**1.研究假设**

从回答的倾向性上来看，Chatgpt生成的对话会比人类的对话在情感上更偏向于中性。

**2.研究方法**

**2.1数据来源**

HC3(https://github.com/Hello-SimpleAI/chatgpt-comparison-detection)是一个人类和ChatGPT对比语料的开源数据集，共包含12835个问题，22259个人类回答，17522个ChatGPT回答。本研究从中筛选出跟心理、情感有关的条目进行分析，包含1099个问题和回答。

**2.2工具**

为了能快速出结果，再加上电脑的性能一般，所以本研究直接用百度云的aip接口进行情感分析任务。在调用aip接口输入文档后，接口会返回5个参数：log\_id（本次请求的id）、sentiment（表示情感极性分类结果，0:负向，1:中性，2:正向）、confidence（表示分类的置信度，取值范围[0,1]）、positive\_prob（表示属于积极类别的概率 ，取值范围[0,1]）、negative\_prob（表示属于消极类别的概率，取值范围[0,1]）。百度aip平台对免费用户有50w次的接口调用权限，对于本次任务来说绰绰有余。

**2.3 数据处理**

所有的数据分析过程都在SPSS中完成。采用单样本T检验来检验人类、ChatGPT回答的情感是否偏离中性情感；采用相关样本T检验来检验在同个问题下，人类、ChatGPT的回答在情感上是否有显著差异。

**3.数据结果**

**3.1数据筛选**

以置信度为指标来筛选数据。由于置信度低于0.5的情感分析结果有较大可能存在误差，所以本研究分别以人类和ChatGPT对话的置信度平均值（）为界限，将低于平均值的对话剔除出本次分析。筛选前共有1099条回答数据，筛选后还剩下456条回答数据。

**3.2 单样本T检验**

以1（中立情绪）为检验值，人类、ChatGPT对话的单样本T检验结果如表1显示。

表1 不同文本情绪的单样本T检验结果

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 变量 | *M* | *SD* | T | df |
| 1 | 人类文本情绪 | 1.05 | .98 | 1.15 | 455 |
| 2 | ChatGPT文本情绪 | 1.66 | .74 | 18.78\*\*\* | 455 |

备注：文本情绪（0=消极，1=中立，2=积极）

\* *p* < .05, \*\* *p* < .01, \*\*\* *p* < .001.

由上表可见，ChatGPT文本情绪显著高于检验值，而人类文本情绪则没有显著差异。

**3.3 相关样本T检验**

由于人类文本和ChatGPT文本都是针对同一个问题进行回复，因此在回答的情绪上存在较高的相关性（表2），因此采用相关样本T检验来探查二者是否在文本情绪上存在显著差异。相关样本T检验的结果如表2显示。

表2 不同文本情绪的相关样本T检验结果

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 变量 | *M* | *SD* | r |  |  | *T* |
| 1 | 人类文本积极情绪 | .52 | .46 | .25\*\*\* | -.29 | .51 | -12.37\*\*\* |
| 2 | ChatGPT文本积极情绪 | .82 | .35 |

备注：文本情绪（表示属于积极类别的概率，取值范围[0,1]）

\* *p* < .05, \*\* *p* < .01, \*\*\* *p* < .001.

由上表可见，人类文本的积极情绪显著低于ChatGPT文本的积极情绪。并且人类文本积极情绪的标准差也大于ChatGPT，说明人类文本的情绪更不稳定。

综上，数据分析的结果不能支持假设。数据分析的结果说明，ChatGPT的文本情绪总体上更加稳定且积极。而人类文本的情绪则更加极化（更积极或更消极）。