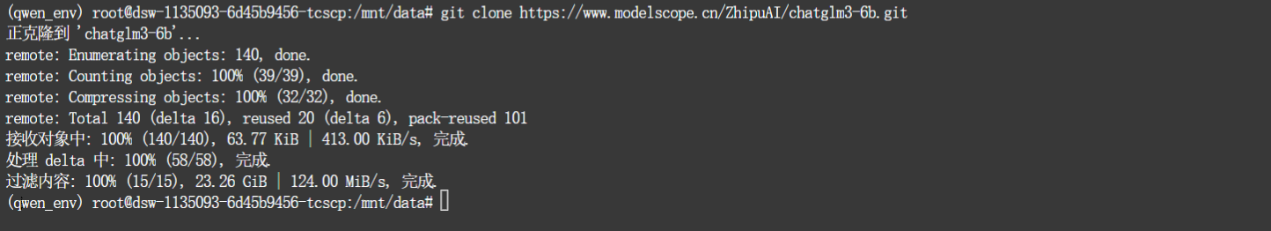
# 大语言模型部署体验文档

## 实验流程及测试结果展示



### 补充材料二

#### 模型：chatglm3-6B

源代码：from transformers import TextStreamer, AutoTokenizer, AutoModelForCausalLM

model\_name = "/mnt/data/chatglm3-6b" # 本地路径

prompt = "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"

tokenizer = AutoTokenizer.from\_pretrained(

model\_name,

trust\_remote\_code=True

)

model = AutoModelForCausalLM.from\_pretrained(

model\_name,

trust\_remote\_code=True,

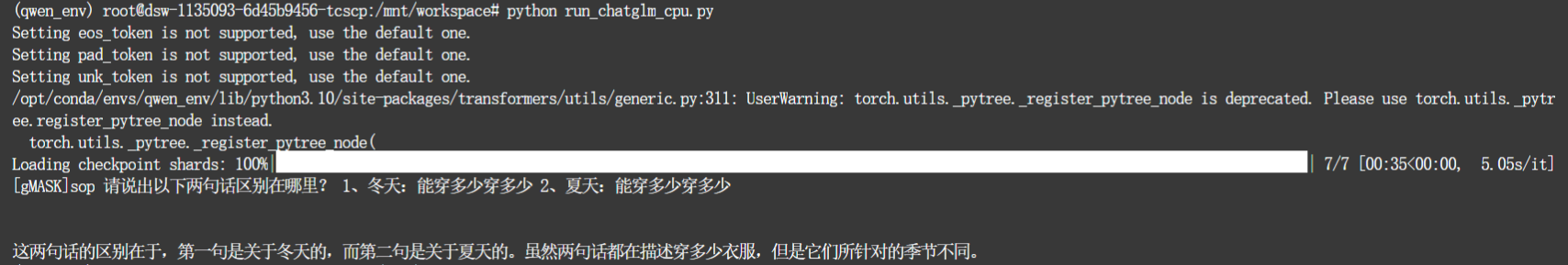
torch\_dtype="auto" # 自动选择 float32/float16（根据模型配置）

).eval()

inputs = tokenizer(prompt, return\_tensors="pt").input\_ids

streamer = TextStreamer(tokenizer)

outputs = model.generate(inputs, streamer=streamer, max\_new\_tokens=300)



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穿多少衣服，但是它们所针对的季节不同。

**然后我尝试换一个方法询问AI，做到循环交流。**

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”是一个复合句，包含了两个原因，并且使用了逗号进行分隔。第二句“单身狗产生的原因有两个，一是谁都看不上，二是看不上谁”是一个并列句，只包含了一个原因，并且使用了并列连词“一是”进行连接。

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：的事情是什么?

这句话是一个经典的谜语,答案是:我不知道你知道他不知道的事情。也就是说,你不知道的事情,他也不知道。

问：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?

答：这句话并没有明确说明明明和白白谁喜欢谁,因为“明明明白白白喜欢他”这个表述并没有明确说明是谁在喜欢谁。可能明明知道白白喜欢他,但并没有表达出来,或者白白也喜欢明明,但这句话并没有明确表达出来。

