

深度学习与自然语言处理实验报告

王鸣震

wawawa866866@163.com

摘要

本文选取了 ChatGPT3.5、GPT4、文心一言三种目前比较流行的大型语言模型，用材料分析、文本生成、逻辑推理的三种自然语言的下游任务来分析对比这三种自然语言模型的性能。由于 ChatGPT3.5 与 GPT4 的数据库训练都止于 2021 年 9 月，因此将 2021 年 9 月以后的数据投喂于这三种大型语言模型最有说服力并且最具备客观性。

1 介绍

大型语言模型是一种特殊类型的 AI，可以理解和生成人类语言。这类模型接受输入（例如文字）并产生相应的输出，通常以更加复杂的文字形式。这些模型是通过在大量文本数据上进行训练而得到的，可以执行包括但不限于语言翻译，问答系统，摘要生成等多样化的任务。OpenAI 的 GPT 系列模型就是大型语言模型的代表。例如，GPT-3 模型具有 1750 亿个参数，它的训练数据来自互联网上的大量文本，包括书籍，网页，新闻等。GPT-4 模型规模则更大，可生成更准确、自然的文本，更好地理解复杂的输入，解决更复杂的任务。

然而，即使是最先进的大型语言模型，仍然需要一些额外的帮助以更好地解决某些特定任务。这就引出了“提示工程”的概念。提示工程是一个设计输入或“提示”的过程，以引导语言模型产生想要的输出。这是一个相当复杂的过程，需要深入理解语言模型的工作原理，并设计出能够引导模型产生特定结果的提示。

总结来说，大型语言模型是一种强大的人工智能工具，可以理解和生成人类语言。然而，为了实现特定的目标或适应特定的任务，我们还需要进行提示工程，以引导模型产生我们想要的结果。未来的研究将会持续深入探索这两个领域，不断优化并发展大型语言模型的性能，提升提示工程的效率，以期在各种应用中更好地服务人类社会。大型语言模型和提示工程是人工智能领域的重要组成部分，它们的发展将带来深远的影响。

2 实验研究

3.1 材料分析

在这里我选取了 2023 届河北省衡水市部分重点高中二模考试历史试题第 19 题材料分析题的第一小题作为分析材料, 将该材料投喂于 GPT4、New Bing 与 ChatGPT, 结合标准的参考答案试对比出它们的不同结果:

题目:

19. (12 分) 阅读材料, 完成下列要求。

材料

按照英国宗教改革前的传统, 商人死后要将其个人财产的 1/3 留给教会, 或将遗产留给商人行会用于接济贫困成员。在近代早期的伦敦, 商人占全部慈善捐助者的 36%, 比例远高于贵族和乡绅, 商人捐的财物也占总数的 56%。商人慈善捐赠的获益者多是本教区内的老弱病残和一无所有的流浪汉等, 甚至有商人直接为穷人购买衣服、食物等生活用品。中世纪后期的南安普顿商业行会章程有规定: 每个成员在集会时必须携带两个塞斯特尔的啤酒给麻风病患者和其他需要照顾的人。16 世纪前期, 亨利八世宗教改革后, 教会财产被没收, 教会作为慈善捐赠机构的功能被政府所取代。商人在参与慈善活动过程中, 其宗教情感逐渐减弱, 社会责任感在不断地增强。1485 年到 1540 年间, 伦敦有 45% 的私人慈善捐款是给予教会的, 而到 16 世纪末则减少到 7%。受加尔文教、清教道德观的影响, 商人对贫民和流浪汉产生了厌恶感, 认为他们的贫困和流浪是由个人道德败坏和行为懒惰而导致的。

——摘编自陈娟、陈勇《略论近代早期英国商人的慈善活动》

(1) 根据材料, 指出英国宗教改革前后, 商人慈善行为发生的转变, 并结合所学知识解读宗教改革对促成这一转变的具体作用。(8 分)

参考答案:

19. (1) 转变: 从直接向教会捐款捐物到通过世俗机构或直接从事慈善活动; 从同情和帮助流浪汉群体到厌弃流浪汉群体; 由将从事慈善当作宗教义务到将慈善视为社会责任担当。(4 分, 答出两点即可)

