

尽管 ChatGPT 在信息提炼、机器翻译、语句生成等多项任务中都能取得较好的效果, 但其能力仍然具有一定局限性. 本文试从文本理解这一语言模型应当具有的重要能力着手, 通过实例展示 ChatGPT 在这一任务上表现的不足之处.

A 文本理解的准确性不足: 以对多义词句语义的理解为例

一词多义、一句多义都是常见而重要的语言现象, 能够结合上下文情境和语法规则, 正确推断词的意义和词性或理解句子的实际意义, 是语言模型在文本理解方面的目标之一. 但是, 以下的例子说明 ChatGPT 并不能完美理解或区分具有多义性词句的准确含义.

A.1 多义词

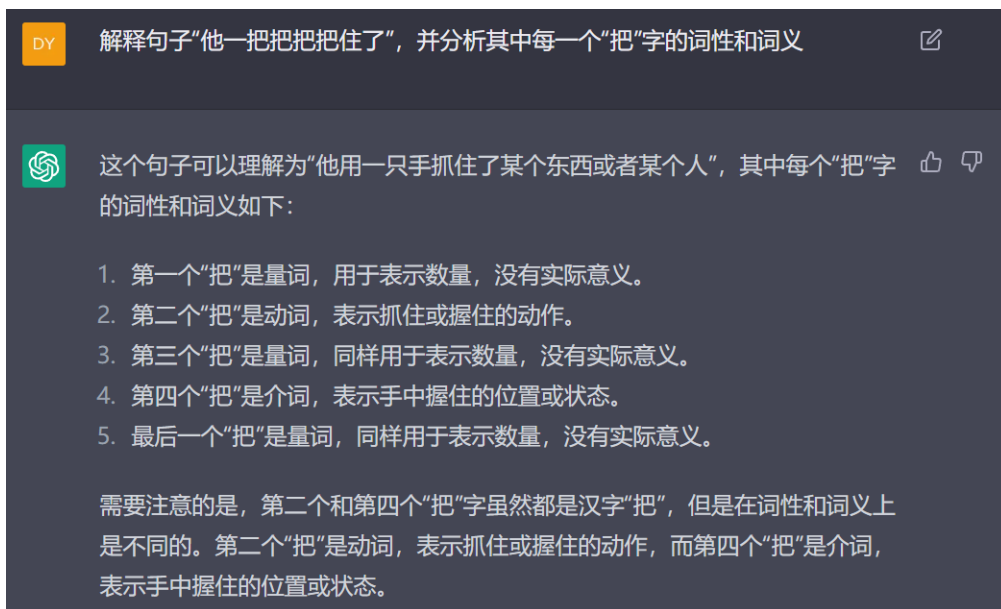


图 1: 不同含义的同一字反复出现对 ChatGPT 的理解带来偏差的例子.

如图1所示, ChatGPT 在理解句子“他一把把把把住了”中的各个“把”字时均错误或不准确, 甚至在确定句子中“把”字的数量时也出现了错误. 该句子的正确分词方式与各“把”字的含义应为

他/一把 (“一把”连用, 表示用手一下子做某动作)/把 (提示宾语前置的助词)/
把 (名词, 车把)/把住了 (动词, 抓握).

Homework 1

自然语言处理基础
2023 年 12 月 13 日

即使对这个句子加入上下文情境, 让”把把把住”的实际意义”握住车把”更为明晰, ChatGPT 仍然无法正确推断出句中“把”字的意义 (图2).

