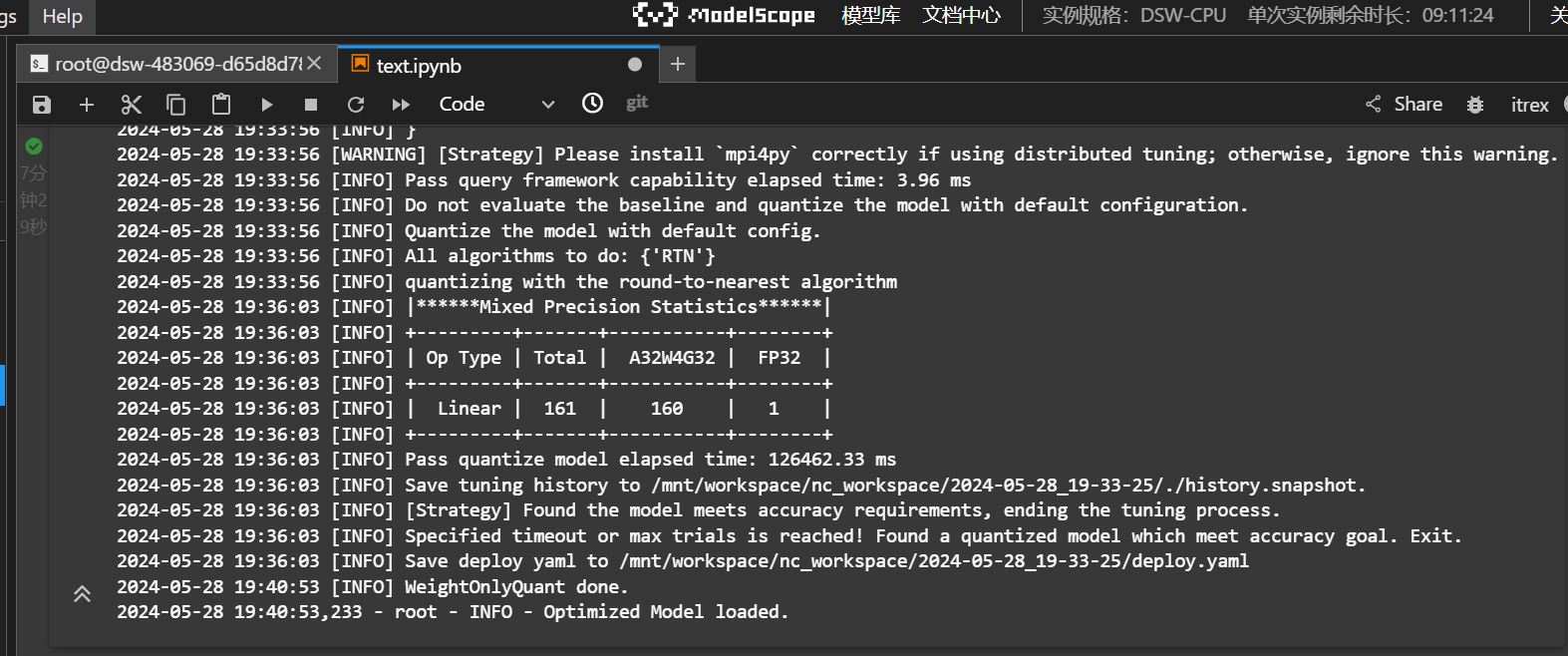


**2.对话与分析(此部分对应作业要求2.3：问答测试结果的相关截图和语言模型之间的横向对比分析)**

**2.1.量化并构建chatbot**

使用智谱ChatGLM3-6B大语言模型：

当我们想要换一个大语言模型时，将另一个大语言模型的路径复制换到相应的位置，再次量化并构建，下面就是换为通义千问Qwen-7B-Chat大语言模型的过程截图：

**2****.2.对话测试和横向分析比较**

问题一.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；

