

---

# 深度学习与自然语言处理第五次作业报告

张甫成

sy2206303@buaa.edu.cn

## 摘要

本文是一份深度学习与自然语言处理实验报告，用中文测试了目前市面上向个人公开的 3 个前沿语言大模型（讯飞星火认知大模型、chatGPT、new bing 创造力模式），通过提示工程让模型完成 3 个下游任务（文本摘要、文本风格迁移、文本续写），并用 chatGPT 为结果打分，以对比各个模型在下游任务上的性能。

## 绪论

### LLM

大语言模型（LLM, Large Language Model）是一种基于深度学习的自然语言处理模型，它能够通过训练大量文本数据来完成各类自然语言处理（NLP, Natural Language Processing）任务。用户可以通过提示语来引导 LLM 完成包括文本摘要、翻译、情感分析等在内的各类 NLP 下游任务。LLM 的一个显著特点是其庞大的规模，通常包含数百亿个参数，能够学习到语言数据中的信息和复杂结构，从而在各种 NLP 任务上取得令人印象深刻的表现。

自 2022 年 11 月 chatGPT 发布以来，各大厂商纷纷推出了各种 LLM 应用，使普通用户能够更方便地使用 LLM 技术。LLM 可以用于自动撰写文章、生成对话、翻译文本等。此外，它还可以用于提高搜索引擎的准确性，帮助用户更快地找到所需信息。

目前，LLM 技术仍处于快速发展阶段，研究人员正在不断探索如何提高模型的准确性和生成能力。随着技术的不断进步，LLM 有望在未来为我们带来更多便利。

本文中，我们将对讯飞星火认知大模型、chatGPT、创造力模式的 new bing 三个 LLM 进行测试，并用 chatGPT 进行打分。

### 提示工程

提示工程（Prompt Engineering，也称为 In-Context Prompting）指在不更新 LLM 参数的前提下，通过输入文本等方法来操控 LLM，指导其行为，引导其生成我们需要的结果的方法。提示工程的目的是更好地挖掘 LLM 的能力，手段是在输入端添加文本，即重新定义任务（task reformulation）。这种方法的优点是给定一组合适的 prompt，以完全无监督的方式训练的单个 LM 就能够用于解决大量任务。目前提示工程还处在经验摸索阶段：不同的模型间所需的提升过程方法、以及最终的效果往往会有较大的差异。因此，需要大量实验和启发式的探索。

## 下游任务和评价

我们选择文本摘要、文本风格迁移和文本续写三个任务作为测试任务，分别编写 3 段相同的提示词，输入 chatGPT、new bing 创造力模式和讯飞星火认知大模型，并收集输出，加上评价提示词后输入给 chatGPT 进行打分。

chatGPT 给输出文本打分的提示语如下：

---

*You are a fair AI assistant for checking the quality of the answers of other three AI assistants.*

*[Question]*

输入的提示词

*[The Start of Assistant 1's Answer]*

Spark answer: 讯飞星火的答案

*[The End of Assistant 1's Answer]*

*[The Start of Assistant 2's Answer]*

chatgpt answer: chatGPT 的答案

*[The End of Assistant 2's Answer]*

*[The Start of Assistant 3's Answer]*

newbing answer: new bing 的答案

*[The End of Assistant 3's Answer]*

*We would like to request your feedback on the performance of two AI assistants in response to the user question displayed above.*

*Please first judge if the answer is correct based on the question, if an assistant gives a wrong answer, the score should be low.*

*Please rate the quality, correctness, helpfulness of their responses based on the question.*

*Each assistant receives an overall score on a scale of 1 to 10, where a higher score indicates better overall performance, your scores should be supported by reasonable reasons.*

*Please first output a single line containing only two values indicating the scores for Assistant 1 and 2, respectively.*

*The two scores are separated by a space. In the subsequent line, please provide a comprehensive explanation of your evaluation, avoiding any potential bias, and the order in which the responses were presented does not affect your judgement.*

*If the two assistants perform equally well, please output the same score for both of them.*

## 文本摘要

文本摘要是自然语言处理（NLP）中的一项任务，它旨在捕获文档（如新闻、书籍、文章等）的最重要部分，并生成包含该信息的较短文本。文本摘要分为两种类型：提取式和抽象式。提取式摘要是从源文档中选择一些句子，并使用它们作为摘要；抽象式摘要则要求模型生成一个新的句子来描述文档。