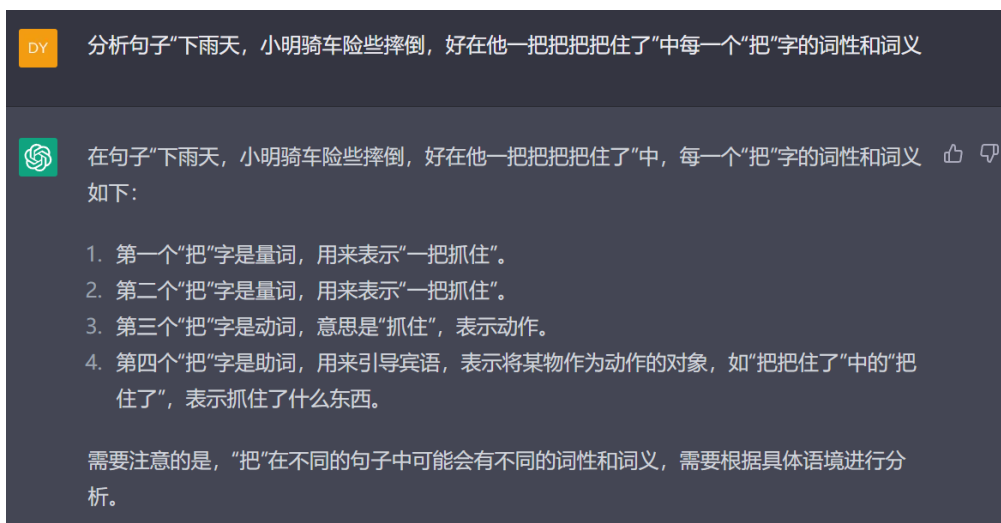


图 2: 加入上下文语境后, ChatGPT 对词义的理解仍然存在问题.

我们将这个句子进一步的简单化, 将助词“把”用同义词“将”代替, 将动词“把住”用同义词“捏住”代替, 从而降低理解句子的困难程度, 但图3所展示的 ChatGPT 的回答内容表明, 它不仅仍然无法理解句中“把”字的含义, 并且生造了“把”字的用法以强行解释词义. 由此可见, ChatGPT 并不具备真正理解或推测词句内容的能力.

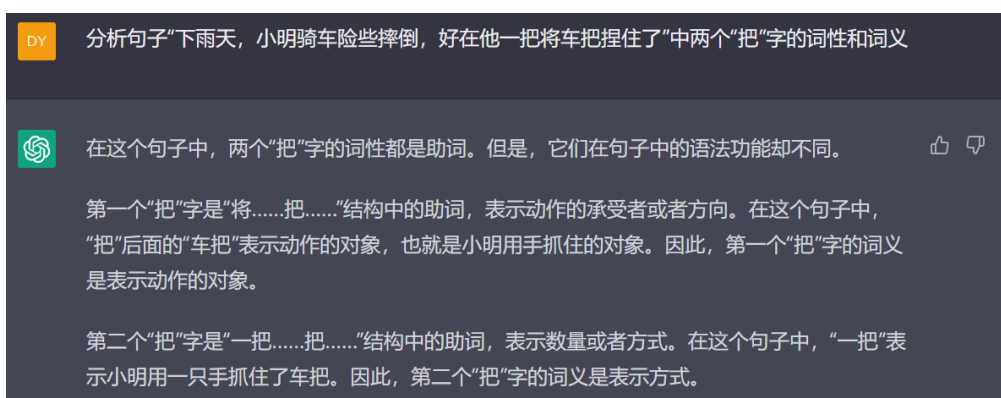


图 3: 减少同义字出现的次数后, ChatGPT 仍无法正确解读词义.

最后, 我们将四个“把”中的三个用同义词替换, 此时, ChatGPT 能够较为准确地理解余下的“把”字的含义, 图4展示了其中一种情况. 这说明图1, 2, 3中所呈现的理解偏差, 并非因

为 ChatGPT 本身不了解“把”字的意义,而是受到了上下文出现的具有不同含义的同一字的影响.

