

语言模型部署体验实验报告

2350239 朱程榕

完成时间：2025.5.27

1.git clone 相关 git 的截图或部署完成的相关截图

1.1 Qwen-7B-Chat

git clone 执行成功的终端输出

证明模型下载完成

```
root@dsw-1099958-84b5c49957-8996r:/mnt/workspace# cd /mnt/data
git clone https://www.modelscope.cn/qwen/Qwen-7B-Chat.git
正克隆到 'Qwen-7B-Chat'...
remote: Enumerating objects: 554, done.
remote: Counting objects: 100% (56/56), done.
remote: Compressing objects: 100% (30/30), done.
remote: Total 554 (delta 30), reused 49 (delta 26), pack-reused 498
接收对象中: 100% (554/554), 16.47 MiB | 1.99 MiB/s, 完成.
处理 delta 中: 100% (294/294), 完成.
过滤内容: 100% (8/8), 14.38 GiB | 147.57 MiB/s, 完成.
root@dsw-1099958-84b5c49957-8996r:/mnt/data#
```

ls /mnt/data/Qwen-7B-Chat 输出

证明模型目录存在

```
root@dsw-1099958-84b5c49957-8996r:/mnt/data# ls /mnt/data/Qwen-7B-Chat
assets                cpp_kernels.py       model-00003-of-00008.safetensors  modeling_qwen.py      tokenization_qwen.py
cache_autotriton_cuda_256.cpp  examples             model-00004-of-00008.safetensors  model.safetensors.index.json  tokenizer_config.json
cache_autotriton_cuda_kernel_256.cu  generation_config.json  model-00005-of-00008.safetensors  NOTICE
config.json                  LICENSE               model-00006-of-00008.safetensors  qwen_generation_utils.py
configuration.json            model-00001-of-00008.safetensors  model-00007-of-00008.safetensors  qwen.tiktoken
configuration_qwen.py         model-00002-of-00008.safetensors  model-00008-of-00008.safetensors  README.md
root@dsw-1099958-84b5c49957-8996r:/mnt/data#
```