作用: 宗教改革剥夺了教会经济特权; 新教财富观鼓励发财致富。(4 分)

提问原文:

按照英国宗教改革前的传统,商人死后要将其个人财产的 1/3 留给教会,或将遗产留给商人行会用于接济贫困成员。在近代早期的伦敦,商人占全部慈善捐助者的 36%,比例远高于贵族和乡绅,商人捐的财物也占总数的 56%。商人慈善捐赠的获益者多是本教区内的老弱病残和一无所有的流浪汉等,甚至有商人直接为穷人购买衣服、食物等生活用品。中世纪后期的南安普顿商业行会章程有规定:每个成员在集会时必须携带两个塞斯特尔的啤酒给麻风病患者和其他需要照顾的人。16 世纪前期,亨利八世宗教改革后,教会财产被没收,教会作为总善捐赠机构的功能被政府所取代。商人在参与慈善活动过程中,其宗教情感逐渐减弱,社会责任感在不断地增强。1485 年到 1540 年间,伦敦有 45% 的私人慈善捐款是给予教会的,而到 16 世纪末则减少到 7%。受加尔文教、清教道德观的影响,商人对贫民和流浪汉产生了厌恶感,认为他们的贫困和流浪是由个人道德败坏和行为懒惰而导致的。

——摘编自陈娟、陈勇《略论近代早期英国商人的慈善活动》

根据材料,指出英国宗教改革前后,商人慈善行为发生的转变,并结合所学知识解读宗教改革对促成这一转变的具体作用。

ChatGPT3.5 回答:



根据所提供的材料和相关知识,可以指出英国宗教改革前后商人慈善行为发生了转变。在宗教改革之前,商人遵循传统宗教观念和行业规定,将一部分财产捐赠给教会或商人行会,用于接济贫困成员和其他需要帮助的人。商人在社会中承担着相当大的慈善责任,对于弱势群体和贫困人士给予帮助和赞助。

然而,宗教改革对商人慈善行为产生了影响。在宗教改革之后,教会的财产被政府没收,教会失去了作为总善捐赠机构的地位,政府取而代之。这导致商人的慈善活动逐渐从教会转向社会其他方面。商人们逐渐减少对教会的捐款,并开始将更多的慈善捐赠直接投入到社会中。商人的宗教情感逐渐减弱,而社会责任感则不断增强。

宗教改革的影响是多方面的。首先,宗教改革废除了宗教的中央集权,使得教会失去了对商人和社会慈善的控制权。这使得商人们有了更多自主决策的权力,可以选择将慈善捐赠用于更广泛的社会目的,而不仅限于教会。

其次,加尔文主义的兴起也对商人慈善行为的转变产生了影响。加尔文主义强调个体的责任和道德行为,商人们受到其影响,开始认为穷人和流浪汉的困境是由于个人道德败坏和行为懒惰所致。这种思想观念使得商人们对于给予贫困人士慈善援助持有较为谨慎的态度,转而强调个人的责任和努力。

总的来说,英国宗教改革前后商人慈善行为发生了转变。在宗教改革之前,商人主要将慈善捐赠投入到教会和商人行会,用于教会和行会成员的救助和援助。而在宗教改革之后,商人的慈善行为逐渐从教会

在宗教改革之后,商人的慈善行为逐渐从教会转向社会其他方面。他们开始将更多的慈善捐赠直接用于帮助社会中的贫困人士和弱势群体,如老弱病残和流浪汉等。

GPT4 回答:



根据这段文字，我们可以看到宗教改革前后英国商人的慈善行为确实发生了明显的转变。在宗教改革前，商人在慈善活动上主要是受教会引导和影响，他们习惯将一部分个人财产留给教会，或者商人行会用来帮助贫困成员，而他们的慈善活动主要是出于宗教的感情和道义责任。然而，随着宗教改革的到来，教会的影响力减弱，商人们的慈善行为逐渐转变为更多地体现出社会责任感，他们对教会的捐赠也大幅减少，而更多地将财产捐赠给社会和个人。