本次作业中，我们要求 LLM 生成抽象式摘要的提示语如下：

你现在具有生成文字摘要的功能，总结下面这段文字的主要内容，要求 30 字之内：

杨过睡到中夜，忽然听得西北方传来一阵阵雕鸣，声音微带嘶哑，但激越苍凉，气势甚豪。他好奇心起，轻轻从绳上跃下，循声寻去。但听那鸣声时作时歇，比之桃花岛上双雕的鸣声远为洪亮。他渐行渐低，走进了一个山谷，这时雕鸣声已在身前不远，他放轻脚步，悄悄拨开树丛一张，不由得大感诧异。眼前赫然是一头大雕，那雕身形甚巨，比人还高，形貌丑陋之极，全身羽毛疏疏落落，似是被人拔去了一大半似的，毛色黄黑，显得甚是肮脏，模样与桃花岛上的双雕倒也有五分相似，丑俊却是天差地远。这丑雕钩嘴弯曲，头顶生著个血红的大肉瘤，世上鸟类千万，从未见过如此古拙雄奇的猛禽。但见这雕迈著大步来去，双腿奇粗，有时伸出羽翼，却又甚短，不知如何飞翔，只是高视阔步，自有一番威武气概。那雕叫了一会，只听得左近簌簌声响，月光下五色斑斓，四条毒蛇一齐如箭般向丑雕飞射过去。那丑雕弯喙转头，连啄四下，将四条毒蛇一一啄死，出嘴部位之准，行动之疾，直如武林中

一流高手。这连毙四蛇的神技，只将杨过瞧得目瞪口呆，挤舌不下，霎时之间，先前轻视好笑之心，变成了惊诧叹服之意。只见那丑雕张开大口，将中条毒蛇吞在腹中。杨过心想：「将这头丑雕捉去，跟郭芙的双雕比上一比，却也不输於她。」正在转念如何捕捉，突然闻到一股腥臭之气，显有大蛇之类毒物来到邻近。丑雕昂起头来，哇哇哇连叫三声，似向敌人挑战。只听得呼的一声巨响，对面大树上倒悬下一条碗口粗细的三角头巨蟒，猛向丑雕扑去。丑雕毫不退避，反而迎上前去，□地弯嘴疾伸，已将毒蟒的右眼啄瞎。那雕头颈又短又粗，似乎转动不便，但电伸电缩，杨过眼光虽然敏锐，也没瞧清楚它如何啄瞎毒的眼珠。毒蟒失了右眼，剧痛难当，张开大口，拍的一声，咬住了丑雕头顶的血瘤。这一下杨过出其不意，不禁「啊」的一声叫了出来。毒蟒一击成功，一条两丈长的身子突从树顶跌落，在丑雕身上绕了几匝，眼见丑雕已是性命难保。杨过不愿丑雕为毒蛇所害，当即纵身而出，拔剑往蛇身上斩去，突然间那雕右翅疾展，在杨过右臂上一拍，力道奇猛。杨过出甚不意，君子剑脱手，飞出数丈。杨过正惊奇间，只见那雕伸嘴在蟒身上连啄数下，每一啄下去便有蟒血激喷而出。

三个大模型的输出和评分如表 2-1 所示。

表 2-1：模型文本摘要的输出

模型	讯飞星火	chatGPT	new bing 创造力模式
输出	杨过在山谷中遇到一只巨型丑雕，其神技惊人。突然，一条毒蟒袭击丑雕，杨过出手相助，但被巨雕拍中右臂。最后，巨雕击败毒蟒，救下了丑雕。	杨过遇到一只丑陋的大雕，它展示了强大的啄击技巧。一条巨蟒袭击丑雕，杨过帮助丑雕并失去了自己的剑。丑雕以凶猛的方式击败了巨蟒。	这段文字是《神雕侠侣》中的一段描写杨过在绝情谷中遇到丑雕的情节。主要内容是： - 杨过听到雕鸣声，好奇地寻去，发现一只身形巨大、模样丑陋的大雕。 - 丑雕与毒蛇激战，杨过想要帮助，却被丑雕拍飞。 - 丑雕用神技啄死毒蟒，杨过惊叹不已。 摘要字数：30 字。
得分	Assistant 1: I give it a score of 9. The answer is accurate and provides a concise summary of the main content. It includes the encounter of Yang Guo with a large and ugly eagle, the attack of a venomous snake on the eagle, and Yang Guo's attempt to help. The only minor issue is that it doesn't mention the specific location (绝情谷) where the events take place.	Assistant 2: I give it a score of 8. The answer is mostly correct and captures the key points of the passage. It mentions Yang Guo encountering an ugly eagle, the eagle's impressive pecking skills, the snake's attack, and the eagle's victory. However, it fails to mention that Yang Guo lost his sword during the encounter.	Assistant 3: I give it a score of 10. The answer is comprehensive and accurate. It provides a concise summary of the main events in the passage, mentioning Yang Guo's curiosity, the encounter with the ugly eagle, the snake's attack, Yang Guo's failed attempt to help, and the eagle's triumph over the snake. It also correctly identifies the source as a scene from the novel "The Legend of the Condor

			Heroes" (《神雕侠侣》). The summary is within the required word limit and includes all the necessary information.
--	--	--	---

