

同濟大學

TONGJI UNIVERSITY

课程报告

课程名称: 人工智能导论

班号: 42041201

学院/专业: 计算机科学与技术学院 软件工程

学生姓名: 杨子龙

指导老师: 邓浩

撰写时间: 2025 年 6 月

一. 问题设置及分析

本次针对 Qwen-7B-Chat 与 ChatGLM3-6B 的对比实验所设计的八个问题，涵盖了语言理解、幽默识别、逻辑推理、语义歧义处理、常识运用及情感共情等多个关键能力维度。每个问题都紧扣中文语言的特点，兼具趣味性与挑战性，能够较全面地评估两个中文大语言模型在复杂语境下的语言生成与理解能力。

从整体来看，问题设置体现出良好的区分性和多样性：如第 1 题与第 5 题聚焦于语境与多义词的动态解析，第 2 与第 4 题则突出了模型对语言幽默和命名歧义的认识能力，第 3 与第 7 题体现逻辑推理与时间感知的处理水平，而第 8 题则考查了模型对用户情绪的回应能力，具有实际人机交互背景的参考价值。

1. 冬天：能穿多少穿多少；夏天：能穿多少穿多少

(1) 考察点：语境理解与双关语识别（“穿多少”在不同季节含义反转）

(2) 合理性：能考查模型是否理解句子虽形式相同但语义相反

2. 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

(1) 考察点：幽默识别、语言模糊性处理（其实只有一个原因）

(2) 合理性：可测试模型是否能识别“语言幽默”与“表述逻辑的荒谬性”

3. 他知道我知道你知道他不知道吗？

(1) 考察点：多重嵌套认知推理（ToM 层级）

(2) 合理性：能考查模型的逻辑深度处理能力

4. 明明明明白白白喜欢他，可她就是不说

(1) 考察点：语言歧义消解、命名实体识别（“明明”“白白”是人名还是副词）

(2) 合理性：极具挑战性

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。

(1) 考察点：上下文歧义解析、语用学能力（语境中的“意思”多种含义）

(2) 合理性：能够测试模型对多义词的灵活运用与对话连贯性理解

6. 为什么冬天水会结冰而夏天不会？

(1) 考察点：常识推理

(2) 合理性：可对比模型对常识理解是否稳定、科学表述是否准确

7. 妈妈比我大 20 岁，我今年 10 岁，她 5 年前几岁？

(1) 考察点：数学常识 + 时间推理

(2) 合理性：考查模型对时间逻辑的理解

8. 我今天很难过，因为我朋友不理我了，你怎么看？

(1) 考察点：情感理解与同理心（共情能力）

(2) 合理性：比较模型是否能给予安慰、建议，以及语气是否自然

二. git clone 完成模型部署 + 官方页介绍

1. 部署 Qwen-7B-Chat

```
root@dsw-1161010-844b4dfbbc-sw68r:/mnt/workspace# 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 | 731.00 KiB/s, 完成
处理 delta 中: 100% (295/295), 完成
正在更新文件: 100% (31/31), 完成
过滤内容: 100% (8/8), 14.38 GiB | 142.57 MiB/s, 完成
```

(1) 训练数据规模与多样性

Qwen-7B-Chat 是在基础 Qwen-7B 大模型(70 亿参数)上,通过监督微调(SFT)和 RLHF 对齐训练得到。其预训练覆盖托管 2.2–2.4 万亿 tokens,多语种内容(中英文)、专业书籍、网络文本、代码、数学等领域

(2) 适用场景与功能能力

- ① 多语言应用：优秀的中英文混合能力。
- ② 多任务场景：表现优于对话、翻译、写作、编程和数学问题。
- ③ 工具支持：支持 ReAct、agent、代码解释等复杂工具调用

(1) 开源与部署友好

- ① 开源：在 Hugging Face、ModelScope 上提供权重，可选择 Int4/Int8/BF16 量化版本。
- ② 长上下文支持：8K token 窗口，且架构设计支持 flash-attention 加速。
- ③ 方便微调部署：适配 LoRA、Q-LoRA，部署兼容 CPU & GPU。

(4) 模型扩展性与生态

- ① 多型号迭代：从 1.8B 到 72B 全线开源，Qwen-Chat 支持代码、数学等专业模型。
- ② 细粒度调优支持：支持 LoRA、Q-LoRA 微调，社区示例丰富。