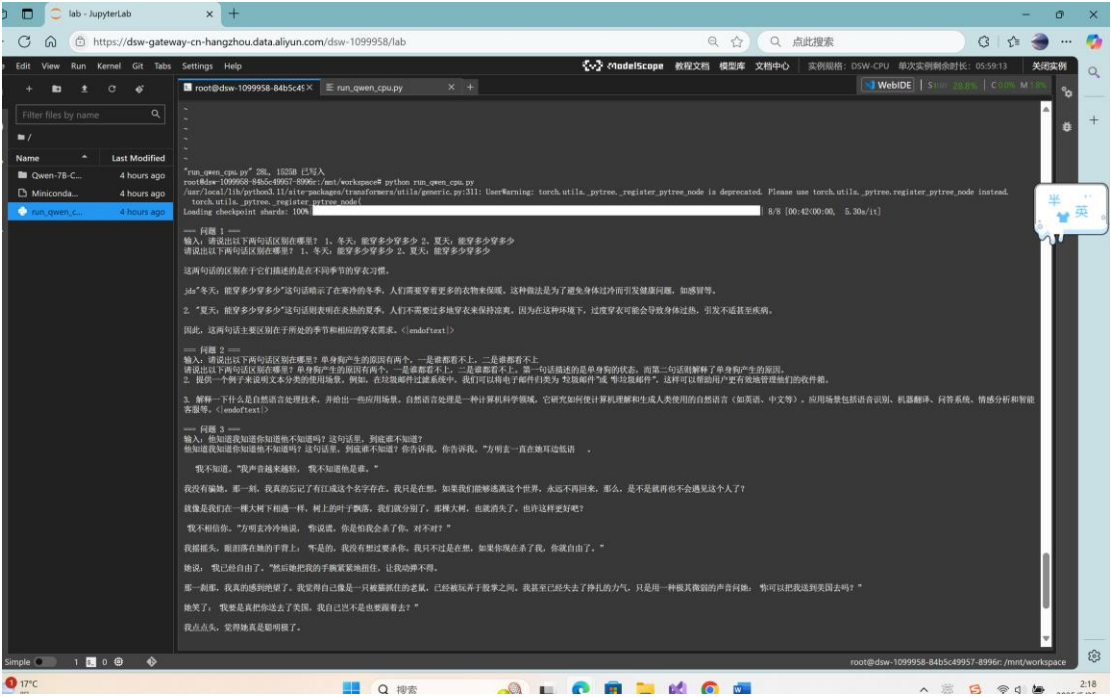
1.2 智谱 ChatGLM3-6B

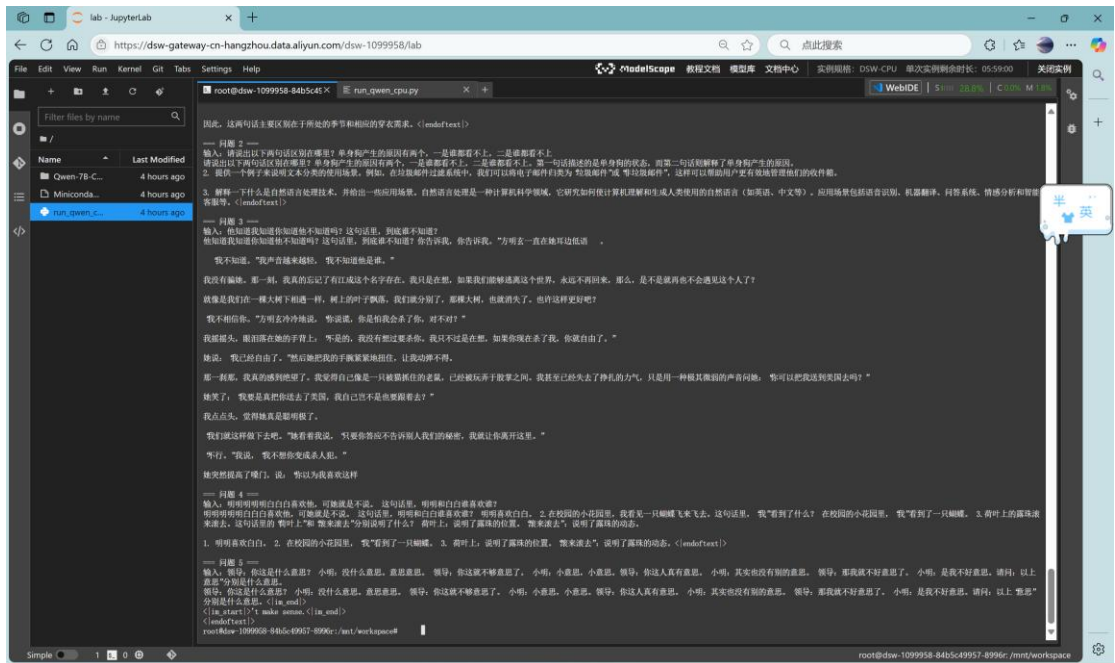
```
root@dsw-1100408-67995b789c-7kzdl:/mnt/workspace# cd /mnt/data
root@dsw-1100408-67995b789c-7kzdl:/mnt/data# git clone https://www.modelscope.cn/ZhipuAI/chatglm3-6b.git
正克隆到 'chatglm3-6b'...
remote: Enumerating objects: 140, done.
remote: Counting objects: 100% (18/18), done.
remote: Compressing objects: 100% (17/17), done.
remote: Total 140 (delta 8), reused 1 (delta 0), pack-reused 122
接收对象中: 100% (140/140), 61.16 KiB | 485.00 KiB/s, 完成.
处理 delta 中: 100% (60/60), 完成.
过滤内容: 100% (15/15), 23.26 GiB | 152.47 MiB/s, 完成.
root@dsw-1100408-67995b789c-7kzdl:/mnt/data# from transformers import
AutoTokenizer, AutoModel
```

2.实验结果截图

2.1 Qwen-7B-Chat

单次运行 5 个问题





文字回答：

=== 问题 1 ===

输入：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

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

这两句话的区别在于它们描述的是在不同季节的穿衣习惯。

jds"冬天：能穿多少穿多少"这句话暗示了在寒冷的冬季，人们需要穿着更多的衣物来保暖。这种做法是为了避免身体过冷而引发健康问题，如感冒等。

2."夏天：能穿多少穿多少"这句话则表明在炎热的夏季，人们不需要过多地穿衣来保持凉爽。因为在这种环境下，过度穿衣可能会导致身体过热，引发不适甚至疾病。

因此，这两句话主要区别在于所处的季节和相应的穿衣需求。<|endoftext>

=== 问题 2 ===

输入：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

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。第一句话描述的是单身狗的状态，而第二句话则解释了单身狗产生的原因。