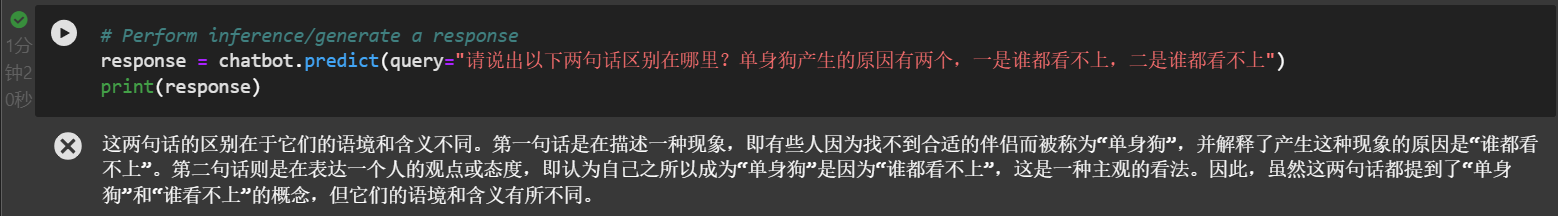
通义千问Qwen-7B-Chat回答：

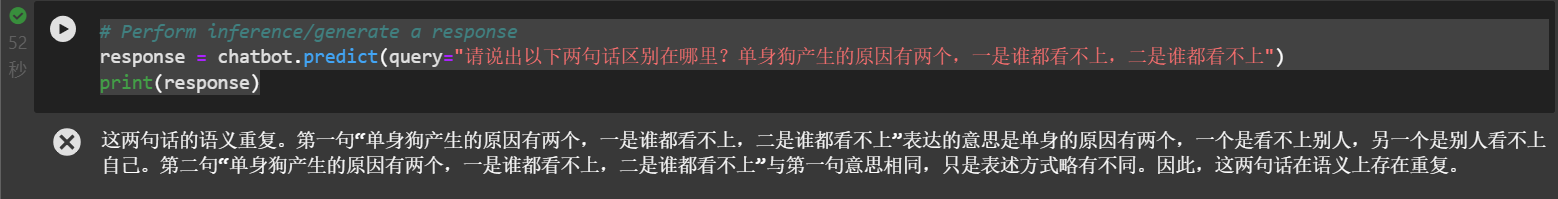
智谱ChatGLM3-6B回答：

对于第一个问题，我们人类的正常理解是前半句的意思是冬天要穿多，夏天要穿少，两句的意思相反；通义千问Qwen-7B-Chat的回答和人类的正常思维相符；智谱ChatGLM3-6B的回答虽然说两句话意思相同，但是后面的解释也与人类思维大差不差。再考虑两者回答所需时间，都是一分零2秒。所以综上所述，对于第一个问题，两种大语言模型都很好地做出了正确的回答。