2. 部署 chatglm3-6b

```
root@dsw-1140855-cdb6c6bcd-pqtq5:/mnt/workspace# git clone https://www.modelscope.cn/ZhipuAI/chatglm3-6b.git
bash: git clone https://www.modelscope.cn/ZhipuAI/chatglm3-6b.git: 没有那个文件或目录
root@dsw-1140855-cdb6c6bcd-pqtq5:/mnt/workspace# git clone https://www.modelscope.cn/ZhipuAI/chatglm3-6b.git
正克隆到 'chatglm3-6b'...
remote: Enumerating objects: 140, done.
remote: Total 140 (delta 0), reused 0 (delta 0), pack-reused 140
接收对象中: 100% (140/140), 47.10 KiB | 354.00 KiB/s, 完成
处理 delta 中: 100% (64/64), 完成
过滤内容: 100% (15/15), 23.26 GiB | 106.16 MiB/s, 完成
```

(1) 训练数据规模与多样性

ChatGLM3-6B 为 6.24B 参数模型，预训练使用多样中英语料，通过改进 Prompt 格式与更足训练步数，在语义、推理、数学、代码等任务中，在子 10B 模型里性能领先

(2) 适用场景与功能能力

- ① 双语对话：优化中文交互质量，兼顾英文场景。
- ② 原生工具支持：支持 Function Call、代码执行、Agent，适用复杂任务流程
- ③ 长文本理解：提供 8K 与 32K 上下文版本，适配长文应用

(3) 开源与部署便利性

- ① 权重公开：Hugging Face 提供 ChatGLM3-6B、Base、32K 等版本
- ② 可量化部署：支持 CPU/GPU 本地量化（如 GGML），并兼容 LoRA 微调

(4) 模型扩展性与生态

- ① 系列布局完善：包括 Base、对话版、32K 长文版，权重公开。
- ② 社区支持丰富：生态含微调示例、Agent 框架教程，共同构建工具调用能力

三. 输出结果

1. Qwen- 7B-Chat 模型测试结果

(1) 这两句话的区别在于所处的季节不同。第一句“冬天：能穿多少穿多少”是在描述冬天气温较低的情况下，人们为了保暖会选择穿较多的衣物；而第二句“夏天：能穿多少穿多少”则是在描述夏天气温较高的情况下，人们为了散热会尽量减少穿着，穿着相对较少。

```

/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:441: UserWarning: torch.utils._pytree.register_pytree_node is deprecated. Please use torch.utils._pytree.register_pytree_node instead.
  torch_pytree.register_pytree_node(
正在加载模型（使用 CPU）...
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.
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:309: UserWarning: torch.utils._pytree.register_pytree_node is deprecated. Please use torch.utils._pytree.register_pytree_node instead.
  torch_pytree.register_pytree_node(
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:309: UserWarning: torch.utils._pytree.register_pytree_node is deprecated. Please use torch.utils._pytree.register_pytree_node instead.
  torch_pytree.register_pytree_node(
Loading checkpoint shards: 100% | 7/7 [00:36:00.00, 5.28s/it]
你：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
ChatGLM: 这两句话的区别在于所处的季节不同。第一句“冬天：能穿多少穿多少”是在描述冬天气温较低的情况下，人们为了保暖会选择穿较多的衣物；而第二句“夏天：能穿多少穿多少”则是在描述夏天气温较高的情况下，人们为了散热会尽量减少穿着，穿着相对较少。
    
```

(2) 这两句话的区别在于表达方式和语义。

第一句话“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”，采用了列举问题的方式进行表述，明确指出了两个导致单身狗产生的原因，并且用了“一是”和“二是”进行分类。

第二句话“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”，表达方式一致，但是使用了“一是”和“二是”进行表述，同样指出了两个原因，并且内容上没有任何区别。

总结来说，两句话在表达方式和表述上略有不同，但是它们所表达的含义是相同的。

```

/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:441: UserWarning: torch.utils._pytree.register_pytree_node is deprecated. Please use torch.utils._pytree.register_pytree_node instead.
  torch_pytree.register_pytree_node(
正在加载模型（使用 CPU）...
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.
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:309: UserWarning: torch.utils._pytree.register_pytree_node is deprecated. Please use torch.utils._pytree.register_pytree_node instead.
  torch_pytree.register_pytree_node(
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:309: UserWarning: torch.utils._pytree.register_pytree_node is deprecated. Please use torch.utils._pytree.register_pytree_node instead.
  torch_pytree.register_pytree_node(
Loading checkpoint shards: 100% | 7/7 [00:32:00.00, 4.63s/it]
你：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
ChatGLM: 这两句话的区别在于表达方式和语义。
第一句话“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”，采用了列举问题的方式进行表述，明确指出了两个导致单身狗产生的原因，并且用了“一是”和“二是”进行分类。
第二句话“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”，表达方式一致，但是使用了“一是”和“二是”进行表述，同样指出了两个原因，并且内容上没有任何区别。
总结来说，两句话在表达方式和表述上略有不同，但是它们所表达的含义是相同的。
    
```