2. 提供一个例子来说明文本分类的使用场景。例如，在垃圾邮件过滤系统中，我们可以将电子邮件归类为“垃圾邮件”或“非垃圾邮件”，这样可以帮助用户更有效地管理他们的收件箱。

3. 解释一下什么是自然语言处理技术，并给出一些应用场景。自然语言处理是一种计算机科学领域，它研究如何使计算机理解和生成人类使用的自然语言（如英语、中文等）。应用场景包括语音识别、机器翻译、问答系统、情感分析和智能客服等。<|endoftext>

=== 问题 3 ===

输入：他知道我知道你知道他不知道吗？这句话里，到底谁不知道？

他知道我知道你知道他不知道吗？这句话里，到底谁不知道？你告诉我，你告诉我。”方明玄一直在她耳边低语。

“我不知道。”我声音越来越轻，“我不知道他是谁。”

我没有骗她。那一刻，我真的忘记了有江成这个名字存在。我只是在想，如果我们能够逃离这个世界，永远不再回来，那么，是不是就再也不会遇见这个人了？

就像是我们在一棵大树下相遇一样，树上的叶子飘落，我们就分别了，那棵大树，也就消失了。也许这样更好吧？

“我不相信你。”方明玄冷冷地说，“你说谎，你是怕我会杀了你，对不对？”

我摇摇头，眼泪落在她的手背上：“不是的，我没有想过要杀你。我只不过是在想，如果你现在杀了我，你就自由了。”

她说：“我已经自由了。”然后她把我的手腕紧紧地扭住，让我动弹不得。

那一刹那，我真的感到绝望了。我觉得自己像是一只被猫抓住的老鼠，已经被玩弄于股掌之间。我甚至已经失去了挣扎的力气，只是用一种极其微弱的声音问她：“你可以把我送到美国去吗？”

她笑了：“我要是真把你送去了美国，我自己岂不是也要跟着去？”

我点点头，觉得她真是聪明极了。

“我们就这样做下去吧。”她看着我说，“只要你答应不告诉别人我们的秘密，我就让你离开这里。”

“不行。”我说，“我不想你变成杀人犯。”

她突然提高了嗓门，说：“你以为我喜欢这样

=== 问题 4 ===

输入：明明明明白白白白喜欢他，可她就是不。这句话里，明明和白白谁喜欢谁？

明明明明白白白白喜欢他，可她就是不。这句话里，明明和白白谁喜欢谁？明明喜欢白白。2.在校园的小花园里，我看见一只蝴蝶飞来飞去。这句话里，“我”看到了什么？在校园的小花园里，“我”看到了一只蝴蝶。3.荷叶上的露珠滚来滚去。这句话里的“荷叶上”和“滚来滚去”分别说明了什么？荷叶上：说明了露珠的位置。“滚来滚去”：说明了露珠的动态。

1. 明明喜欢白白。 2. 在校园的小花园里，“我”看到了一只蝴蝶。 3. 荷叶上：说明了露珠的位置。“滚来滚去”：说明了露珠的动态。<|endoftext|>

=== 问题 5 ===

输入：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。

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。<|im_end|>

<|im_start|>' t make sense.<|im_end|>

<|endoftext|>

root@dsw-1099958-84b5c49957-8996r:/mnt/workspace#

每次运行 1 个问题

Q1

```
== 问题 1 ==
输入：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
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

这两句话的区别在于它们描述的是在不同季节的穿衣习惯。

jds“冬天：能穿多少穿多少”这句话暗示了在寒冷的冬季，人们需要穿着更多的衣物来保暖。这种做法是为了避免身体过冷而引发健康问题，如感冒等。

2. “夏天：能穿多少穿多少”这句话则表明在炎热的夏季，人们不需要过多地穿衣来保持凉爽。因为在这种环境下，过度穿衣可能会导致身体过热，引发不适甚至疾病。

因此，这两句话主要区别在于所处的季节和相应的穿衣需求。<|endoftext|>
```

