

AI 对比分析报告

2353926-赵泽远

公开链接: [ai-hw4/report.pdf at main · gfjdh/ai-hw4](#)

测试问题:

- 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.明明明明白白白喜欢他，可她就是不。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。

以下是对 Qwen-7B-Chat、DeepSeek-llm-7B-Chat、ChatGLM3-6B 三个大模型在中文语言理解测试中的横向对比分析，基于五个典型问题的回答表现：

1. 语义歧义：季节歧义句解析 ("能穿多少穿多少")

Qwen: 这两句话的主要区别在于它们所描述的季节不同。第一句话冬天：能穿多少穿多少"强调的是在寒冷的冬

季，人们应该尽可能地多穿衣服以保暖。而第二句话夏天：能穿多少穿多少则是在炎热的夏季，人们应该选择轻薄透气的衣服来避免过度出汗和不适感。因此，这两句话所传达的信息是完全不同的。

Deepseek: 冬天：能穿多少穿多少（可以理解为多穿点衣服保暖，根据自己身体情况来调节衣服的厚度。）

夏天：能穿多少穿多少（可以理解为少穿点衣服降温，同样根据自己身体情况来调节衣服的厚度。）

Chatglm3: 这两句话的区别在于所处的季节不同。第一句是关于冬天的穿着建议，表示在冬天应该尽可能多穿衣服，以保暖；而第二句是关于夏天的穿着建议，同样表示在夏天应该尽可能多穿衣服，以避免炎热。这两句话所针对的季节不同，但传达的意思相同，即要根据所处的季节选择适当的穿搭方式。

模型	正确性	表现
Qwen	✓	准确区分冬夏语境：冬天"多穿"保暖 vs 夏天"少穿"降温
DeepSeek	✓	简洁点出核心：强调"根据身体调节厚度"，冬夏目标相反
ChatGLM3	✗	严重错误：认为两句话"传达意思相同"（未识别歧义）

2. 语义歧义：同词异义幽默 ("谁都看不上")

Qwen: 这两句话的主要区别在于句子的语气和重点。第一句话使用了他"来强调产生单身狗的原因有很多，但并没有明确指出哪些

是主要原因；而第二句话则直接指出产生单身狗的两个主要原因是也看不上"和自己也看不上”。因此，第二句话更加具体和直接地表达了单身狗产生的原因，而第一句话则较为模糊和概括。

Deepseek: 第一句和第二句的区别在于看不上"的意思和用法不同。在第一句中，看不上指的是看不起”，表示对某人或某事持有轻视、鄙视的态度，认为对方不够好、不够优秀。因此，第一句的意思是单身狗产生的原因有两个，一是谁都看不上，即单身狗认为周围的人都不够优秀，不符合自己的标准；二是谁都看不上，即单身狗对周围的人都有轻视、鄙视的态度。在第二句中，看不上”指的是不想看”，表示对某人或某事没有兴趣，不想关注或接触。因此，第二句的意思是单身狗产生的原因有两个，一是谁都看不上，即单身狗对周围的人都没有兴趣，不想去了解或交往；二是谁都看不上，即单身狗对周围的人都没有兴趣，不想去了解或交往。因此，两句话的区别在于看不上"的用法和意义不同。

Chatglm3: 这两句话都描述了单身狗的情况，但使用了不同的修辞手法。第一句话使用了“是谁都看不上”，强调单身狗认为所有人都存在一些不足之处，因此他们难以找到合适的伴侣。这种说法暗示了单身狗对自己要求过高，对伴侣期望过高，从而导致他们一直单身。第二句话使用了唯都看不上”，表达单身狗对自己的不满和失望。他们认为自己在某些方面存在不足，因此难以得到别人的认可。这种说法强调的是单身狗对自己缺乏自信和自尊心，从而导致他们难以建立健康的人际关系。总之，这两句话都描述了单身狗的困境，但第一句话更强调他们对自己和伴侣的要求过高，而第二句话更强调他们对自己缺乏自信和自尊心。

模型	正确性	表现
Qwen	✗	混淆结构：将两句话拆解为"模糊概括"和"具体原因", 未抓住核心双关
DeepSeek	✓	明确区分"看不上=看不起"（挑剔他人）， "看不上=无人青睐"
ChatGLM3	✗	过度解读：用"自尊心""要求过高"等心理分析替代语言双关本质

3. 逻辑推理：多层逻辑嵌套 ("谁不知道")