(3) 这句话是一个经典的谜语或者绕口令,它的意思是:一个人知道另外两个人知道的事情,但是第三个人并不知道。因此,答案是“第三个人不知道”。

```
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:441: UserWarning: torch.utils._pytree._register_pytree_node is deprecated. Please use torch.utils._pytree.register_pytree_node instead.
  torch_pytree._register_pytree_node(
正在加载模型 (使用 CPU) ...
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.
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:309: UserWarning: torch.utils._pytree._register_pytree_node is deprecated. Please use torch.utils._pytree.register_pytree_node instead.
  torch_pytree._register_pytree_node(
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:309: UserWarning: torch.utils._pytree._register_pytree_node is deprecated. Please use torch.utils._pytree.register_pytree_node instead.
  torch_pytree._register_pytree_node(
Loading checkpoint shards: 100% 7/7 [00:30:00.00, 4.40s/it]

你: 他知道他知道他知道他不知道吗? 这句话里, 到底谁不知道?
ChatGLM: 这句话是一个经典的谜语或者绕口令, 它的意思是: 一个人知道另外两个人知道的事情, 但是第三个人并不知道。因此, 答案是 “第三个人不知道”。
```

(4) 这句话中, 明明白白白白喜欢他, 但白白并没有明确表达出来, 无法确定谁喜欢谁。

```
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:441: UserWarning: torch.utils._pytree._register_pytree_node is deprecated. Please use torch.utils._pytree.register_pytree_node instead.
  torch_pytree._register_pytree_node(
正在加载模型 (使用 CPU) ...
Setting eos_token is not supported, use the default one.
Setting pad_token is not supported, use the default one.
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:309: UserWarning: torch.utils._pytree._register_pytree_node is deprecated. Please use torch.utils._pytree.register_pytree_node instead.
  torch_pytree._register_pytree_node(
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:309: UserWarning: torch.utils._pytree._register_pytree_node is deprecated. Please use torch.utils._pytree.register_pytree_node instead.
  torch_pytree._register_pytree_node(
Loading checkpoint shards: 100% 7/7 [00:28:00.00, 4.05s/it]

你: 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?
ChatGLM: 这句话中, 明明白白白白喜欢他, 但白白并没有明确表达出来, 所以无法确定谁喜欢谁。
```

(5)

1. 领导: 你这是什么意思? - 这个问题表示领导对某种行为或话语表示困惑或不理解, 想了解对方的意思。
2. 意思意思 - 这是一个口语表达, 意思是不太正式或不太正式地表示某种意思, 通常用于玩笑或表示亲昵。
3. 小意思, 小意思。 - 这是一个口语表达, 意思是比较小的费用或礼物, 表示对别人的感谢或表示友好的态度。
4. 这人真有意思 - 这是一个表示对某个人或某个事物表示赞赏和有趣的评价。
5. 其实也没有别的意思 - 这是一个表示对某个话题或问题已经说了很多, 但是并没有特别的意思或者目的。
6. 那我就不好意思了 - 这是一个表示对自己行为言论造成对方困扰或不适的道歉和认错。

```
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:441: UserWarning: torch.utils._pytree._register_pytree_node is deprecated. Please use torch.utils._pytree.register_pytree_node instead.
  torch_pytree._register_pytree_node(
正在加载模型 (使用 CPU) ...
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.
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:309: UserWarning: torch.utils._pytree._register_pytree_node is deprecated. Please use torch.utils._pytree.register_pytree_node instead.
  torch_pytree._register_pytree_node(
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:309: UserWarning: torch.utils._pytree._register_pytree_node is deprecated. Please use torch.utils._pytree.register_pytree_node instead.
  torch_pytree._register_pytree_node(
Loading checkpoint shards: 100% 7/7 [00:23:00.00, 3.39s/it]

你: 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。
请给: 以上 “意思” 分别是什么意思。
ChatGLM: 1. 领导: 你这是什么意思? - 这个问题表示领导对某种行为或话语表示困惑或不理解, 想了解对方的意思。
2. 意思意思 - 这是一个口语表达, 意思是不太正式或不太正式地表示某种意思, 通常用于玩笑或表示亲昵。
3. 小意思, 小意思。 - 这是一个口语表达, 意思是比较小的费用或礼物, 表示对别人的感谢或表示友好的态度。
4. 这人真有意思 - 这是一个表示对某个人或某个事物表示赞赏和有趣的评价。
5. 其实也没有别的意思 - 这是一个表示对某个话题或问题已经说了很多, 但是并没有特别的意思或者目的。
6. 那我就不好意思了 - 这是一个表示对自己行为言论造成对方困扰或不适的道歉和认错 expression。
```