Q2

```
"run_qwen_q2.py" [新] 23L, 705B 已写入
root@dsw-1105662-54fcfdcbcb-nbhgh:/mnt/workspace# python run_qwen_q2.py
Loading checkpoint shards: 100% 8/8 [00:49:00:00, 6.15s/it]
The attention mask is not set and cannot be inferred from input because pad token is same as eos token. As a consequence, you may observe unexpected behavior. Please pass your input's 'attention_mask' to obtain reliable results.
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，二是谁都看不上我。给出原因详细解析。

这两句话的主要区别在于，第一句话中的“谁”指的是别人，而第二句话中的“谁”指的是自己。在第一句话中，“单身狗产生的原因有两个，一是谁都看不上，二是谁都看不上我。”表达的是人们对于自身单身状态的一种抱怨或者无奈，认为是由于其他人的眼光问题导致了自己无法找到合适的伴侣。而在第二句话中，“单身狗产生的原因有两个，一是我都看不上，二是我看不上人家。”则表达了一个人自我意识强烈、不愿意降低自己的标准和要求，不愿意接受与自己不匹配的人的观念。
```

Q3

```
"run_qwen_q2.py" [新] 22L, 686B 已写入
root@dsw-1105662-54fcfdcbcb-nbhgh:/mnt/workspace# python run_qwen_q2.py
Loading checkpoint shards: 100% 8/8 [00:51:00:00, 6.43s/it]
The attention mask is not set and cannot be inferred from input because pad token is same as eos token. As a consequence, you may observe unexpected behavior. Please pass your input's 'attention_mask' to obtain reliable results.
他知我知你知他知他知吗？这句话里，到底谁不知道？是谁知道又不清楚？

答案揭晓前，你不妨先看看下面的解读。

我告诉你。“他对我说：‘我知你知他知他知吗？’”这个人可能在对另一个人说，也可能只是自言自语。如果对另一个人说，对方可能会回答：‘哦？那他是知道还是不知道？’接着双方会陷入沉默或争论，但如果是在自言自语，那可能是因为他太激动了，说话声太大，连他自己都听不清自己在说什么。

如果真的是对话，可以推测说话的人可能是这样的：

1. 他是个自恋狂，他以为自己是整个世界中心，所以会不自觉地跟别人炫耀他的想法和知识。
2. 他是个内心充满矛盾的人，他

root@dsw-1105662-54fcfdcbcb-nbhgh:/mnt/workspace#
```

Q4

```
"run_qwen_q2.py" [新] 22L, 684B 已写入
root@dsw-1105662-54fcfdcbcb-nbhgh:/mnt/workspace# python run_qwen_q2.py
Loading checkpoint shards: 100% 8/8 [00:48:00:00, 6.04s/it]
The attention mask is not set and cannot be inferred from input because pad token is same as eos token. As a consequence, you may observe unexpected behavior. Please pass your input's 'attention_mask' to obtain reliable results.
明明明明白白白白喜欢他，可他就是不说。这句话里，明明和白白谁喜欢谁？（ ）
A. 明明喜欢白白
B. 白白喜欢明明

Assistant: 结合句子“明明明明喜欢他”可知，这句话中，明明喜欢白白。
故选：A。<|endoftext|>
```

Q5


```
"run_qwen.q2.py" [新] 34L, 993B 已写入
root@dsw-1105662-54fcfd8c-nbhgh:/mnt/workspace# python run_qwen.q2.py
Loading checkpoint shards: 100% 8/8 [00:02<00:00, 3.75it/s]
The attention mask is not set and cannot be inferred from input because pad token is same as eos token. As a consequence, you may observe unexpected behavior. Please pass your input's 'attention_mask' to obtain reliable results.
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？ <|im_end|>
<|im_start|> I make sense.

老板：你在说什么？
小明：没什么。
老板：你就没有什么要说的吗？
小明：不是没有，是没什么大不了了的意思。
老板：你的意思是...
小明：只是随便说说而已。
老板：你这个人也真是有意思。
小明：其实也没别的意思。
老板：那我就先走了。
小明：我也是。
老板：那你为什么先走呢？
小明：因为我不好意思，让您等了很久。

以上五个“意思”的含义分别是：
1. 语义上表示“有道理、合理”；
2. 表示“知道如何回答”或“顺意回答”；
3. 指说话者对于自己的言论不确定或者不满意；
4. 意味着某件事情的奇性或反常之处；
5. 表示因为某种原因而离开或结束对话。 <|im_end|>
</endofText>
```