Qwen: 选项：A.他不知道 B.我不知道 C.你知道
首先，我们需要理解这句话的意思。它的意思是他知道我知道你知道他不知道吗？”也就是说，四个人中，有两个人知道某个事情，另一个人知道另外两个人知道的事情，而最后一个人则不知道全部的事情。
所以，我们可以将这个句子简化为以下形式：
知道的人数：4-1=3
未知的人数：1
根据这个简化后的结构，我们就可以确定，选项 A 和 B 都是错误的，因为它们描述了知道的人数（2 个或 0 个）而不是正确的人数（3 个）。因此，正确的答案应该是选项 C，即“你知道”。
Deepseek: 他知道我知道你知道他不知道吗？这句话里，到底谁不知道谁啊？是不是挺绕的？这其实就是一种语言上的绕口令，也是句式上的“循环”。这种句式，通常叫做循环疑问句”。

循环疑问句的构成,一般都是这样的:

- 1.问句开始用升调,问到对方。
 - 2.如果对方不知道,就会回答"yes",如果对方知道,就会回答 No”。
 - 3. 回答了 “Yes”，问句的回答就变成 No”，回答了 No”，问句的回答就变成 Yes”。
- 比如：你吃过饭了吗？这句话就是典型的循环疑问句。如果对方不知道自己吃过饭了没有，就会回答：，我没吃过。”如果对方已经吃过饭了，就会回答：我吃过了。”

循环疑问句的作用,是为了让对方更好地理解自己的意思,也让自己更好地理解对方的回答。在英语中,循环疑问句通常用在礼貌的场合,比如询问别人的身份、职业、年龄等。需要注意的是,循环疑问句在口语中比较常见,但是在正式的场合中使用,可能会显得有些不够正式。

Chatglm3: 这句话是一个经典的谜语或者绕口令,它的含义是: 一个人知道另外一个人知道某个秘密,但这个人自己并不知道这个秘密。因此,在这个句子中,三个人都知道这个秘密,但只有第三个人不知道自己知道这个秘密。

模型	正确性	表现
Qwen	✗	完全错误: 强行数学化 (4-1=3), 得出"你知道"的荒谬结论
DeepSeek	✗	回避问题: 大篇幅解释"循环疑问句"概念, 未直接回答"谁不知道"
ChatGLM3	✓	唯一正确: 指出"第三个人 (他) 不知道自己知道秘密"的嵌套逻辑

4. 逻辑推理：人名关系 ("明明""白白"谁喜欢谁)

Qwen: 这句话的意思是: 明明很明白自己喜欢白白, 但是她就是不说出来。因此, 明明喜欢白白。

Deepseek: 明明明明明白白白喜欢他, 可她就是不。这句话里, 明明和白白谁喜欢谁? 显然是明明喜欢白白, 对不对? 那如果我说这句话的完整版是 “明明明明明白白白喜欢他, 可是她就是不说。” 那这句话就是明明白白喜欢他, 对吗? 所以, 我们不要去纠结这句话里的 “明明”和 “白白 “谁喜欢谁, 因为两个版本都在表达同样的意思。就如同这个故事一样, 我们不要去纠结它是否真的有意义, 因为每个读者都会有自己的解读。所以, 不要怕别人否定你的想法, 因为每个人都是独一无二的, 你的想法也是独一无二的。无论别人怎么想, 只要你觉得有意义, 那就足够了。

Chatglm3: 这句话中, 明明是知道白白喜欢他, 但是明明并没有表达出自己的感情。因此, 可以理解为明明对白白有感情, 但是没有表达出来。

模型	正确性	表现
Qwen	✓	正确解析: 明明明白 (白白喜欢他) → 白白喜欢"他"
DeepSeek	✗	偷换概念: 将问题扭曲为"版本解读差异", 回避实质关系
ChatGLM3	✓	正确结论: 白白喜欢"他", 但误增"明明对白白有感情"的过度解读

5. 语境解析：多义词"意思"语境解析

- Qwen:**
- 1. “我这是什么意思？” — 这句话的意思是询问对方说话或行为的含义。
 - 2. “没什么意思。意思意思。” — 这句话的意思是没有特别的意思, 只是随便说说。
 - 3. “你这就不够意思了。” — 这句话的意思是对对方的行为或言语表示不满, 认为不够礼貌或尊重。
 - 4. “你这人真有意思。” — 这句话的意思是对对方的行为或言语感到有趣或好奇。
 - 5. “其实也没有别的意思。” — 这句话的意思是否认之前说过的话有其他意思, 只是单纯的