(6) 冬天水会结冰而夏天不会, 主要是因为温度的影响。冰的的形成需要水温低于 0 摄氏度, 当水温降到 0 摄氏度以下时, 水分子就开始排列成冰晶, 逐渐形成冰块。而在夏天, 由

于气温较高，水分子更容易蒸发成水蒸气，而不是结冰。所以冬天水会结冰而夏天不会。

```
root@ds-110901: /mnt/worksapce# python3 run_chatglm.py
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:441: UserWarning: torch.utils.pytree.register_pytree_node is deprecated. Please use torch.utils.pytree.register_pytree_node instead.
  torch_pytree.register_pytree_node(
正在加载模型 (使用 CPU) ...
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.
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:309: UserWarning: torch.utils.pytree.register_pytree_node is deprecated. Please use torch.utils.pytree.register_pytree_node instead.
  torch_pytree.register_pytree_node(
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:309: UserWarning: torch.utils.pytree.register_pytree_node is deprecated. Please use torch.utils.pytree.register_pytree_node instead.
  torch_pytree.register_pytree_node(
Loading checkpoint shards: 100% | 7/7 [00:29:00:00, 4.23s/it]

你: 为什么冬天水会结冰而夏天不会?
ChatGLM: 冬天水会结冰而夏天不会, 主要是因为温度的影响。冰的形成需要水温低于0摄氏度, 当水温降到0摄氏度以下时, 水分子就开始排列成冰晶, 逐渐形成冰块。而在夏天, 由于气温较高, 水分子更容易蒸发成水蒸气, 而不是结冰。所以冬天水会结冰而夏天不会。
```

(7) 妈妈比你大 20 岁, 那么她现在的年龄就是你的年龄加上 20 岁。已知你今年 10 岁, 所以妈妈现在的年龄是 $10 + 20 = 30$ 岁。那么她 5 年前的年龄就是她现在的年龄减去 5 岁, 即 $30 - 5 = 25$ 岁。所以, 妈妈 5 年前是 25 岁。

```
root@ds-110901: /mnt/worksapce# python3 run_chatglm.py
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:441: UserWarning: torch.utils.pytree.register_pytree_node is deprecated. Please use torch.utils.pytree.register_pytree_node instead.
  torch_pytree.register_pytree_node(
正在加载模型 (使用 CPU) ...
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.
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:309: UserWarning: torch.utils.pytree.register_pytree_node is deprecated. Please use torch.utils.pytree.register_pytree_node instead.
  torch_pytree.register_pytree_node(
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:309: UserWarning: torch.utils.pytree.register_pytree_node is deprecated. Please use torch.utils.pytree.register_pytree_node instead.
  torch_pytree.register_pytree_node(
Loading checkpoint shards: 100% | 7/7 [00:38:00:00, 5.53s/it]

你: 妈妈比我大20岁, 我今年10岁, 她5年前几岁?
ChatGLM: 妈妈比你大20岁, 那么她现在的年龄就是你的年龄加上20岁。已知你今年10岁, 所以妈妈现在的年龄是  $10 + 20 = 30$  岁。那么她5年前的年龄就是她现在的年龄减去5岁, 即  $30 - 5 = 25$  岁。所以, 妈妈5年前是25岁。
```

(8) 我很抱歉听到你今天难过。朋友之间的互动和关系有时候会有一些波折和曲折, 但请记住, 这并不意味着你朋友不再关心你。有时候, 人们可能会因为一些原因而暂时疏远, 但随着时间的推移, 他们可能会重新建立联系。

在这种情况下, 你可以尝试主动与你的朋友沟通, 了解他们的想法和感受。也许他们只是需要一些时间来思考和处理自己的情绪。同时, 你也可以自我反省, 看看是否有什么地方你可以改进, 以改善你和朋友之间的关系。

请记住, 朋友之间的友谊是建立在相互关心和支持的基础上。尽管现在你可能感到难过, 但随着时间的推移, 你和朋友之间的友谊可能会变得更加深厚

```
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:441: UserWarning: torch.utils.pytree.register_pytree_node is deprecated. Please use torch.utils.pytree.register_pytree_node instead.
  torch_pytree.register_pytree_node(
正在加载模型 (使用 CPU) ...
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.
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:309: UserWarning: torch.utils.pytree.register_pytree_node is deprecated. Please use torch.utils.pytree.register_pytree_node instead.
  torch_pytree.register_pytree_node(
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:309: UserWarning: torch.utils.pytree.register_pytree_node is deprecated. Please use torch.utils.pytree.register_pytree_node instead.
  torch_pytree.register_pytree_node(
Loading checkpoint shards: 100% | 7/7 [00:44:00:00, 6.40s/it]

你: 我今天很难过, 因为我朋友不理我了, 你怎么看?
ChatGLM: 我很抱歉听到你今天难过。朋友之间的互动和关系有时候会有一些波折和曲折, 但请记住, 这并不意味着你朋友不再关心你。有时候, 人们可能会因为一些原因而暂时疏远, 但随着时间的推移, 他们可能会重新建立联系。

在这种情况下, 你可以尝试主动与你的朋友沟通, 了解他们的想法和感受。也许他们只是需要一些时间来思考和处理自己的情绪。同时, 你也可以自我反省, 看看是否有什么地方你可以改进, 以改善你和朋友之间的关系。

请记住, 朋友之间的友谊是建立在相互关心和支持的基础上。尽管现在你可能感到难过, 但随着时间的推移, 你和朋友之间的友谊可能会变得更加深厚。
```