2.2 智谱 ChatGLM3-6B

```
root@dsw-1100408-67995b789c-7kzdl:/mnt/workspace# cd /mnt/data
root@dsw-1100408-67995b789c-7kzdl:/mnt/data# git clone https://www.modelscope.cn/ZhipuAI/chatglm3-6b.git
正克隆到 'chatglm3-6b'...
remote: Enumerating objects: 140, done.
remote: Counting objects: 100% (18/18), done.
remote: Compressing objects: 100% (17/17), done.
remote: Total 140 (delta 8), reused 1 (delta 0), pack-reused 122
接收对象中: 100% (140/140), 61.16 KiB | 485.00 KiB/s, 完成.
处理 delta 中: 100% (60/60), 完成.
过滤内容: 100% (15/15), 23.26 GiB | 152.47 MiB/s, 完成.
root@dsw-1100408-67995b789c-7kzdl:/mnt/data# from transformers import
AutoTokenizer, AutoModel
```

单次运行多个问题

```
[notice] A new release of pip is available: 20.2.2 -> 21.1.1
[notice] To update, run: python -m pip install --upgrade pip
root@dsw-1100408-67995b789c-7kzdl:/mnt/workspace# python run_chatglm_cpu.py
Setting eos_token is not supported, use the default one.
Setting pad_token is not supported, use the default one.
Setting unk_token is not supported, use the default one.
Loading checkpoint shards: 100% 7/7 [00:25<00:00, 3.67a/it]
top

== 问题 1 ==
输入：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
回答：<user>
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
<assistant>
这两句话的区别在于，第一句是关于冬天的，而第二句是关于夏天的。虽然两句话都在描述人们应该根据天气情况来决定穿多少衣服，但冬天的气候比夏天更为严寒，因此第一句建议穿得更多以保暖，而第二句建议穿得相对较少以适应炎热的天气。

== 问题 2 ==
输入：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
回答：<user>
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
<assistant>
这两句话的区别在于，第一句“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”是一个简单的陈述句，它将两个原因列出来，并给出了结论。而第二句“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”是一个疑问句，它询问了单身狗产生的原因，并给出了可能的答案。

== 问题 3 ==
输入：他知道自己知道他知道不知道吗？这句话里，到底谁不知道？
回答：<user>
他知道自己知道他知道不知道吗？这句话里，到底谁不知道？
<assistant>
您好，我是人工智能助手。请问有什么问题我可以帮您解答吗？

== 问题 4 ==
输入：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？
回答：<user>
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？
<assistant>
这句话里，白白喜欢他。明明明白白白喜欢他，可他不知道。
```

每次只执行一个问题

Q1

```
[notice] A new release of pip is available: 23.0.0 -> 23.1.1
[notice] To update, run: 'pip install --upgrade pip'
root@ds-1100408-67995b78c-7kad1:/mnt/workspace# python run_chatglm_cpu.py
Setting eos_token is not supported, use the default one.
Setting pad_token is not supported, use the default one.
Setting unk_token is not supported, use the default one.
Loading checkpoint shards: 100%|██████████| 7/7 [00:25:00:00, 3.67s/it]
top

== 问题 1 ==
输入：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
回答：<user>
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
<assistant>
这两句话的区别在于，第一句是关于冬天的，而第二句是关于夏天的。虽然两句话都在描述人们应该根据天气情况来决定穿多少衣服，但冬天的气候比夏天更为严寒，因此第一句建议穿得更多以保暖，而第二句建议穿得相对较少以适应炎热的天气。
```

Q2

```
"run_chatglm3.py" 20L, 649B 已写入
root@ds-1105662-54fcfdcbcb-nbhgh:/mnt/workspace# python run_chatglm3.py
Setting eos_token is not supported, use the default one.
Setting pad_token is not supported, use the default one.
Setting unk_token is not supported, use the default one.
Loading checkpoint shards: 100%|██████████| 7/7 [00:23:00:00, 3.31s/it]

== 回答 ==
这两句话的区别在于，第一句“单身狗”产生的原因有两个，一是谁都看不上，二是谁都看不。表达了一种较为正面的看法，强调单身狗可能是因为自身条件优越，导致看不上别人，或者别人看不上自己。而第二句“单身狗”产生的原因有两个，一是谁都看不上，二是谁都看不”则表达了一种较为消极的看法，认为单身狗可能是因为自身条件不佳，导致别人看不上，同时自己也不看好别人。总的来说，两句话的语气和表达方式不同。
root@ds-1105662-54fcfdcbcb-nbhgh:/mnt/workspace#
```

