

# 人工智能导论第四次作业

## 一、项目公开可访问链接

[https://github.com/Chenhongliang777/models\\_compare](https://github.com/Chenhongliang777/models_compare)

## 二、模型部署结果图

### 2.1 Baichuan2-7B-Chat 部署结果图

```
response = model.chat(tokenizer, messages)
print(response)
messages.append({"role": "assistant", "content": response})

messages.append({"role": "user", "content": "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。"})
response = model.chat(tokenizer, messages)
print(response)
messages.append({"role": "assistant", "content": response})

messages.append({"role": "user", "content": "他知道我知道你知道他不知道吗？这句话里，到底谁不知道。"})
response = model.chat(tokenizer, messages)
print(response)
messages.append({"role": "assistant", "content": response})

messages.append({"role": "user", "content": "明明明明明白白白喜欢他，可能就是不说。这句话里，明明和白白谁喜欢谁。"})
response = model.chat(tokenizer, messages)
print(response)
messages.append({"role": "assistant", "content": response})

messages.append({"role": "user", "content": "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：那我就不客气了。"})
response = model.chat(tokenizer, messages)
print(response)
```

/root/miniconda3/envs/chatbot\_env/lib/python3.11/site-packages/tqdm/auto.py:21: TqdmWarning: IPProgress not found. Please update jupyter and ipywidgets. See [https://ipywidgets.readthedocs.io/en/stable/user\\_install.html](https://ipywidgets.readthedocs.io/en/stable/user_install.html)

from .autonotebook import tqdm as notebook\_tqdm

Downloading Model from <https://www.modelscope.cn> to directory: /mnt/workspace/.cache/modelscope/models/baichuan-inc/Baichuan2-7B-Chat

2025-05-19 00:40:37,008 - modelscope - INFO - Use user-specified model revision: v1.0.5

/root/miniconda3/envs/chatbot\_env/lib/python3.11/contextlib.py:105: FutureWarning: `torch.backends.cuda.sdp\_kernel()` is deprecated. In the future, this context manager will be removed. Please see `torch.nn.attention.sdpa\_kernel()` for the new context manager, with updated signature.

self.gen = func(\*args, \*\*kwargs)

### 2.2 Qwen-7B-Chat 部署结果图

```
/root/miniconda3/envs/chatbot_env/lib/python3.11/site-packages/tqdm/auto.py:21: TqdmWarning: IPProgress not found. Please update jupyter and ipywidgets. See https://ipywidgets.readthedocs.io/en/stable/user\_install.html
```

from .autonotebook import tqdm as notebook\_tqdm

Downloading Model from <https://www.modelscope.cn> to directory: /mnt/workspace/.cache/modelscope/models/Qwen/Qwen-7B-Chat

2025-05-19 00:12:43,524 - modelscope - INFO - Use user-specified model revision: v1.0.5

Loading checkpoint shards: 100% [██████████] 8/8 [00:16<00:00, 2.01s/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.

## 三、问题测试结果

对两个大模型分别询问相同的五个问题

### 3.1 Baichuan2-7B-Chat 的回答

1.请说出以下两句话区别在哪里？冬天：能穿多少穿多少 2、夏天：能穿多少穿多少

这两句话的主要区别在于所描述的季节不同。第一句“冬天：能穿多少穿多少”是针对冬季的穿衣建议，强调在寒冷的冬天里要穿得保暖；而第二句“夏天：能穿多少穿多少”是针对夏季的穿衣建议，强调在炎热的夏天里要穿得透气和舒适。

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

这两句话的主要区别在于它们描述了两种不同的单身狗产生的原因。第一句“一是谁都看不上”是指一个人因为对自己的条件或外貌过于自信，看不起别人，因此无法找到伴侣；第二句“二是谁都看不上”是指一个人因为对别人的要求过高，无法接受比自己条件差的人，因此无法找到伴侣。