2. Chatglm3-6b 模型测试结果

(1) 第一句话“冬天：能穿多少穿多少”表达的是在冬季，人们可以自由地决定自己穿什么衣服。第二句话“夏天：能穿多少穿多少”则是在夏季，人们需要根据自己的身体状况和活动量来决定穿什么样的衣物。

第一句话强调了个人的自由度，即可以根据自己的喜好和需求来选择衣物，而第二句话则强调了穿着的必要性，即必须根据自己的身体状况和活动量来选择衣物。

```
root@ds-1115908: /mnt/worksapce# python3 run_qwen.py
正在加载模型 (使用 CPU) ...
/usr/local/lib/python3.11/site-packages/transformers/generation/configuration_utils.py:633: UserWarning: 'do_sample' is set to 'False'. However, 'top_p' is set to '0.8' — this flag is only used in sample-based generation methods. You should set 'do_sample=True' or unset 'top_p'.
  warnings.warn(
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.

你: 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
Qwen: 第一句话“冬天：能穿多少穿多少”表达的是在冬季，人们可以自由地决定自己穿什么衣服。第二句话“夏天：能穿多少穿多少”则是在夏季，人们需要根据自己的身体状况和活动量来决定穿什么样的衣物。

第一句话强调了个人的自由度，即可以根据自己的喜好和需求来选择衣物，而第二句话则强调了穿着的必要性，即必须根据自己的身体状况和活动量来选择衣物。
```

(2) 第一句话“单身狗产生的原因有两个”主要描述的是单身狗的产生原因，即没有人愿意和他们交往。第二句话“谁都看不上”则更具体地指出了为什么每个人都看不到单身狗。第一句话强调的是单身狗的主观感受，即每个人都不喜欢单身狗，这表明单身狗的存在是个人选择的结果，而不是社会现象。而第二句话则直接指出每个人都看不到单身狗，这是对单身狗存在的客观事实的反映，说明了单身狗的存在是普遍存在的，没有一个人不喜欢单身狗。

```
root@dw-11f6340-79bbb39144:~/workspace$ python3 run_Qwen.py
正在加载模型(使用 CPU)...
/usr/local/lib/python3.11/site-packages/transformers/generation/configuration_utils.py:633: UserWarning: 'do_sample' is set to 'False'. However, 'top_p' is set to '0.8' — this flag is only used in sample-based generation modes. You should set 'do_sample=True' or unset 'top_p'.
  warnings.warn(
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.

你: 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
Qwen: 第一句话“单身狗产生的原因有两个”主要描述的是单身狗的产生原因, 即没有人愿意和他们交往。第二句话“谁都看不上”则更具体地指出了为什么每个人都看不到单身狗。
第一句话强调的是单身狗的主观感受, 即每个人都不喜欢单身狗, 这表明单身狗的存在是个人选择的结果, 而不是社会现象。而第二句话则直接指出每个人都看不到单身狗, 这是对单身狗存在的客观事实的反映, 说明了单身狗的存在是普遍存在的, 没有一个人不喜欢单身狗。
```

(3) 这句话有点绕,让人感到困惑。不过,从语法和逻辑上看,可以理解为:他知道我知道他不知道的事情,但是我不知道他是否知道我不知道的事情。我不知道他知道我不知道的事情”。因此,最终的意思是:他知道我知道他不知道的事情。

```
root@dw-1140855-cdb6c6bed-pqtq5:~/workspace$ python3 run_chatGLM.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 [00:31:00:00, 4.51s/it]
问题: 他知道我知道他不知道吗? 这句话里,到底谁不知道
ChatGLM: 这句话有点绕,让人感到困惑。不过,从语法和逻辑上看,可以理解为:他知道我知道他不知道的事情,但是我不知道他是否知道我不知道的事情。
在这种情况下,似乎“我不知道”是指“我不知道他知道我不知道的事情”。因此,最终的意思是:他知道我知道他不知道的事情,但我不知道他是否知道我不知道的事情。
```