宗教改革对这一转变的推动主要有以下几个方面：

1. 教会财产的没收：在亨利八世的宗教改革中，教会的财产被没收，教会失去了其作为主要慈善组织的能力。这使得商人们必须寻找其他途径来进行他们的慈善活动，从而促使他们更多地将注意力转向社会和个人。
2. 社会观念的转变：宗教改革也引发了社会观念的变革。在新教的影响下，加尔文主义和清教道德观开始流行，这些观念认为个人的贫困和困境是由其个人的道德败坏和懒惰导致的。这改变了商人们对贫困者和流浪汉的看法，也影响了他们的慈善行为。
3. 商人自身的变化：随着商业的发展和资本主义的崛起，商人的社会地位和影响力逐渐提高。他们开始意识到自身对社会的责任，从而在慈善活动上更加积极。与此同时，他们对教会的依赖性减少，使得他们的慈善行为更加自主和多元。

总的来说，英国的宗教改革对商人慈善行为的转变起到了重要的推动作用。

文心一言回答：



对话 发现 NEW



期英国商人的慈善活动》根据材料并用New Bing搜索,指出英国宗教改革前后,商人慈善行为发生的转变,并结合所学知识解读宗教改革对促成这一转变的具体作用。

英国宗教改革前,商人的慈善行为主要是捐赠给教会或商人行会,用于接济贫困成员。而在宗教改革后,教会财产被没收,慈善行为逐渐从教会转向政府,商人的宗教情感也逐渐减弱,社会责任感增强。同时,商人对贫民和流浪汉的态度也发生了转变,认为他们的贫困和流浪是由个人道德败坏和行为懒惰导致的。

宗教改革对促成这一转变的具体作用在于,亨利八世的宗教改革打击了天主教会的权威,使得教会作为慈善捐赠机构的功能被政府取代。此外,加尔文教的清教主义道德观也影响了商人的慈善行为,认为个人道德败坏和行为懒惰导致了贫困和流浪,从而使得商人逐渐将慈善行为从教会转向政府,并更加注重社会责任感的体现。



有问题尽管问我~



在材料分析这一模块，通过比较这三种大型语言模型可以看出，目前这三种流行的大型语言模型都有它们各自的不足，尽管 ChatGPT3.5 可以详细论述英国宗教改革前后商人慈善行为发生的转变，与参考答案已经比较接近了，但对宗教改革促成这一转变的具体作用完全没有提及，因此有一半的分数没有拿到；而 GPT4 对英国宗教改革前后商人慈善行为发生的转变有了更加清晰的论述，但对宗教改革促成这一转变的作用只是说了一句起到了重要的推动作用就匆匆一笔带过，没有具体提及，因此一半的分数也没有拿到；而文心一言虽然对于英国宗教改革前后商人慈善行为发生的转变与宗教改革促成这一转变的作用均有了详细的描述，回答得很全面，但并没有像 GPT4 一样分点分段论述，因此相比之下在清晰性上没有像 GPT4 一样一目了然。

3.2 文本生成

在雅思考试学习中，往往背单词是基本且重要的一贯，但是在背单词时，有很多单词显得晦涩难懂，这时可以发挥运用大型语言模型文本生成编故事的能力来背单词，在这里我选取了 5 个难度相对大一点的英文单词，要求 ChatGPT3.5、GPT4 与文心一言围绕着这五个单词编一个小故事，分析对比它们文本生成的能力。

提问：

请你当我万能的老师。用以下列表中的新单词编一个有趣的英文小故事，50单词以内。我需要在每次新单词出现时都在后面标注中文意思。然后将这个故事翻译成中文，我需要中文翻译的时候遇到对应的新单词也标注一下。最后列出所有新单词，词性和对应的中文意思：
Scourge, intrigue, shipshape, blemish, palpable

ChatGPT3.5 回答：



在一个船舶上，船员们忙着打扫船只，保持整洁有序。船长发现了一个小瑕疵，这引起了他的注意。他感到很兴奋，因为这是他心中一个令人着迷的故事的开端。这个故事是关于一个曾经被折磨着的小岛，它受到了一个可怕的麻烦（Scourge）。船长感到这个故事的真实性非常明显（palpable），他迫不及待地想要了解更多。

故事中的小岛被一个恶名昭著的海盗团伤害得体无完肤。这个团伙是如此邪恶，以至于人们都不敢提起他们的名字。船长被这个故事深深吸引（intrigue），他想要亲自解救这个小岛。他决定与船员们一起，整装待发，确保船只的完美（shipshape）状态，以便开始这个令人激动的冒险。