问题二.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；

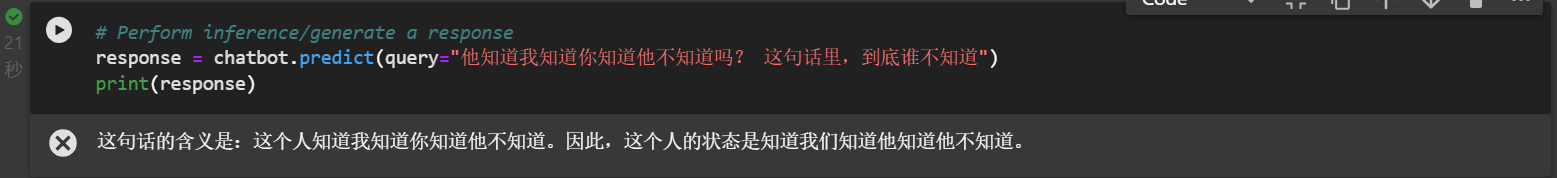
通义千问Qwen-7B-Chat回答：

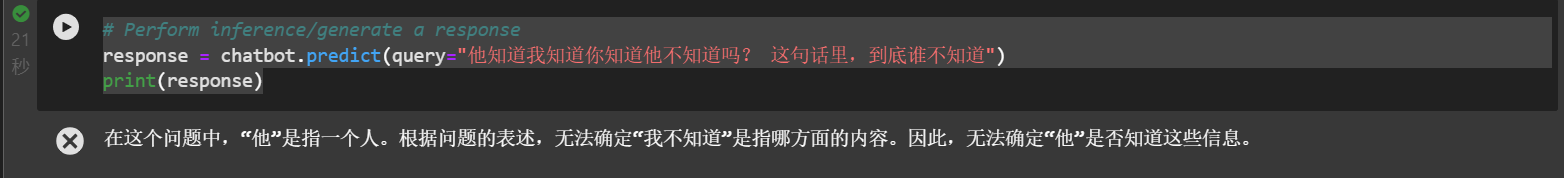


智谱ChatGLM3-6B回答：

对于第二个问题，我们人类的正常理解是人单身的原因，可能是自己看不上他人，也有可能是他人看不上自己。通义千问Qwen-7B-Chat的回答与智谱ChatGLM3-6B的回答在解释上都和人类的正常思维相符，但是ChatGLM3-6B回答中的两句话语意重复我有点没看明白，可能是暂时性出错吧。再考虑两者回答所需时间，通义千问Qwen-7B-Chat是一分零二秒，ChatGLM3-6B是52秒。综上所述，通义千问Qwen-7B-Chat回答的更加容易理解、准确，ChatGLM3-6B虽然回答快些，但是出了点不影响整体意思的小问题。

问题三.他知道我知道你知道他不知道吗？这句话里，到底谁不知道；

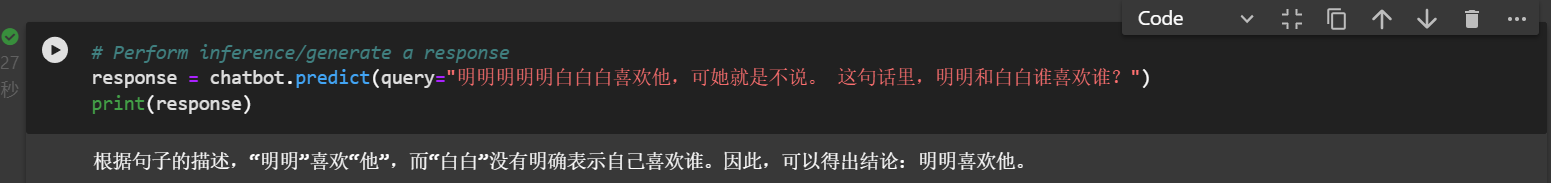
通义千问Qwen-7B-Chat回答：

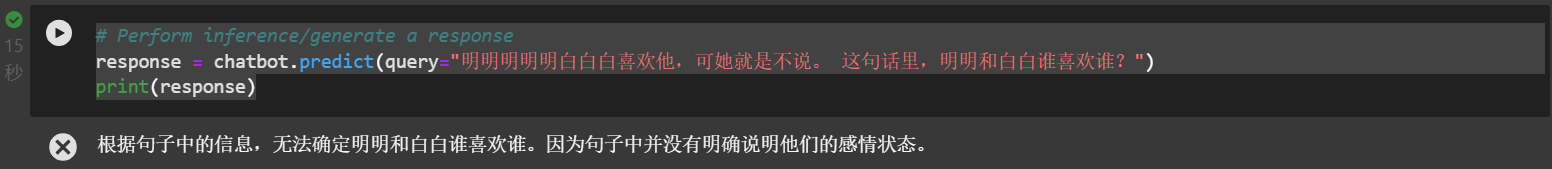
智谱ChatGLM3-6B回答：

对于第三个问题，我们人类的正常理解应该是“他”不知道，这句话大致意思是，“你”知道“他不知道”这一事实，我又知道“你知道他不知道”这一事实，“我”想问“他”到底知不知道“我知道你知道他不知道”这一事实。通义千问Qwen-7B-Chat的回答没有最出很好的解释，而且给出的答案是错的，因为他并不知道；智谱ChatGLM3-6B也没回答上来，而且感觉答非所问。再考虑两者回答所需时间，两者都是21秒。综上所述，两者对这道题的回答均不理想，但是我决定通义千问Qwen-7B-Chat回答地更深入些，因为它回答了“他”到底知不知道这一问题，虽然最后回答错了；智谱ChatGLM3-6B回答的更不着边际，就是它肯还没懂这句话是在问它什么。

问题四.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?