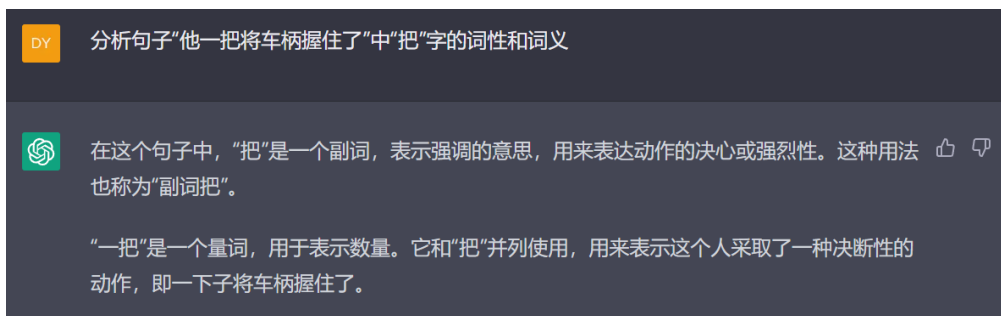


图 4: 当多义词只出现一次时, ChatGPT 对词义的解释相对准确.

A.2 多义句

我们使用句子“这份报告,我写不好”的多义性对 ChatGPT 进行测试. 对这个句子,可以有以下两种分词与解释方法

解释一. 这份报告,我/写不好 (表示认为自己在主客观原因,无法完美完成该动词表示的事项或任务,侧重于表达时间、资源或个人能力因素等的不足引起的“写得不佳”)

解释二. 这份报告,我写/不好 (“不好”用于表达自己不适合承担某项工作,侧重于表达身份、地位、经历等特殊因素带来的“不便行事”)

首先,让 ChatGPT 直接尝试解释这个句子的多义性,得到图5的结果. 从中我们可以看出,尽管 ChatGPT 列出了四条它认为不同的解读方式,但它所给出的各项意义实质上是相同的,也即上面的**解释一**.

但是,无法列举并不代表无法理解相应的意义. 为了进一步检验 ChatGPT 理解该句子多义性能力,我们可以为两种解释增加前文语境,并在此基础上让它进行理解. ChatGPT 生成的结果见图6. 从中我们可以发现,虽然加入了“我也牵扯在这事中”这一具有比较强的对说话者身份、经历等情况的暗示,ChatGPT 的依然倾向采用**解释一**对这句话进行理解.

但是,使用句中不带标点或停顿的“我写不好”表达**解释二**的含义并不合适. 为了探明 ChatGPT 是否有理解**解释二**的能力,将“我写不好”扩充为“让我写,不太好”后,再次进行测试,结果如图7所示. 我们发现,ChatGPT 将这种表达方式理解为说话者缓和的语气,而非

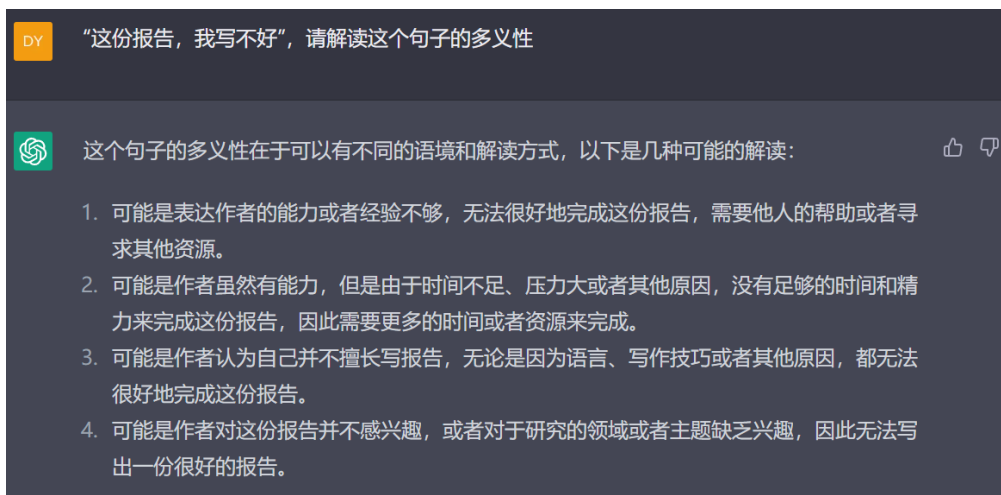


图 5: ChatGPT 直接解释句子的多义性时，其解读意义相互交叉、重复。

对因自身经历而不便写报告的说明。ChatGPT 在这个特殊例子下的表现也说明，它在部分情况下可能会因无法正确理解多义句的含义（即使提供了上下文的信息供推断），而导致生成不正确的回答。