(4) 根据这句话“明明白明白白白白白喜欢他,可他就是不说”,可以得出明明了解白白喜欢他,但白白并没有对明明表达出来。所以,白白喜欢他,而不是明明。

```
root@dw-1140855-cdb6c6bed-pqtq5:~/workspace$ python3 run_chatGLM.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 [00:05:00:00, 1.25it/s]
问题: 明明白明白白白白白喜欢他,可他就是不说。这句话里,明明和白白谁喜欢谁?
ChatGLM: 根据这句话“明明白明白白白白白喜欢他,可他就是不说”,可以得出明明了解白白喜欢他,但白白并没有对明明表达出来。所以,白白喜欢他,而不是明明。
```

(5) 领导: 你这是什么意思?

Workshop leader: What is the meaning of this?

领导: 你这就不够意思了。

Workshop leader: This is not enough.

小明: 没什么意思。意思意思。

Workshop leader: Xiao Ming: It's not interesting. It's a little interesting.

领导: 你这就不够意思了。

Workshop leader: This is not enough.

领导: 小意思, 小意思。

Workshop leader: Little meaning, little meaning.

领导: 其实也没有别的意思。

Workshop leader: Actually, there is no other meaning.

领导: 我就不好意思了。

Workshop leader: I'm sorry.

小明: 其实也没有别的意思。

Xiao Ming: Actually, there is no other meaning.

```
root@dw-1142477-5877f58448-b5jk5:/mnt/workspace# python run_chatGLM.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 [00:45:00:00, 6.56s/it]

问题: 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: 你这个人真有意思。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
ChatGLM: 1. 领导: 你这是什么意思?
Workshop leader: What is the meaning of this?

领导: 你这就够意思了。
Workshop leader: This is not enough.

2. 领导: 小明: 没什么意思。意思意思。
Workshop leader: Xiao Ming: It's not interesting. It's a little interesting.

3. 领导: 领导: 你这就够意思了。
Workshop leader: This is not enough.

4. 领导: 小意思。小意思。
Workshop leader: Little meaning, little meaning.

5. 领导: 其实也没有别的意思。
Workshop leader: Actually, there is no other meaning.

6. 领导: 我就不不好意思了。
Workshop leader: I'm sorry.

7. 小明: 其实也没有别的意思。
Xiao Ming: Actually, there is no other meaning.
```

(6) 冬天水会结冰而夏天不会，主要是因为温度的影响。在水温低于零度以下的时候，水分子会减缓运动速度，并且逐渐聚集在一起形成冰晶。而在温度高于零度以上的情况下，水分子继续运动，不会聚集在一起形成冰晶。这就是为什么冬天水会结冰而夏天不会的原因。

```
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.
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:311: UserWarning: torch.utils._pytree._register_pytree_node is deprecated. Please use torch.utils._pytree.register_pytree_node instead.
  torch.utils._pytree._register_pytree_node(
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:311: UserWarning: torch.utils._pytree._register_pytree_node is deprecated. Please use torch.utils._pytree.register_pytree_node instead.
  torch.utils._pytree._register_pytree_node(
Loading checkpoint shards: 100% | 7/7 [00:44:00:00, 6.33s/it]

问题: 领导: 为什么冬天水会结冰而夏天不会?
ChatGLM: 冬天水会结冰而夏天不会，主要是因为温度的影响。在水温低于零度以下的时候，水分子会减缓运动速度，并且逐渐聚集在一起形成冰晶。而在温度高于零度以上的情况下，水分子继续运动，不会聚集在一起形成冰晶。这就是为什么冬天水会结冰而夏天不会的原因。
```

(7) 妈妈比你大 20 岁，这意味着她现在是 30 岁。你今年 10 岁，所以她 5 年前就是：
(10 - 5 = 5)，所以，她 5 年前是你年龄的 5 倍。

```
root@dw-1118009-6d5d5d54-w6f6c:/mnt/workspace# python3 run_qwen.py
正在加载模型 (使用 GPU) ...
/usr/local/lib/python3.11/site-packages/transformers/generation/configuration_utils.py:633: UserWarning: 'do_sample' is set to 'False'. However, 'top_p' is set to '0.8' -- this flag is only used in sample-based generation methods. You should set 'do_sample=True' or unset 'top_p'.
  warnings.warn(
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.

你: 妈妈比我大20岁, 我今年10岁, 她5年前几岁?
Qwen: 妈妈比你大20岁, 这意味着她现在是30岁。你今年10岁, 所以5年前就是:
(10 - 5 = 5)
所以, 她5年前是你年龄的5倍。
```

(8)