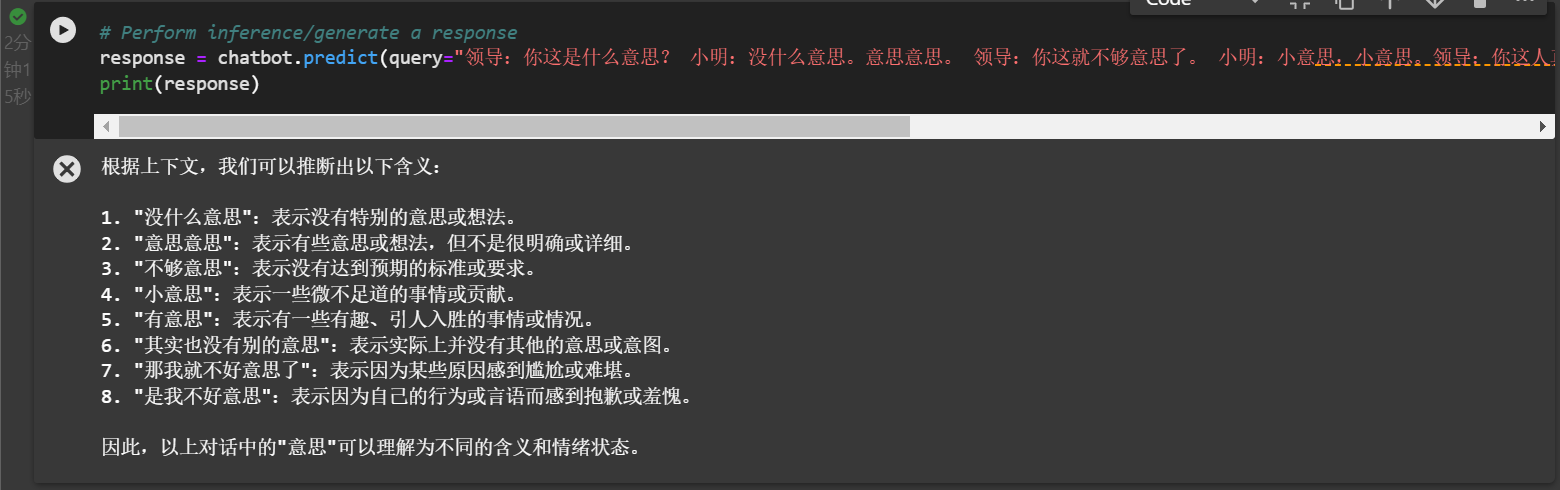
通义千问Qwen-7B-Chat回答：



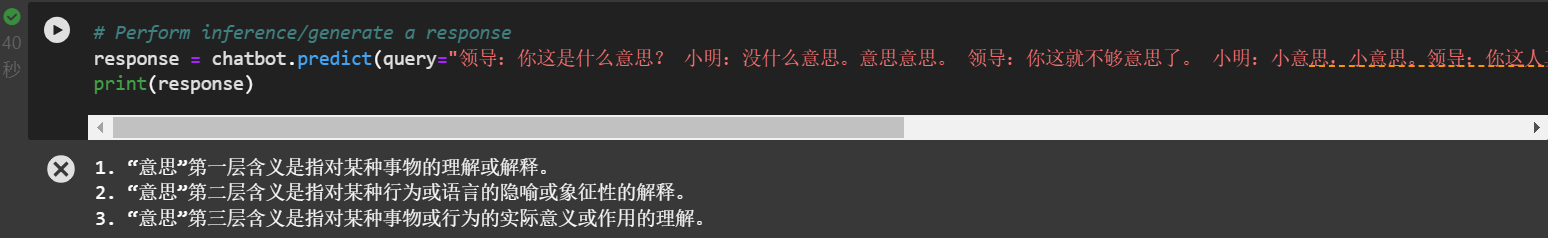
智谱ChatGLM3-6B回答：

对于第四个问题，我们人类的正常理解应该是白白喜欢明明。我们从句子中可以知道明明是男生，白白是女生，白白喜欢明明但是白白没有说出口。通义千问Qwen-7B-Chat的回答是错的；智谱ChatGLM3-6B的回答感觉又犯了前面一个问题的错误，就是没有读懂我在问什么，它连谁喜欢谁都没有做出判断。再考虑两者回答所需时间，通义千问Qwen-7B-Chat是27秒，智谱ChatGLM3-6B是15秒。综上所述，我觉得依旧是通义千问Qwen-7B-Chat回答更好，虽然他回答错误，但是它至少是在回答这个问题的具体内容；但是智谱ChatGLM3-6B直接连判断都做不出来了。

问题五.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。

通义千问Qwen-7B-Chat回答：

智谱ChatGLM3-6B回答：

对于第五个问题，确实有点绕，以下是我作为人类的理解：领导一开始问小明行为的意图(第一个意思)，小明回答没啥重要的(第二个意思)，只是想表达一下对领导的感激(第三个意思意思)，领导回答小明这行为可能不太好(第四个意思)，小明回答这只是小问题(比如送小礼物啥的，第五个意思)，领导回复小明这人不错，懂得找领导套近乎关系(第六个意思)，小明回答其实也只是想感谢一下领导，没有想叫领导帮忙照顾啥的(第七个意思)，领导回复那他这样白白收礼但不帮助小明他不太好意思(第八个意思)，小明谦称说领导不要不好意思是我自己送礼不好意思。