新单词列表：

1. Scourge [名词] - 灾祸，祸害
2. Intrigue [动词] - 引起兴趣，激起好奇心
3. Shipshape [形容词] - 井然有序的，整洁的
4. Blemish [名词] - 瑕疵，污点
5. Palpable [形容词] - 明显的，可感知的

GPT4 回答：



英文小故事：

The scourge (灾难) of pirates, filled with intrigue (阴谋), kept the shipshape (井然有序的) vessel from reaching its destination. Despite this, the crew's palpable (可触知的) courage left no blemish (污点) on their honor.

中文翻译：

海盗的灾难 (scourge)，充满阴谋 (intrigue)，阻止了井然有序的 (shipshape) 船只到达其目的地。尽管如此，船员们的可触知的 (palpable) 勇气没有给他们的荣誉留下任何污点 (blemish)。

所有新单词，词性和对应的中文意思：

1. Scourge (名词): 灾难
2. Intrigue (名词): 阴谋
3. Shipshape (形容词): 井然有序的
4. Palpable (形容词): 可触知的
5. Blemish (名词): 污点

文心一言回答：



对话

发现 NEW



请你当我万能的老师。用以下列表中的新单词编一个有趣的英文小故事，50单词以内。我需要在每次新单词出现时都在后面标注中文意思。然后将这个故事翻译成中文，我需要中文翻译的时候遇到对应的新单词也标注一下。最后列出所有新单词，词性和对应的中文意思：Scourge, intrigue, shipshape, blemish, palpable

List of New Words:

1. Scourge - n. 灾祸，瘟疫
2. Intrigue - v. 激发兴趣，引起好奇心
3. Shipshape - adj. 整洁有序的
4. Blemish - n. 瑕疵，污点
5. Palpable - adj. 明显的，可感知的

Story:

The scourge of the dirty house had reached a palpable level. The blemishes on the once-shipshape floors were now intriguing, as they

有问题尽管问我~





对话 发现 NEW



2. Intrigue - v. 激发兴趣，引起好奇心

3. Shipshape - adj. 整洁有序的

4. Blemish - n. 瑕疵，污点

5. Palpable - adj. 明显的，可感知的

Story:

The scourge of the dirty house had reached a palpable level. The blemishes on the once-shipshape floors were now intriguing, as they told the story of our chaotic lives. It was up to me to scour the mess and bring back the shipshape order.

中文翻译：

家里脏乱的灾难已经到了明显的程度。曾经整洁有序的地板上，瑕疵现在变得有趣起来，因为它们讲述了我们混乱的生活。我得清理这乱七八糟的东西，带回有条理的秩序。



有问题尽管问我~



从这三种结果可以看出，ChatGPT3.5 的文本生成能力最好，逻辑条理最为清晰；其次文心一言，虽然逻辑与连贯性略显牵强，但是总体也比较满意；相比之下 GPT4 在文本生成上的表现不是那么令人满意，逻辑牵强，语言连贯性也不是特别地强。也许是近期 GPT4 的用户增幅较大，所以算力没有跟上用户的需求，针对这一问题在后续的版本应该会得到一些改善。

3.3 逻辑推理

题目：

S先生、P先生、Q先生他们知道桌子的抽屉里有16张扑克牌：红桃A、Q、4 黑桃J、8、4、2、7、3 草花K、Q、5、4、6 方块A、5。约翰教授从这16张牌中挑出一张牌来，并把这张牌的点数告诉 P先生，把这张牌的花色告诉Q先生。这时，约翰教授问P先生和Q先生：你们能从已知的点数或花色中推知这张牌是什么牌吗？于是，S先生听到如下的对话：