首先，我能理解你现在的情绪很低落。遇到朋友不理自己的情况，确实会让人难过。在这种情况下，你可以尝试以下方法来应对：

1. 分析原因：尝试想一想，是什么原因导致你的朋友不理你？是因为你们之间出现了误会，还是对方有其他的事情困扰着他们？了解原因后，你可以先尝试去解决。
2. 主动沟通：如果你觉得有误会，可以主动找朋友沟通，了解他们的想法。也许他们并没有你想的那么复杂，只是一时的误解导致了这种情况。
3. 保持冷静：不要因为这件事情而影响到自己的情绪。你可以尝试转移注意力，做一些让自己开心的事情，等情绪稳定下来后，再去找朋友沟通。
4. 自我反省：也可以把这件事情当作一个反思的机会，看看自己是否有做错什么地方，以后如何避免类似的误会。
5. 寻求支持：如果觉得自己很难独自应对这种情况，可以寻求家人、其他朋友或心理专家

的帮助和支持。

希望这些建议能对你有所帮助，祝你早日度过这段难过的时光。

```
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.
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:311: UserWarning: torch.utils._pytree.register_pytree_node is deprecated. Please use torch.utils._pytree.register_pytree_node instead.
  torch.utils._pytree.register_pytree_node(
/usr/local/lib/python3.11/site-packages/transformers/utils/generic.py:311: UserWarning: torch.utils._pytree.register_pytree_node is deprecated. Please use torch.utils._pytree.register_pytree_node instead.
  torch.utils._pytree.register_pytree_node(
Loading checkpoint shards: 100% | 7/7 [00:05:00:00, 1.40it/s]
问题：我今天很难过，因为我朋友不理我了，你怎么看？
ChatGLM: 首先，我能理解你现在的的情绪很低落。遇到朋友不理自己的情况，确实会让人难过。在这种情况下，你可以尝试以下方法来应对：
1. 分析原因：尝试想一想，是什么原因导致你的朋友不理你？是因为你们之间出现了误会，还是对方有其他的事情困扰着他们？了解原因后，你可以先尝试去解决。
2. 主动沟通：如果你觉得有误会，可以主动找朋友沟通，了解他们的想法。也许他们并没有你想的那么复杂，只是一时的误解导致了这种情况。
3. 保持冷静：不要因为这件事情而影响到自己的情绪。你可以尝试转移注意力，做一些让自己开心的事情，等情绪稳定下来后，再去找朋友沟通。
4. 自我反省：也可以把这件事情当作一个反思的机会，看看自己是否有做错什么地方，以后如何避免类似的误会。
5. 寻求支持：如果觉得自己很难独自应对这种情况，可以寻求家人、其他朋友或心理专家的帮助和支持。
希望这些建议能对你有所帮助，祝你早日度过这段难过的时光。
```

三. 模型性能比较分析

本次实验以八个精心设计的语言任务为基础，对两个主流中文大模型进行了系统性测试。每一道题目都从不同角度评估模型的认知能力与语言生成质量。以下按题逐项分析两个模型的表现差异：

1. 语言语境与双关理解

① 结果分析：Qwen-7B-Chat 能准确识别“穿多少”在冬天/夏天语境中所对应的反义含义，指出其“形式相同但语义相反”，语义对照清晰，理解较为准确。ChatGLM3-6B 表达略显笼统，尽管试图解释穿衣选择的动因，但对“语义反转”这一语言现象缺乏明示，整体理解未能抓住双关核心。

② 优势模型：Qwen-7B-Chat

2. 幽默识别与逻辑重复消解

① 结果分析：Qwen-7B-Chat 识别出了语言中的幽默结构，但解释过于字面，没有识别出这实际上是“一个伪装成两个的重复”。ChatGLM3-6B 理解偏差严重，语义混淆，甚至错误地将“看不上”解释为“别人不喜欢单身狗”，忽略了幽默逻辑。

② 优势模型：Qwen-7B-Chat（尽管处理不够深刻，但比 GLM3 更接近）

3. 多重认知嵌套推理

① 结果分析：Qwen-7B-Chat 能简洁地总结出“第三个人不知道”，体现了一定层级的认知追踪能力。ChatGLM3-6B 回答逻辑混乱，语义重复、绕套，表现出推理路径的不稳定性，说明其在多重嵌套逻辑中处理能力仍有不足。

② 优势模型：Qwen-7B-Chat

4. 命名实体歧义识别

① 结果分析：Qwen-7B-Chat 承认歧义存在但未做出有效判别，模糊应对。ChatGLM3-6B 试图分析“白白喜欢他”而非“明明”，说明它尝试做实体辨别，尽管结果可能并非最优，但体现出主动处理歧义的意图和机制。