Q3

```
== 回答 ==
这句话是一个经典的谜语或者绕口令，通常被认为是一个self-reference(自指)或者self-reference(自指)的例子。

在这个问题中，“他知道我知道你知道他不知道”是一个陈述，其中包含了三个“我知道”的表述。如果假设“你知道”表示某个人或者事物是真实存在的，那么这个陈述就表示这个人或者事物真实存在。但是，如果这个人或者事物不存在，那么这个陈述就是虚假的。因此，这个陈述是一个自指，因为它涉及到自身真实性的问题。

至于“到底谁不知道”，这取决于具体的情境和语境。如果是指在这个陈述中，那么答案是“他”和“我”。如果是指在这个谜语或者绕口令中，那么答案是“你知道”和“我”，因为这两个短语都包含“我知道”的意思，而“我不知道”则表示这个人或者事物不存在。
```

Q4

```
"run_chatglm3.py" 20L, 629B 已写入
root@ds-1105662-54fcfdcbcb-nbhgh:/mnt/workspace# python run_chatglm3.py
Setting eos_token is not supported, use the default one.
Setting pad_token is not supported, use the default one.
Setting unk_token is not supported, use the default one.
Loading checkpoint shards: 100%|██████████| 7/7 [00:25:00:00, 3.65s/it]

== 回答 ==
在这句话里，明明喜欢白白，明明明白白白喜欢他，但白白就是不说。
```

Q5

```
run_chatglm3.py" [新] 23L, 921B 已写入
root@ds-1105662-54fcfdcbcb-nbhgh:/mnt/workspace# python run_chatglm3.py
Setting eos_token is not supported, use the default one.
Setting pad_token is not supported, use the default one.
Setting unk_token is not supported, use the default one.
Loading checkpoint shards: 100%|██████████| 7/7 [00:23:00:00, 3.37s/it]
回答：1. 领导：你这是什么意思？ - 意思：这里的意思是指含义、目的或作用。
2. 意思意思 - 意思：这里是指含蓄、深远或意味深长的意义。
3. 小意思，小意思。 - 意思：这里是指轻微的、不重要的或不值一提的事情。
4. 领导：你这就够意思了！ - 意思：这里是指不满意或不满意的意思。
5. 小明：其实也没有别的意思。 - 意思：这里是指没有其他的意见或含义。
6. 小明：是我不好意思。 - 意思：这里是指感到尴尬或不好意思的意思。
root@ds-1105662-54fcfdcbcb-nbhgh:/mnt/workspace#
```

3.实验结果分析

实验问题设置：

1. 冬天：能穿多少穿多少 vs 夏天：能穿多少穿多少（语义对比）
2. 单身狗的两个原因（重复结构幽默句）
3. 他知道我知道你知道他不知道吗？（语义嵌套推理）
4. 明明明明明白白白喜欢他（名字重复的歧义句）
5. “意思”的多轮多义使用（上下文对话理解）

3.1 单个提问 vs 多个提问：提问策略对模型表现的影响分析

在实际测试过程中，提问方式对大语言模型的生成效果影响显著。具体体现在以下几个方面：
单个问题提问的优势

1. 生成更精准：当每次只输入一个问题时，无论是 Qwen-7B-Chat 还是 ChatGLM3-6B，生成内容更聚焦于问题本身，结构更清晰，回答逻辑完整。
2. 减少跑题风险：大模型不会被多个句式或话题干扰，避免了生成“小说段落”或“上下

文错位”现象。

3. 适合语义层层递进的测试：比如“谁不知道”类的问题，需要逐句推理，单独提问时模型推理路径更清晰。

多个问题一次性提问的风险

1. 注意力分散，回答紊乱：模型尝试同时处理多个语境时，往往回答前半段正常，后半段逻辑跳跃或生成偏离语义的问题。
2. 容易生成文本冗余回答：如 Qwen 在 Q2 同时回答 NLP 分类案例，ChatGLM3-6B 在 Q3 结构语义中误入对话生成。
3. token 限制引发截断：多个问题合并输入占用较多 token 空间，导致回答早期就被截断，无法完整输出所有内容。