最后，我们尝试向 ChatGPT 描述**解释二**的意义并举例进行阐释，但当我们让 ChatGPT 再次解释“这份报告，让我写，并不好”时，它所给出的回答最终回归到了**解释一**上，见图8。基于此，我们可以猜想，在 ChatGPT 包含“动词 + 不好”形式短语的训练数据中，“不好”的含义为**解释一**的语料数目远远超过**解释二**，导致它只学习到了**解释一**。这一例子表明，以 ChatGPT 为代表的语言大模型可能存在“大数据”掩盖“小数据”的不足之处，这也应当成为未来模型改善的一个需要考虑的方面。

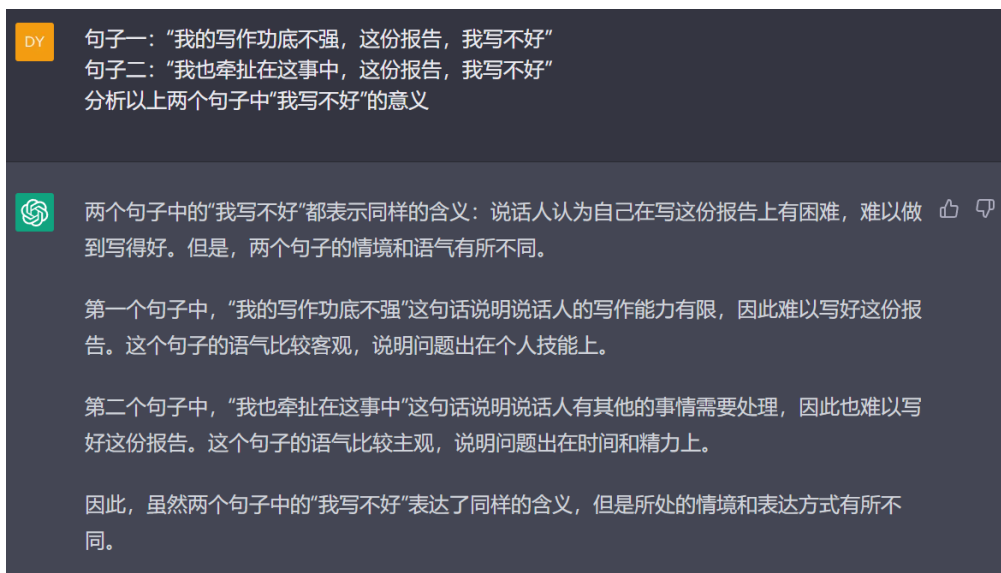


图 6: 在有了上文语境的基础上, ChatGPT 仍然采用**解释一**来理解两句话的句义.

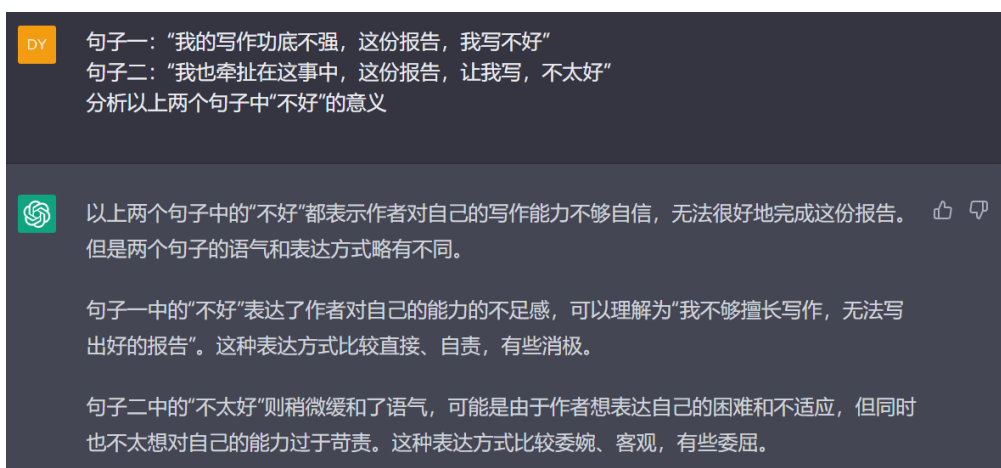


图 7: 将句子二的含义表达得更为明显后, ChatGPT 仍无法解读出说话者因身份、经历等而不适合写报告的内涵.