P先生：我不知道这张牌。

Q先生：我知道你不知道这张牌。

P先生：现在我知道这张牌了。

Q先生：我也知道了。

请问：这张牌是什么牌？

参考答案：

应该是方块5

P先生：我不知道这张牌。可知 不是J 2 7 3 8 K 6

Q先生：我知道你不知道这张牌。可知 应该是红桃或方块

P先生：现在我知道这张牌了。

Q先生：我也知道了。可知应该是方块5

ChatGPT3.5 回答：

根据对话内容，我们可以逐步分析并确定这张牌是什么。

P先生在第一句话中说他不知道这张牌，这意味着这张牌的点数在所有点数中不唯一。这意味着这张牌不是A、Q、4这三种点数之一。

Q先生在第二句话中说他知道P先生不知道这张牌。这意味着这张牌的花色在所有花色中不唯一。这意味着这张牌不是红桃、黑桃、草花这三种花色之一。

现在我们来看第三句话。P先生说他现在知道这张牌了。这意味着通过Q先生的回答，P先生能够确定这张牌的点数。这暗示着这张牌的点数在剩下的点数中是唯一的。

由于这张牌的花色不是红桃、黑桃、草花，剩下的唯一可能就是方块。所以这张牌的花色是方块。

最后，根据Q先生的最后一句话，他说他也知道了这张牌，这意味着通过P先生的回答，Q先生能够确定这张牌的花色。由于这张牌的花色是方块，Q先生可以确定这张牌的花色。

综上所述，这张牌的点数和花色分别为方块的5，所以这张牌是方块5。

GPT4 回答：



这个问题是一个经典的推理问题，通过一系列的信息交换和推理，我们可以得出答案。我们可以按照以下步骤进行推理：

1. 首先，P先生说他知道这张牌。这意味着点数不能是唯一的，否则P先生就能确定这张牌了。所以，这张牌不能是红桃A、黑桃J、草花K，因为这些牌的点数在所有牌中是唯一的。
2. 然后，Q先生说他知道P先生不知道这张牌。这意味着这张牌的花色也不能有唯一的点数。所以，这张牌不能是红桃Q、红桃4、黑桃2、黑桃7、黑桃3，因为这些花色的牌有唯一的点数。
3. 此时，我们剩下的牌是：黑桃8、4、草花Q、5、4、6、方块A、5。在这些牌中，P先生现在能确定这张牌，所以这张牌的点数在剩下的牌中必须是唯一的。因此，这张牌不能是黑桃4、草花4、草花6，因为这些牌的点数在剩下的牌中不是唯一的。
4. 最后，我们剩下的牌是：黑桃8、草花Q、5、方块A、5。在这些牌中，Q先生现在也能确定这张牌，所以这张牌的花色在剩下的牌中必须是唯一的。因此，这张牌只能是方块5，因为方块是在剩下的牌中唯一的花色。

所以，这张牌就是方块5。

文心一言回答：



对话 发现 NEW



翰教授从这16张牌中挑出一张牌来，并把这张牌的点数告诉 P 先生，把这张牌的花色告诉 Q 先生。这时，约翰教授问 P 先生和 Q 先生：你们能从已知的点数或花色中推知这张牌是什么牌吗？于是，S 先生听到如下的对话： P 先生：我不知道这张牌。 Q 先生：我知道你不知道这张牌。 P 先生：现在我知道这张牌了。 Q 先生：我也知道了。 请问：这张牌是什么牌？

根据对话，我们可以推断出这张牌的点数是唯一的，否则 P 先生无法确定这张牌是什么牌。同时，这张牌的花色也是唯一的，因为如果这张牌的花色有不止一种的可能性，那么 Q 先生就不能确定 P 先生无法确定这张牌是什么牌。

因此，这张牌的点数是唯一的，花色也是唯一的。在16张牌中，只有一张牌的点数是唯一的，花色也是唯一的，那就是方块5。因此，这张牌是方块5。

有问题尽管问我~



从这三种大型语言模型可以看出，在逻辑推理方面，三种语言模型各有各的长处，文心一言的回答相对简洁，但却给出了正确答案；ChatGPT3.5 与 GPT4 的答案虽然冗长些，但条理思路也很缜密。

结论

ChatGPT3.5、GPT4、文心一言这三种不同大型语言模型处理相同的下游任务各有各的长处，在逻辑推理与材料分析方面 GPT4 相对优势大一些；在文本生成方面 ChatGPT3.5 的优势相对大一些；在这三个方面文心一言给出的答案虽然相对简洁，但综合性、均衡性把握得很好，给出的答案结果也很令人满意。