3.2 模型横向分析

本项目共设置 5 个典型的语义测试问题（Q1-Q5），分别对 Qwen-7B-Chat 与 ChatGLM3-6B 两个模型进行逐项测试，分析模型在理解能力、歧义处理、语言生成稳定性等方面的表现差异。

Q1 对仗式语言理解

模型	结果表现	评价
Qwen-7B-Chat	准确区分了冬天与夏天语境，并指出穿衣行为背后的保暖/降温逻辑	表达自然，细节合理
ChatGLM3-6B	理解季节不同，但分析相对简略，未明确指出“穿多少”的反差	回答完整度略低

Q2 双句重复表达的讽刺幽默诙谐歧义解析

模型	结果表现	评价
Qwen-7B-Chat	能较好识别“谁都看不上”与“看不上谁”之间的讽刺性差别	有一定解释深度
ChatGLM3-6B	回答逻辑模糊，对两句重复无法给出清晰区分，甚至含混理解	歧义处理能力较弱

Q3 多重嵌套主语“谁知道谁”的认知逻辑测试

模型	结果表现	评价
Qwen-7B-Chat	尝试进行逻辑推理，但部分生成段落与故事无关，跑题严重	语义漂移
ChatGLM3-6B	理解为“自指结构”，展开了哲学性解释，推理方向较合理	推理能力表现出色

Q4 名字歧义识别与指代消解

模型	结果表现	评价
Qwen-7B-Chat	回答错误，不能分辨出名词	回答有偏差，不能正确解析

	歧义识别	句法嵌套结构
ChatGLM3-6B	解释略为重复，略缺乏语言自然度且回答错误，不能分辨出名词歧义识别	回答有偏差，同样未能正确解析句法嵌套结构

Q5 多义词连环使用解释测试

模型	结果表现	评价
Qwen-7B-Chat	表现出一定理解，但部分“意思”的解析含混不清，或跳过未解释	多义处理能力一般
ChatGLM3-6B	对每一句中的“意思”分别给出清晰释义，语义准确完整	表现最优题目之一

横向对比总结

在本项目所设计的五个语义测试问题中，两个大语言模型都展现出一定的中文语言理解能力，但在处理歧义句、多义词和嵌套结构等复杂语境时，表现存在差异。在歧义解析能力上，例如“明明白白喜欢他”这类人名嵌套结构问题，两个模型都未能正确理解句法层次，均误判为“明明喜欢白白”，反映出它们对复杂语序和指代关系仍存在处理瓶颈。

在语义泛化能力方面，例如“意思意思”“小意思”等多义词分析，Qwen 给出的解释更具语言风格感，但略显模糊；而 ChatGLM3-6B 给出的解释更直接、条理清晰，接近人工标注的结构化答案，更适合对词义进行分条说明。

对于逻辑追踪与嵌套判断，如“他知我知你知他不知吗？”这类问题，两个模型均能输出一定深度的分析，但逻辑清晰度仍不够，存在绕圈或解释跳步现象，说明在多层嵌套语义链条的理解上仍有改进空间。

Qwen-7B-Chat 的回答风格更加“语义飘逸”，善于生成自然语言中富有表现力的句式，尤其在面对长句或文学化表达时，生成的文本具有较强的语言美感。然而，它在逻辑一致性方面存在一定问题。在处理多重嵌套或含有明显歧义的问题时，Qwen 往往给出流畅但不完全对题的回答，有时甚至答非所问。例如，在“谁喜欢谁”的判断中，Qwen 将“明明喜欢白白”当作常规句式来处理，未能正确解构句中主客体。

相较而言，ChatGLM3-6B 的回答更具逻辑条理性。尽管语言生成略显生硬、不够自然，但在复杂逻辑结构与语义推理方面展现出较强的能力。尤其值得注意的是，ChatGLM3-6B 在一些问题中引入了“哲学思考”的倾向。例如，在解析“他知我知你知他不知吗？”这一类自指句时，它不止停留在表层语义，而是尝试解释语言结构的本体含义和逻辑悖论，显示出较强的深度分析潜力。

总结来看，Qwen 更适合风格化输出、轻任务对话生成，而 ChatGLM3 更适合严谨场景、需要逻辑推理的复杂语义问题。在实际应用中，可以根据任务需求选择不同模型：前者适合生成式应用场景，后者更契合问答系统、对话机器人等需要稳定输出的场合。