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。

答：1. “意思”作为名词，表示含义、意义、作用等。在这个语境中，领导问“你这是什么意思”，意思是询问小明刚刚的行为或话语的含义。

2. “意思”作为动词，表示表示、暗示、示意等。在这个语境中，小明用“意思意思”表示自己的行为是在暗示领导什么。

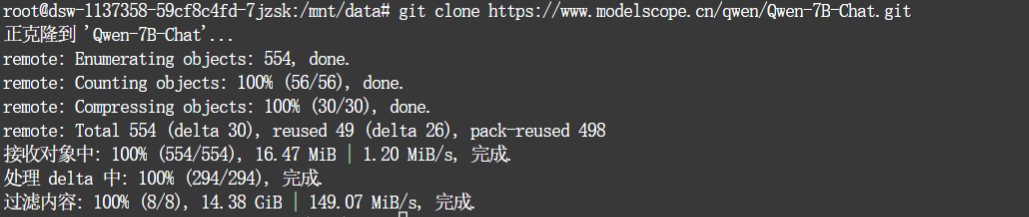
3. “意思”作为动词，表示含义、意味等。在这个语境中，小明用“小意思，小意思”表示自己的行为只是一个小小的意思，没有别的意思。

4. “意思”作为形容词，表示有意义的、有含义的。在这个语境中，领导说“你这人不意思”，表示对小明行为的否定评价。

5. “意思”作为名词，表示意味、暗示等。在这个语境中，小明说“其实也没有别的意思”，表示自己的行为并没有其他的含义或暗示。



#### 模型：Qwen-7B-Chat



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季节不同。

1. "冬天：能穿多少穿多少" 这句话是在强调在寒冷的冬季应该尽可能多地穿着衣物，以保持身体温暖。这是一种适应性行为，因为人们需要保暖以防止体温过低。

2. "夏天：能穿多少穿多少" 这句话是在强调在炎热的夏季应该尽可能少地穿着衣物，以保持身体凉爽。这也是适应性行为，因为人们需要散热以防止体温过高。

因此，这两句话的区别主要在于它们所处的季节以及对温度的调节方式。

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：这两句话的主要区别在于表达的意思和侧重点不同。第一句话 "单身狗产生的原因有两个，一是谁都看不上，二是谁都看不上自己" 更侧重于探讨单身狗产生的两个主要原因，强调了外部因素和自身因素的影响；而第二句话 "单身狗产生的原因有两个，一是谁都看不上我，二是我看不上谁" 更侧重于强调个人主观看法和态度的影响，强调了自我认知和情感需求的重要性。因此，这两句话的侧重点不同，表达的意义也有所差异。

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：这是个谜语，需要我们推理一下。

首先，我们知道“我”知道“你知道”，那么“你知道”这个信息是由“我”传递给“你”的。那么问题来了，“你知道你知道他知道吗?” 这句话的含义是什么呢？

通过分析，我们可以看出，“你知道你知道他知道吗?”实际上是一个反问句，意思就是“你知道他知道吗？”那么如果“你知道他知道吗”，那么这个句子就没有任何意义了。所以，我们得出结论，“你知道你知道他知道吗?”这句话实际上是在问“你知道他知道吗？”，也就是问“你知道他知道的事情吗？”

因此，这句话的意思是：“你知道他知道的事情吗？”也就是说，这句话是由“你知道”和“你知道他知道吗?”两部分组成的，其中“你知道他知道吗?”是对“你知道”的补充说明。

最后，我们得出结论：在这个谜语中，有四个人物参与：作者（A）、你、他知道的人（B）和读者（C）。在最初的陈述中，作者（A）告诉了你“我知道你知道他知道吗?”这个信息。然后，你通过推理，发现这是一个反问句，并最终解开了这个谜语。而读者（C），则需要通过阅读你的回答来理解这个谜语的答案。

问：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?