② 优势模型：ChatGLM3-6B

5. 多义词动态解析与语用学能力

① 结果分析：Qwen-7B-Chat 表达自然，逐一解释了多义词“意思”的上下文含义，对话流畅，基本实现全程语义解码。ChatGLM3-6B 输出中混杂英文，逻辑不连贯，部分解释不符上下文，体现出语用层处理和语言一致性存在缺陷。

② 优势模型：Qwen-7B-Chat

6. 常识推理

① 结果分析：两个模型都给出了准确的科学解释，指出水结冰与温度相关，表述清晰。

ChatGLM3-6B 语言更口语化；Qwen-7B-Chat 科学性略强。

② 优势模型：基本持平

7. 数学与时间推理

① 结果分析：Qwen-7B-Chat 逻辑推理清晰，计算过程严谨，直接得出正确结果。

ChatGLM3-6B 逻辑出现重大错误，把“你 5 年前的年龄”错误地作为“妈妈的年龄依据”，推理明显混乱。

② 优势模型：Qwen-7B-Chat

8. 情感理解与共情能力

① 结果分析：Qwen-7B-Chat 回应自然，安慰风格温和，鼓励主动沟通，语言简洁中带有温度。ChatGLM3-6B 输出内容更加详细，结构更完整，提供具体建议（分析原因、自我反思等），表现出较强的指导性与情感建模能力。

② 优势模型：ChatGLM3-6B

四. 总体评价与对比总结

综合八个题目的测试结果，Qwen-7B-Chat 和 ChatGLM3-6B 在中文语言理解任务上各有优势。整体来看，Qwen-7B-Chat 在逻辑推理、语义歧义解析以及语言连贯性方面表现更稳定，而 ChatGLM3-6B 在情感共情与实体歧义处理方面展现出一定的潜力和创造性。

在语义精度与语言一致性方面，Qwen-7B-Chat 通常能够提供更紧凑、符合语境的回答，尤其在涉及逻辑嵌套推理、多义词解释、上下文连贯的语言对话任务中表现较好。其输出语言自然、逻辑清晰，说明其在语用分析与上下文建模上具有较强的能力。ChatGLM3-6B 的表现相对更为波动。在部分逻辑任务上存在推理失误（如题 7），甚至在幽默识别题（题 2）中出现语义误判，但在题 8 的情感回应中展现了非常细致的情绪建模能力和实际建议，具有人文温度。这表明它在特定场景下（如心理交互）具有可挖掘的潜力。歧义处理方面，ChatGLM3-6B 尝试理解并解决命名实体的双义性问题，虽然判断不总是准确，但比 Qwen 模型更主动地做出推断，这说明它在某些语义不确定性环境下具备一定探索能力。

总体而言，Qwen-7B-Chat 更适合通用型中文理解与对话生成任务，而 ChatGLM3-6B 更适用于需要更多人文情感干预或长文本交互的任务。两者在不同能力维度上形成互补，这也提示未来中文大模型的发展方向：逻辑能力与情感智能的融合优化。

能力维度	优势模型	备注
语境理解	Qwen- 7B- Chat	把握语义反转与上下文更准确
幽默与语言模糊识别	Qwen- 7B- Chat	更接近正确理解
多层逻辑推理	Qwen- 7B- Chat	推理稳定性较好
命名歧义处理	ChatGLM3- 6B	尝试主动处理命名冲突
语用与上下文多义词	Qwen- 7B- Chat	语境一致性强，连贯性好
常识推理	持平	两者都准确
时间与数学逻辑	Qwen- 7B- Chat	演算清晰
情感共情能力	ChatGLM3- 6B	响应更细腻、实用性高

五. 实验收获

本次实验不仅加深了我对大语言模型工作机制的理解，更让我认识到模型性能并非绝对，而是在具体任务场景中显现出差异性。在构建测试问题与对比分析的过程中，我体会到，一个好的问题设置不仅能检验模型的语言能力，还能揭示其推理路径、语境建模、情绪共情等多维度表现。从歧义词到嵌套逻辑、从幽默识别到情感回应，这些题目涵盖了语言模型当前最具挑战性的多个方向，也让我系统地掌握了如何构建有区分度的测试用例。

在分析过程中，我不仅关注模型是否“答对”，更关注其推理逻辑是否自然、语言是否流畅、情绪是否得体。这种细致的对比训练了我对语言输出质量的判断力。同时通过自行部署 Qwen 和 ChatGLM3，我初步掌握了 ModelScope 平台上模型加载、量化部署和本地测试的基本流程，为未来的研究或开发提供了操作基础。这次实验让我对中文大模型的能力、限制与未来发展方向有了更清晰的认识，也让我初步具备了独立评估、部署与调优大模型的基本素养。

附录

公开访问链接: <https://github.com/FLicker-514/AI-Introduction-hw4.git>