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

这句话里，“他”不知道。

4.明明明明明白白白喜欢他，可她就是不说话。这句话里，明明和白白谁喜欢谁？

这句话里，“明明”喜欢“白白”。

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。

在这句话中，领导对小明客气，而小明对领导不客气。

### 3.2 Qwen-7B-Chat 的回答

1.请说出以下两句话区别在哪里？冬天：能穿多少穿多少 2、夏天：能穿多少穿多少

问：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。  
答：这两句话的区别在于它们描述的是在不同的季节里对衣服的穿着要求。第一句话是在冬天的描述，暗示应该尽可能多地穿衣服来保暖；而第二句话是在夏天的描述，暗示应该尽可能少地穿衣服来保持凉爽。

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

问：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。  
答：这两句话在表达上有一些相似之处，都是在描述单身狗产生的原因。然而，它们之间在语义上存在一定的差异。第一句话“谁都看不上”表示单身狗找不到合适的对象，而第二句话“谁都看不上”则表示单身狗本身不具备吸引异性的能力。因此，这两句话在描述单身狗产生的原因方面存在一定的分歧。

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

问：他知道我知道你知道他不知道吗？这句话里，到底谁不知道。  
答：这句话的语义比较复杂，需要仔细分析。从字面意义上来看，这句话可以理解为：他知道我知道你知道他知道。但是，根据句子的结构和语义，这句话可能有多种解释。一种解释是，句子的主语是“他”，他知道自己不知道；另外一种解释是，句子的主语是“你知道”，你知道他知道，但是你不知道他知道你知道。因此，这句话的未知数比较多，具体是指谁知道谁不知道，需要根据上下文来判断。

4.明明明明明白白白喜欢他，可她就是不说话。这句话里，明明和白白谁喜欢谁？

问：明明明明明白白白喜欢他，可她就是不说话。这句话里，明明和白白谁喜欢谁。  
答：从字面意义上来看，这句话可以理解为：她喜欢他，但是她就是不说话。但是，根据句子的结构和语义，这句话可能有多种解释。一种解释是，明明喜欢他，但是她没有说出口；另外一种解释是，白白喜欢他，但是她没有说出口。因此，这句话的未知数比较多，具体是指谁喜欢谁，需要根据上下文来判断。

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。

问：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：那我就客气了。 小明：不客气，不客气。  
请问，这句话里，谁对谁客气，谁对谁不客气。  
答：在这段对话中，小明在面对领导的质问时，两次使用了“意思意思”来表示自己没有别的意思。而领导则在多次使用“有意思”来评价小明。因此，可以说小明对领导客气，而领导对小明不客气。

## 四、大模型横向对比

### 4.1 对话内容对比

对于难度较低的问题一、二，两个模型的回答都比较详细正确，不分伯仲。

从第三问开始，由于难度上升，两个模型开始出现不同，Baichuan2-7B-Chat 会直接给出自己的判断而不加解释；而 Qwen-7B-Chat 给出判断则更为谨慎，往往会附上自己的思考过程，且对于自己不确定的问题更倾向于不做判断。从结果上看，Baichuan2-7B-Chat 答对了第三问，四、五问答错；Qwen-7B-Chat 没有回答三、四问，却答对了第五问。相比之下，似乎是 Qwen-7B-Chat 的总体回答更好（起码不会误导提问者），不过，这也可能与调用模型时的参数设置有关，缺乏控制变量。

### 4.2 模型架构

**Baichuan2-7B-Chat**：基于 Transformer 架构，该架构在处理序列到序列任务方面表现出色，通过多头自注意力机制和位置编码来捕捉句子中的依赖关系和上下文信息。其特别针对对话生成和上下文理解进行了优化，使得模型在对话系统中的表现尤为突出，能够处理复杂的对话上下文，生成自然且连贯的对话内容。