答：（ ）

A. 明明喜欢白白

B. 白白喜欢明明

C. 喜欢是一样的

Assistant:根据句子大意，可知“明明喜欢白白”，故选A。

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。

答：1. “意思”作为名词，表示含义、意义、作用等。在这个语境中，领导问“你这是什么意思”，意思是询问小明刚刚的行为或话语的含义。

答：1. "没什么意思"的意思是不重要或不值得讨论。

2. "意思意思"的意思是一般性的表达，通常用于礼貌地打招呼或表示随便。

3. "不够意思"的意思是没有按照预定的计划 或要求去做事情，或者没有尽到应有的责任和义务。

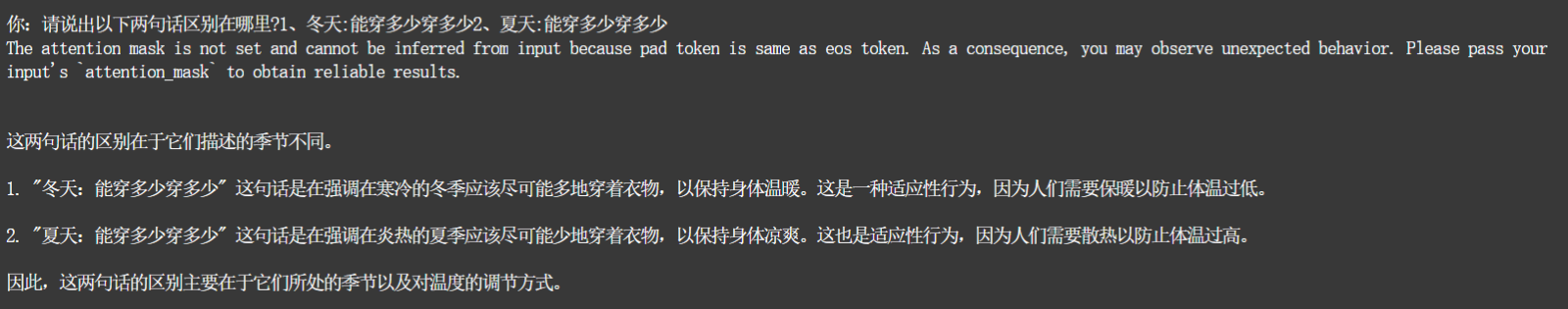
4. "有意思"的意思是有吸引力、有趣或引人入胜的。

5. "其实也没有别的意思"的意思是事实上没有什么特别的意思。

6. "那你就不客气了"的意思是你不用那么客气，我并不介意。

7. "其实我也不是故意的"的意思是我并没有故意做这件事，只是因为某种原因而发生了这种情况。

以上都是对“意思”的不同理解。在不同的语境中，“意思”可能有不同的含义。





### 问题-答案汇总表

|  |  |  |  |
| --- | --- | --- | --- |
| 编号 | 问题 | ChatGLM3-6B 原文回答 | Qwen-7B-Chat 原文回答 |
| 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 | 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穿多少衣服，但是它们所针对的季节不同。 | 这两句话的区别在于它们描述的季节不同。 1. "冬天：能穿多少穿多少" 这句话是在强调在寒冷的冬季应该尽可能多地穿着衣物，以保持身体温暖。这是一种适应性行为，因为人们需要保暖以防止体温过低。 2. "夏天：能穿多少穿多少" 这句话是在强调在炎热的夏季应该尽可能少地穿着衣物，以保持身体凉爽。这也是适应性行为，因为人们需要散热以防止体温过高。 因此，这两句话的区别主要在于它们所处的季节以及对温度的调节方式。 |
| 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 | 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，二是谁都看不上”是一个复合句，包含了两个原因，并且使用了逗号进行分隔。第二句“单身狗产生的原因有两个，一是谁都看不上，二是看不上谁”是一个并列句，只包含了一个原因，并且使用了并列连词“一是”进行连接。 | 这两句话的主要区别在于表达的意思和侧重点不同。第一句话 "单身狗产生的原因有两个，一是谁都看不上，二是谁都看不上自己" 更侧重于探讨单身狗产生的两个主要原因，强调了外部因素和自身因素的影响；而第二句话 "单身狗产生的原因有两个，一是谁都看不上我，二是我看不上谁" 更侧重于强调个人主观看法和态度的影响，强调了自我认知和情感需求的重要性。因此，这两句话的侧重点不同，表达的意义也有所差异。 |
| 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 | 这句话是一个经典的谜语,答案是:我不知道你知道他不知道的事情。也就是说,你不知道的事情,他也不知道。 | 这是个谜语，需要我们推理一下。 首先，我们知道“我”知道“你知道”，那么“你知道”这个信息是由“我”传递给“你”的。那么问题来了，“你知道你知道他知道吗?” 这句话的含义是什么呢？ 通过分析，我们可以看出，“你知道你知道他知道吗?”实际上是一个反问句，意思就是“你知道他知道吗？”那么如果“你知道他知道吗”，那么这个句子就没有任何意义了。所以，我们得出结论，“你知道你知道他知道吗?”这句话实际上是在问“你知道他知道吗？”，也就是问“你知道他知道的事情吗？” 因此，这句话的意思是：“你知道他知道的事情吗？”也就是说，这句话是由“你知道”和“你知道他知道吗?”两部分组成的，其中“你知道他知道吗?”是对“你知道”的补充说明。 最后，我们得出结论：在这个谜语中，有四个人物参与：作者（A）、你、他知道的人（B）和读者（C）。在最初的陈述中，作者（A）告诉了你“我知道你知道他知道吗?”这个信息。然后，你通过推理，发现这是一个反问句，并最终解开了这个谜语。而读者（C），则需要通过阅读你的回答来理解这个谜语的答案。 |
| 4 | 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? | 这句话并没有明确说明明明和白白谁喜欢谁,因为“明明明白白白喜欢他”这个表述并没有明确说明是谁在喜欢谁。可能明明知道白白喜欢他,但并没有表达出来,或者白白也喜欢明明,但这句话并没有明确表达出来。 | （ ） A. 明明喜欢白白 B. 白白喜欢明明 C. 喜欢是一样的 答：根据句子大意，可知“明明喜欢白白”，故选A。 |
| 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. “意思”作为名词，表示含义、意义、作用等。在这个语境中，领导问“你这是什么意思”，意思是询问小明刚刚的行为或话语的含义。 2. “意思”作为动词，表示表示、暗示、示意等。在这个语境中，小明用“意思意思”表示自己的行为是在暗示领导什么。 3. “意思”作为动词，表示含义、意味等。在这个语境中，小明用“小意思，小意思”表示自己的行为只是一个小小的意思，没有别的意思。 4. “意思”作为形容词，表示有意义的、有含义的。在这个语境中，领导说“你这人不意思”，表示对小明行为的否定评价。 5. “意思”作为名词，表示意味、暗示等。在这个语境中，小明说“其实也没有别的意思”，表示自己的行为并没有其他的含义或暗示。 | 1. "没什么意思"的意思是不重要或不值得讨论。 2. "意思意思"的意思是一般性的表达，通常用于礼貌地打招呼或表示随便。 3. "不够意思"的意思是没有按照预定的计划 或要求去做事情，或者没有尽到应有的责任和义务。 4. "有意思"的意思是有吸引力、有趣或引人入胜的。 5. "其实也没有别的意思"的意思是事实上没有什么特别的意思。 6. "那你就不客气了"的意思是你不用那么客气，我并不介意。 7. "其实我也不是故意的"的意思是我并没有故意做这件事，只是因为某种原因而发生了这种情况。 以上都是对“意思”的不同理解。在不同的语境中，“意思”可能有不同的含义。 |