解释。

6. “那我就不好意思了。” —这句话的意思是由于自己的不当言行感到尴尬或歉意。

7. “是我不好意思。” —这句话的意思是承认自己的错误或过失。

Deepseek: 解析：这段对话中，意思”一词出现了多次，意思分别是：

1. “你这是什么意思？”中，“意思”指的是用意、意图、目的等。

2. “小明：没什么意思。意思意思。”中，“意思意思”指的是表示一下，有讨好、奉承的意思。

3. “领导：你这就不够意思了。”中，“不够意思”指的是不够意思，即不够朋友，不够义气。

4. “小明：小意思，小意思。”中，“小意思”指的是小事，不必放在心上。

5. “领导：你这人真有意思。”中，“有意思”指的是有趣，有吸引力。

6. “其实也没有别的意思。”中，“别的意思”指的是除了这个之外的其他意思，即无他意。

7. “不好意思。”中，“不好意思”指的是感到不好意思，即感到羞愧或不好意思。

8. “问我：那我就不好意思了。”中，“不好意思”指的是感到不好意思，即感到羞愧或不好意思。

9. “我：是我不好意思。”中，“是我不好意思”指的是感到羞愧或不好意思的是我。

Chatglm3: 1、“意思”作为名词，可以表示含义、意义、意思、意蕴等。在这个语境中，领导问小明称这是什么意思”，意思是询问小明网钢的行为或言论的含义或目的

2.“意思”作为动词，可以表示表示、暗示、意味等。在这个语境中，小明用意思意思”表示自己的行为并没有特定的含义或目的，只是随口说而已。

3.“意思”作为动词，可以表示含义、意味等。在这个语境中，小明用小意思，小意思”表示自己行为的目的是为了表达敬意或感激之情。

4.“意思”作为形容词，可以表示意义、含义等。在这个语境中，领导认为小明这个人很有趣，因为他能够从小明的话语中意味出很多含义。

5.“意思”作为名词，也可以表示意味、意味等。在这个语境中，小明用其实也没有别的意思”表示自己并没有其他特定的目的或含义。

模型	正确性	表现
Qwen	✗	漏掉 2 处：未解析"意思意思"（送礼）和最后"不好意思"（道歉）
DeepSeek	✓	覆盖全部 9 处，精准辨别"意思意思=奉承""不好意思=羞愧"等
ChatGLM3	✗	混淆词性：将动词（如"意思意思"）错误归类为名词，漏解关键语境

综合能力对比分析

模型	逻辑推理	语义歧义	语境解析	总得分
Qwen	●●○○○	●●●○○	●●●○○	8
DeepSeek	●○○○○	●●●●●	●●●●●	11
ChatGLM3	●●●●●	●○○○○	●●●○○	9

核心优势总结：

DeepSeek: 中文多义词解析强，在歧义句、多义词场景下表现近乎完美，但逻辑推理较弱

ChatGLM3: 复杂逻辑推理能力突出，但语义和文化理解较弱。

Qwen: 表现均衡但无突出项。

附录 1：环境配置

```
# 创建新环境
python -m venv chatglm_env
# 激活环境
source chatglm_env/bin/activate
# 安装基础依赖
pip install torch==2.3.1 torchvision==0.18.1 torchaudio==2.3.1 --index-url
https://download.pytorch.org/whl/cu118
pip install transformers==4.38.2
pip install sentencepiece accelerate
```

下载模型

```
cd /mnt/data
# 清华智谱系列
git clone https://www.modelscope.cn/ZhipuAI/chatglm3-6b.git
# 千问系列
git clone https://www.modelscope.cn/qwen/Qwen-7B-Chat.git
# 深度求索 (DeepSeek)
git clone https://www.modelscope.cn/deepseek-ai/deepseek-llm-7b-chat.git
```

创建运行脚本

```
# 进入工作目录
cd /mnt/workspace
# 创建推理脚本（对应模型命名）
vim run_xxx.py
```

修改 python 脚本内容

```
from transformers import TextStreamer, AutoTokenizer, AutoModelForCausalLM

model_path = "/mnt/data/Qwen-7B-Chat" # 可以根据需要更换模型
prompt = "这里填问题"
tokenizer = AutoTokenizer.from_pretrained(model_path, trust_remote_code=True)
model = AutoModelForCausalLM.from_pretrained(
    model_path,
    trust_remote_code=True,
    torch_dtype="auto" # 自动选择 float32/float16
).eval()
inputs = tokenizer(prompt, return_tensors="pt").input_ids
streamer = TextStreamer(tokenizer)
model.generate(inputs, streamer=streamer, max_new_tokens=300)
```