**Qwen-7B-Chat**：采用 Transformer 架构，注重多语言处理和对话能力，在多语言对话生成和理解方面进行了优化，能够处理多种语言的输入和输出，生成高质量的多语言文本。

### 4.3 参数量

**Baichuan2-7B-Chat**：拥有 7 亿参数，相对较大的参数量使其能够捕捉更复杂的语言模式和细微差别，尤其在对话生成和上下文理解方面，可处理更复杂的对话场景，生成更流畅、连贯的文本。

**Qwen-7B-Chat**：同样具有 7 亿参数，这为其在多语言处理和对话任务中提供了强大的能力，能够支持多种语言的对话生成和理解，生成自然且连贯的多语言文本。

### 4.4 性能与用途

**Baichuan2-7B-Chat**：

- 中文对话生成与理解：在中文对话系统中表现卓越，生成的对话自然流畅，具有良好的上下文关联性，适用于中文环境下的对话任务，如客服机器人、虚拟助理等需要长时间对话和复杂对话场景的应用。
- 高效上下文理解：能够有效理解并保持对话上下文，使生成的对话更加连贯，确保对话的连贯性和一致性。
- 流畅性和自然性：强调对话的自然性和流畅性，在生成连续对话时表现出色，可生成生动、自然的对话内容。

**Qwen-7B-Chat**：

- 多语言对话能力：在多语言对话方面表现出色，能够处理多种语言的输入和输出，支持跨语言交流和国际化应用，适用于需要多语言支持的场景，如国际交流、多语言客服等。

- 文本生成与问答：可生成高质量的文本，适用于内容创作、智能问答等任务，能够提供准确、有帮助的回答。
- 代码生成与解释：具备一定的代码生成和解释能力，可协助用户进行编程相关的任务，如代码编写、调试等。

4.5 训练数据与优化

**Baichuan2-7B-Chat**：训练数据主要集中在中文语料库，通过自注意力机制、数据增强等优化技术提升中文对话理解和生成能力，使其在中文对话场景中具有较高的准确性和适应性。

**Qwen-7B-Chat**：利用多语言语料库进行训练，采用深度学习优化技术，如自注意力机制等，提升多语言对话能力和文本生成质量，以更好地处理多种语言的输入和输出。

4.6 应用场景

**Baichuan2-7B-Chat**：适用于中文对话系统，如中文环境下的客服机器人、虚拟助理、教育与培训等，能够为用户提供个性化的对话服务和信息支持。

**Qwen-7B-Chat**：适用于多语言环境下的各种应用场景，包括跨语言交流、内容创作、智能问答、代码辅助等，可满足用户在不同语言和领域的多样化需求。

以下是 ChatGLM3-6B 和 Baichuan-7B-Chat 在模型架构、参数量、性能与用途、训练数据与优化、以及应用场景方面的详细对比表格：

方面	Baichuan2-7B-Chat	Qwen-7B-Chat
模型架构	基于 Transformer 架构，特别针对对话生成和上下文理解进行了优化。	采用 Transformer 架构，在多语言对话生成和理解方面进行了优化。
参数量	7 亿参数	7 亿参数
性能与用途	在中文对话生成和理解、客服机器人、文本生成、对话系统中表现卓越，特别适合中文环境下的复杂对话任务。	在多语言对话、文本生成、问答系统、代码生成与解释等方面表现出色，适用于多种语言环境下的对话任务，能够生成高质量的文本，提供准确的回答和代码辅助。
训练数据与优化	训练数据主要集中在中文语料库，通过自注意力机制、数据增强等优化技术提升中文对话理解和生成能力，使其在中文对话场景中具有较高的准确性和适应性。	利用多语言语料库进行训练，采用深度学习优化技术，如自注意力机制等，提升多语言对话能力和文本生成质量，以更好地处理多种语言的输入和输出。
应用场景	适用于中文对话系统，如中文环境下的客服机器人、虚拟助理、教育与培	适用于多语言环境下的各种应用场景，包括跨语言交流、内容创作、智能问

	训等，能够为用户提供更个性化的对话服务和信息支持。	答、代码辅助等，可满足用户在不同语言和领域的多样化需求。
--	---------------------------	------------------------------

## 五、总结

总结来说，Baichuan2-7B-Chat 在中文对话任务中表现出色，生成自然流畅的对话内容，适合中文环境下的复杂对话场景；而 Qwen-7B-Chat 在多语言对话和代码生成等方面更具优势，适用于多语言场景和需要代码辅助的应用。在选择模型时，应根据具体的应用需求和场景来决定使用哪个模型。