通义千问Qwen-7B-Chat的回答是挺不错的，每一个“意思”的意思都解释得较为清楚而且与人类思维大致相符；智谱ChatGLM3-6B的回答就差点意思，只是笼统地解释“意思”这个词在众多语境中的意思，没有代入具体的分析，好像它没看懂问题的文本一样。再考虑两者回答所需时间，通义千问Qwen-7B-Chat是两分十五秒，智谱ChatGLM3-6B是40秒。综上所述，我认为通义千问Qwen-7B-Chat回答的更好，虽然它花点时间更多，但是回答质量更高。

**3.总结分析(此部分对应作业要求3：语言模型之间的横向对比分析)**

通义千问-7B(Qwen-7B)是阿里云研发的通义千问大模型系列的70亿参数规模的模型。它是基于Transformer的大语言模型, 在超大规模的预训练数据上进行训练得到。ChatGLM3是智谱AI和清华大学 KEG 实验室联合发布的对话预训练模型。两者都是当代优秀的大语言模型。

在本次项目的五个问题中，两种大语言模型都各有所长。对于比较简单的问题(比如问题一和问题二)来说，两者都做出了较好的回答，但是智谱ChatGLM3-6B可能在一些小细节的表述上不太到位，但是整体的意思是正确的；对于中等难度的题目(比如问题三和问题四)来说，通义千问Qwen-7B-Chat虽然知道你在问什么，但是给出了错误的结果，ChatGLM3-6B好像看不懂问题所在，没有给出确切的答案，而是给出“无法做出判断”类似这样的不确定性语句；对于难度较高的题目(比如问题五)，通义千问Qwen-7B-Chat可以给出较为正确的答案，但是回答的速度并不快，ChatGLM3-6B的回答就比较没有贴合实际的语境，给出的回答太过于笼统泛化，可能是它没有充分地利用上下文语境，没有很好地理解所问问题大意的关键所在。

两者都是优秀的大语言模型，经过这次项目实验，我给出基于个人开发的意见：通义千问Qwen-7B-Chat可能需要在问题回复正确性上多加训练改进；ChatGLM3-6B可能需要在问题的深刻理解上多加训练改进。