运行脚本

```
python run_xxx.py
```

附录 2：截图

Qwen

```
root@dsw-1133167-7f9bfd96b6-4w9zr:/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 | 550.00 KiB/s, 完成.
处理 delta 中: 100% (294/294), 完成.
过滤内容: 50% (4/8), 5.59 GiB | 58.68 MiB/s
过滤内容: 100% (8/8), 14.38 GiB | 50.84 MiB/s, 完成.
```

Qwen git clone 截图

Qwen 问答截图：

```
(qwen_env) root@dsw-1133038-7b8f49cf77-hzf6d:/mnt/workspace# python run_qwen_cpu.py
/opt/conda/envs/qwen_env/lib/python3.10/site-packages/transformers/utils/generic.py:311: Use
rWarning: torch.utils._pytree._register_pytree_node is deprecated. Please use torch.utils._p
ytree.register_pytree_node instead.
  torch.utils._pytree._register_pytree_node(
```

Loading checkpoint shards: 100%

8/8 [00:37<00:00, 4.63s/it]

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。

这两句话的主要区别在于它们所描述的季节不同。第一句话‘冬天：能穿多少穿多少’强调的是在寒冷的冬季，人们应该尽可能地多穿衣服以保暖。而第二句话‘夏天：能穿多少穿多少’则是在炎热的夏季，人们应该选择轻薄透气的衣服来避免过度出汗和不适感。因此，这两句话所传达的信息是完全不同的。</endof

Loading checkpoint shards: 100%

8/8 [00:08<00:00, 1.04s/it]

回答：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，二是谁都看不上自己。这两句话的主要区别在于句子的语气和重点。第一句话使用了‘他’来强调产生单身狗的原因有很多，但并没有明确指出哪些是主要原因；而第二句话则直接指出产生单身狗的两个主要原因是‘谁也看不上’和‘自己也看不上’。因此，第二句话更加具体和直接地表达了单身狗产生的原因，而第一句话则较为模糊和概括。</endof

Loading checkpoint shards: 100%

8/8 [00:08<00:00, 1.03s/it]

回答：3. 他知道我知道你知道他不知道吗？这句话里，到底谁不知道呢？

选项：

- A. 他不知道
- B. 我不知道
- C. 你知道

首先，我们需要理解这句话的意思。它的意思是‘他知道我知道你知道他不知道吗？’也就是说，四个人中，有两个人知道某件事情，另一个人知道另外两个人知道的事情，而最后一个人则不知道全部的事情。

所以，我们可以将这个句子简化为以下形式：

知道的人数：4 - 1 = 3

未知的人数：1

根据这个简化后的结构，我们就可以确定，选项 A 和 B 都是错误的，因为它们描述了知道的人数（2个或0个）而不是正确的人数（3个）。因此，正确的答案应该是选项 C，即‘你知道’。</endof


```
Loading checkpoint shards: 100% | 2/2 [00:04<00:00, 2.01s/it]
回答: Setting `pad_token_id` to `eos_token_id`:100001 for open-end generation.
<|begin_of_sentence|>明明明明白白喜欢他，可她就是不说话。这句话里，明明和白白谁喜欢谁？显然是明明喜欢白白，对不对？那如果我说这句话的完整版是“明明明明白白喜欢他，可她就是不说话。”那这句话就是明明白白喜欢他，对吗？所以，我们不要去纠结这句话里的“明明”和“白白”谁喜欢谁，因为两个版本都在表达同样的意思。就如同这个故事一样，我们不要去纠结它是否真的有意义，因为每个读者都会有自己的解读。所以，不要怕别人否定你的想法，因为每个人都是独一无二的，你的想法也是独一无二的。无论别人怎么想，只要你觉得有意义，那就足够了。<|end_of_sentence|>root@dsw-1133167-7f9bfd96b6-4w9zr:/mnt/workspace#
```

```
Loading checkpoint shards: 100% | 2/2 [00:04<00:00, 2.05s/it]
回答: Setting `pad_token_id` to `eos_token_id`:100001 for open-end generation.
<|begin_of_sentence|>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。
解析：这段对话中，“意思”一词出现了多次，意思分别是：
1. “你这是什么意思？”中，“意思”指的是用意、意图、目的等。
2. “小明：没什么意思。意思意思。”中，“意思意思”指的是表示一下，有讨好、奉承的意思。
3. “领导：你这就不够意思了。”中，“不够意思”指的是不够意思，即不够朋友，不够义气。
4. “小明：小意思，小意思。”中，“小意思”指的是小事，不必放在心上。
5. “领导：你这人真有意思。”中，“有意思”指的是有趣，有吸引力。
6. “其实也没有别的意思。”中，“别的意思”指的是除了这个之外的其他意思，即无他意。
7. “不好意思。”中，“不好意思”指的是感到不好意思，即感到羞愧或不好意思。
8. “问我：那我就不好意思了。”中，“不好意思”指的是感到不好意思，即感到羞愧或不好意思。
9. “我：是我不好意思。”中，“是我不好意思”指的是感到羞愧或不好意思的是我。<|end_of_sentence|>root@dsw-1133167-7f9bfd96b6-4w9zr:/mnt/workspace#
```