## 文大语言模型问答表现的横向对比分析：以 ChatGLM3-6B 与 Qwen-7B-Chat 为例

## 摘要

近年来，中文大语言模型在自然语言理解任务中表现出显著进步。为更细致地评估其在语言幽默、歧义解析及语境推理等复杂语用场景中的能力，本文以 ChatGLM3-6B 与 Qwen-7B-Chat 两个开源模型为研究对象，设计五组富含语义挑战的中文问答题，涵盖文字游戏、代词嵌套、多义词理解等维度。通过对模型回答进行多指标评估与对比分析，结果显示 Qwen-7B-Chat 在准确性、逻辑性、表达清晰度等方面整体优于 ChatGLM3-6B，后者则在基础表达与日常语义任务中具有一定优势。研究为中文语言模型的实用选择与能力定位提供参考。

## 1 引言

近年来，中文类 Chat 模型（如 ChatGLM 系列、Qwen 系列、Baichuan 系列等）被广泛应用于教育、写作、对话系统等领域。然而，人类语言时常出现幽默、风趣的口吻，具有“一词多义”“歧义性质”的遣词造句。对此，我想研究一下部分大模型对语言幽默、语义歧义处理等复杂语境的横向比较这一发方向。为此，本文选取 ChatGLM3-6B 与 Qwen-7B-Chat 两款主流开源模型，从多轮复杂语义问答角度出发，对其语言理解能力展开系统分析。

## 2 评估方法

### 2.1 模型选择

* **ChatGLM3-6B**：清华 KEG 实验室与智谱 AI 联合发布的中文优化语言模型，强调轻量化与多模态适配。
* **Qwen-7B-Chat**：阿里云旗下推出的 Qwen 系列对话模型，具备较强中文表达能力，特别强调多轮对话理解。

### 2.2 测试设计

本实验共设计 5 组中文问答，问题语言具有如下特点：

|  |  |
| --- | --- |
| 题号 | 测试内容简述 |
| Q1 | “冬天/夏天能穿多少穿多少”的语义对比 |
| Q2 | 幽默句式“单身狗”中主谓宾错位分析 |
| Q3 | 多层嵌套代词句“他知道我知道你知道他不知道吗？” |
| Q4 | 拟声重复句“明明明明明白白白喜欢他”语义还原 |
| Q5 | 连续使用多义词“意思”的句群解析 |