Homework 1

自然语言处理基础
2023 年 12 月 13 日

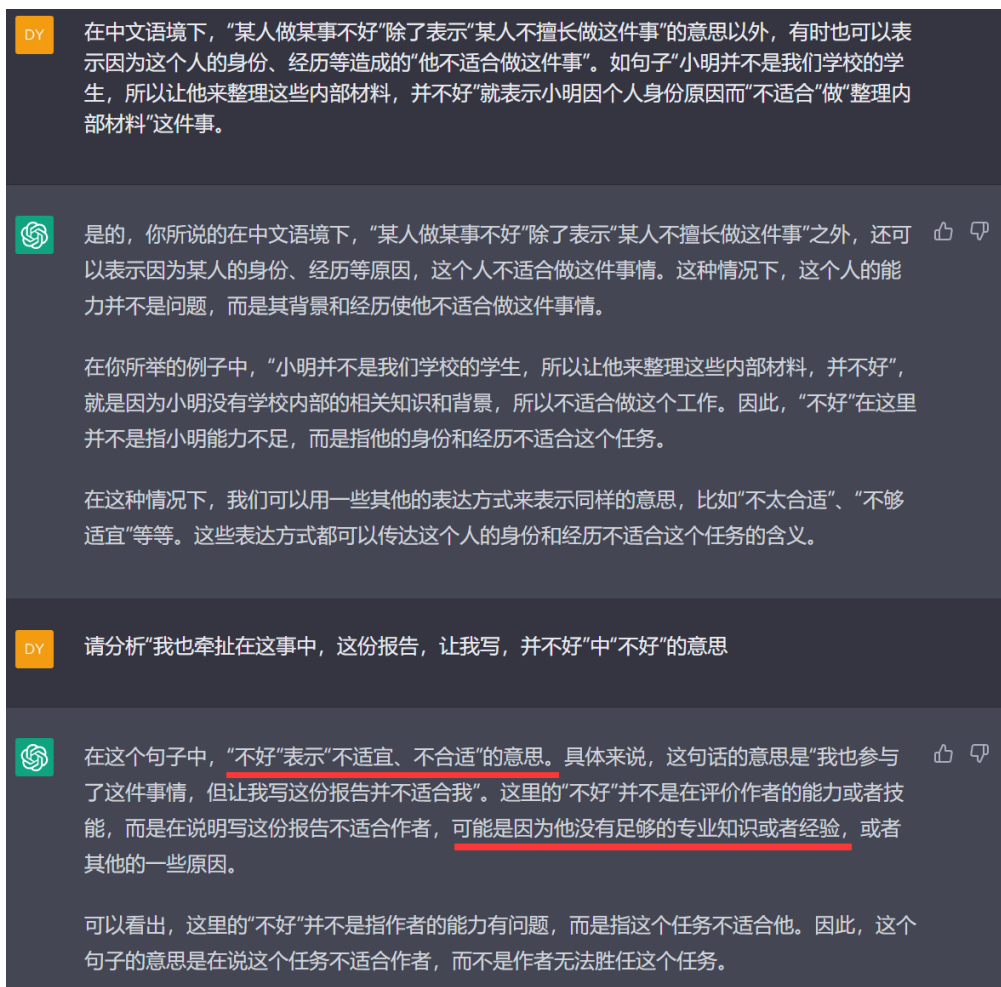


图 8: 即使向 ChatGPT 针对解释二进行了阐述, 它最终的理解仍然落于解释一.