Chatglm3

```
root@dsw-1133167-7f9bfd96b6-4w9zr:/mnt/workspace# 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 | 579.00 KiB/s, 完成。
处理 delta 中: 100% (60/60), 完成。
过滤内容: 46% (7/15), 10.75 GiB | 6.43 MiB/s
过滤内容: 100% (15/15), 23.26 GiB | 52.45 MiB/s, 完成。
root@dsw-1133167-7f9bfd96b6-4w9zr:/mnt/workspace#
```

Chatglm3 git clone 截图

Chatglm3 问答截图：

```
Loading checkpoint shards: 100% | 7/7 [00:02<00:00, 2.90it/s]
回答: 这两句话的区别在于所处的季节不同。第一句是关于冬天的穿着建议，表示在冬天应该尽可能多穿衣服，以保暖；而第二句是关于夏天的穿着建议，同样表示在夏天应该尽可能多穿衣服，以避免炎热。这两句话所针对的季节不同，但传达的意思相同，即要根据所处的季节选择适当的穿搭方式。
```

```
Loading checkpoint shards: 100% | 7/7 [00:02<00:00, 3.26it/s]
回答: 这两句话都描述了单身狗的情况，但使用了不同的修辞手法。
```

第一句话使用了“是谁都看不上”，强调单身狗认为所有人都存在一些不足之处，因此他们难以找到合适的伴侣。这种说法暗示了单身狗对自己要求过高，对伴侣期望过高，从而导致他们一直单身。

第二句话使用了“谁都看不上”，表达单身狗对自己的不满和失望。他们认为自己在某些方面存在不足，因此难以得到别人的认可。这种说法强调的是单身狗对自己缺乏自信和自尊心，从而导致他们难以建立健康的人际关系。

总之，这两句话都描述了单身狗的困境，但第一句话更强调他们对自己和伴侣的要求过高，而第二句话更强调他们对自己缺乏自信和自尊心。

```
Loading checkpoint shards: 100% | 7/7 [00:02<00:00, 3.34it/s]
回答: 这句话是一个经典的谜语或者绕口令，它的含义是：一个人知道另外一个人知道某个秘密，但这个人自己并不知道这个秘密。因此，在这个句子中，三个人都知道这个秘密，但只有第三个人不知道自己知道这个秘密。
```

```
Loading checkpoint shards: 100% | 7/7 [00:02<00:00, 3.27it/s]
回答: 这句话中，明明知道白白喜欢他，但是明明并没有表达出自己的感情。因此，可以理解明明对白白有感情，但是没有表达出来。
```

```
Loading checkpoint shards: 100% | 7/7 [00:32<00:00, 4.66s/it]
回答: 1. “意思”作为名词，可以表示含义、意义、意思、意蕴等。在这个语境中，领导问小明“你这是什么意思”，意思是询问小明刚刚的行为或言论的含义或目的。
2. “意思”作为动词，可以表示表示、暗示、意味等。在这个语境中，小明用“意思意思”表示自己的行为并没有特定的含义或目的，只是随口说而已。
3. “意思”作为动词，可以表示含义、意味等。在这个语境中，小明用“小意思，小意思”表示自己行为的目的是为了表达敬意或感激之情。
4. “意思”作为形容词，可以表示意义、含义等。在这个语境中，领导认为小明这个人很有趣，因为他能够从小明的话语中意味出很多含义。
5. “意思”作为名词，也可以表示意味、意味等。在这个语境中，小明用“其实也没有别的意思”表示自己并没有其他特定的目的或含义。
```