### 2.3 评估维度

评估采用人工标注与内容分析结合的方法，从如下个维度对比两个模型的表现：

1. 准确性 / 正确性
2. 语言流畅性 / 表达清晰度
3. 分析深度 / 广度
4. 合理性 / 逻辑性
5. 覆盖度 / 完整性
6. 创造性 / 多样性
7. 回答风格
8. 安全性 / 审慎性
9. 回答长度控制能力

## 3 评估结果与分析

### 3.1 维度对比表

**表 1：ChatGLM3-6B 与 Qwen-7B-Chat 在十个维度下的性能对比**

|  |  |  |  |
| --- | --- | --- | --- |
| 维度 | ChatGLM3-6B 表现 | Qwen-7B-Chat 表现 | 小结 |
| 准确性 | 回答基本合理，个别语义误判 | 回答普遍准确，语义识别优秀 | Qwen 更稳健 |
| 表达流畅性 | 偶有重复，结构一般 | 语言自然、层次清晰 | Qwen 表达更规范 |
| 分析深度 | 偏向字面解读和句式解读，缺乏语用层扩展 | 能从行为/文化层面解构语义 | Qwen 更具深度 |
| 逻辑性 | 个别逻辑链跳跃或断裂 | 多层代词与反语结构解析能力强 | Qwen 更为严密 |
| 完整性 | 回答不稳定，某些问题可以回答全面，某些回答过于简单 | 大多数全面解释所有多义用法，存在个例回答格式令人困惑 | Qwen 覆盖更全 |
| 创造性 | 保守、模板化明显 | 分析中融入推理/心理动因等推测 | Qwen 更富启发性 |
| 回答风格 | 偏工具化、实用导向 | 略具人类思维、结构清晰 | 风格差异明显 |
| 安全性 | 全文无敏感/攻击内容 | 同上 | 表现一致 |
| 长度控制 | 有时冗长、重复句式 | 回答紧凑，语言密度高 | Chat回答比较简短，Qwen回答较长。 |

### 3.2 代表问题详细对比

以下为 Q1 ~ Q5 的代表性对比分析：

#### Q1：“冬天/夏天能穿多少穿多少”语义比较

* **ChatGLM3**：仅指出季节差异，也就是用词方面的差别。对于相同的文字，未解释“能穿多少”语义反差，无法进行深入语义思考。
* **Qwen**：将表达解构为“行为适应”的语言反讽，可以理解语言的深刻含义，同时解释充分。

#### Q2：“单身狗产生的两个原因”

* **ChatGLM3**：误将“看不上”与“被看不上”理解为并列含义，仅仅解释了语句的结构，没有进行语义理解。
* **Qwen**：准确区分自我认知与他人评价间的幽默对仗，正确理解的语言含义，但是解释太过冗长，同时有些生涩。

#### Q3：多层代词“知道”句解析

* **ChatGLM3**：语义链混乱，解析失败，回答让人感觉莫名其妙。
* **Qwen**：逐层拆解角色视角，逻辑清晰，但是回答过于冗长，存在众多重复性语句。

#### Q4：“明明明明明白白白喜欢他”

* **ChatGLM3**：表示无法判断主谓关系，但是存在少部分词汇拆解正确。
* **Qwen**：判断“明明”喜欢“白白”，语义还原成功，但是回答过于简略，缺乏合理解释。

#### Q5：“你什么意思”连续句义分析

* **ChatGLM3**：精准定位问题词汇，但是存在部分解释没有解释正确或者逃避回答的情况。
* **Qwen**：逐句解释“意思”用法，结构清晰，但是定位不够精确，同时解释也和人类逻辑有部分差异。

## 4 综合评分

**表 2：两个模型在评分维度下的平均表现（满分 5 分）**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 模型 | 准确性 | 表达清晰 | 深度 | 逻辑性 | 完整性 | 创造性 | 平均得分 |
| ChatGLM3-6B | 3.0 | 3.5 | 2.5 | 2.5 | 2.5 | 2.0 | **3.0** |
| Qwen-7B-Chat | 4.5 | 4.5 | 4.5 | 4.5 | 4.5 | 4.0 | **4.4** |

## 5 结论与展望

本文围绕 5 个具有语言复杂性的问题，对 ChatGLM3-6B 与 Qwen-7B-Chat 两个中文大语言模型进行了系统横向对比。结果显示，**Qwen-7B-Chat 在处理多义歧义句式、语境推理和语言表达方面更具优势**，体现出对中文语用现象更高的适应能力。相较之下，**ChatGLM3-6B 虽具实用性，但在面对幽默与逻辑嵌套句时易出现理解偏差**。

未来研究可在更广泛的语言现象、多轮对话、长文本生成等方面扩展评估维度，进一步探索中文大模型的语言能力边界与演化路径。