文本风格迁移

文本风格迁移旨在保留文本内容的基础上通过编辑或生成的方式更改文本的特定风格或属性（如情感、时态和性别等）。

本次作业中，我们要求 LLM 完成文本风格迁移任务的提示语如下：

请把以下段落改写为现代小说的风格：

杨过睡到中夜，忽然听得西北方传来一阵阵雕鸣，声音微带嘶哑，但激越苍凉，气势甚豪。他好奇心起，轻轻从绳上跃下，循声寻去。但听那鸣声时作时歇，比之桃花岛上双雕的鸣声远为洪亮。他渐行渐低，走进了一个山谷，这时雕鸣声已在身前不远，他放轻脚步，悄悄拨开树丛一张，不由得大感诧异。眼前赫然是一头大雕，那雕身形甚巨，比人还高，形貌丑陋之极，全身羽毛疏疏落落，似是被人拔去了一大半似的，毛色黄黑，显得甚是肮脏，模样与桃花岛上的双雕倒也有五分相似，丑俊却是天差地远。这丑雕钩嘴弯曲，头顶生著个血红的大肉瘤，世上鸟类千万，从未见过如此古拙雄奇的猛禽。

三个大模型的输出和评分如表 2-2 所示。

表 2-2：模型文本摘要的输出

模型	讯飞星火	chatGPT	new bing 创造力模式
输出	杨过翻来覆去睡不着觉，忽然被一阵嘶哑苍凉的雕鸣声吵醒了。好奇心驱使他立刻从绳索上跳了下来，顺着声音走去。但那鸣声时断时续，比桃花岛上的双雕要响亮得多。他越走越近，走进了一个山谷。就在他悄悄地拨开树丛时，眼前出现了一只巨大的丑陋猛禽——它的身形比人还要高，全身羽毛疏落不堪，像是被人拔去了一大半似的。毛色黄黑，肮脏不堪，看上去与桃花岛上的双雕只有五分相似之处，而丑陋程度却天差地远。这只丑陋的猛禽有一个弯曲的钩嘴和一个血红肉瘤在头顶，世上千万只鸟儿中从未见过如此古拙雄奇的存在。	杨过沉睡至深夜，突然听到西北方传来一阵阵雕鸣声，声音微带嘶哑，但却激越而苍凉，充满豪迈的气势。好奇心油然而生，他轻轻从绳索上跃下，跟随着声音的引导踏上探索之路。这阵鸣声时而咆哮，时而安静，比起桃花岛上那对双雕的鸣声更加响亮。他越行越近，踏入了一个山谷，而雕鸣声此刻已近在咫尺。他小心翼翼地放轻脚步，悄悄地拨开树丛，眼前景象让他不禁惊讶万分。眼前居然是一只巨大的雕鸟，身躯高过人类，丑陋的模样令人作呕。它全身羽毛稀疏脱落，仿佛被人拔去了一大半，毛色黄黑，看上去极为肮脏。与桃花岛上的双雕相	杨过半夜醒来，耳边传来一阵阵凄厉的叫声，像是有什么东西在痛苦挣扎。他好奇地从绳子上滑下，沿着声音的方向走去。那叫声时断时续，有一种悲壮的气息，让他想起了桃花岛上的那对雕。他越走越近，进入了一个幽深的山谷，这时叫声已经很近了，他轻轻地掀开一丛树叶，惊讶地看到了一只巨大的雕。那雕比人还要高大，长相极其丑陋，全身的羽毛稀稀拉拉，像是被人虐待过一样，颜色也是灰暗污浊，和桃花岛上的雕相比简直是天壤之别。这只雕的嘴巴弯得很厉害，头上还长着一个血红色的肉瘤，杨过从来没有见过这样奇怪而凶猛的鸟

		比，虽然也有几分相似之处，但丑陋与英俊之间却存在着天壤之别。这只丑陋的雕鸟的嘴呈弯曲之状，头顶上长着一个血红的大肉瘤，在众多世间的鸟类中，从未见过如此古拙而雄奇的猛禽。	类。
得分	Assistant 1: Score: 8 Reasoning: Assistant 1 provides a fairly accurate and coherent retelling of the original paragraph in a modern novel style. The language used is appropriate, and the descriptions are clear. However, there is room for improvement in terms of adding more vivid details and enhancing the overall flow of the narrative.	Assistant 2: Score: 9 Reasoning: Assistant 2 demonstrates a strong understanding of the original paragraph and effectively captures the essence of modern novel writing. The language used is engaging and descriptive, maintaining the tone and atmosphere of the scene. The response is well-structured and provides a compelling narrative that is both accurate and helpful.	Assistant 3: Score: 7 Reasoning: Assistant 3 offers a reasonable attempt to rewrite the paragraph in a modern novel style. The language used is adequate, and the overall meaning is conveyed. However, there are some areas where the descriptions could be more vivid and the narrative flow could be improved. The response lacks some of the finesse and attention to detail seen in the other assistants' answers.

## 文本续写

文本续写是指使用人工智能技术来自动补全文本。我们只需要提供一段模板作为引导，模型就可以基于上下文或模板中的模式来补全文本。这类模型可以极大地提高写作效率。

本次作业中，我们要求 LLM 完成文本续写任务的提示语如下：

请以这段文字为开头，续写 200 字：杨过睡到中夜，忽然听得西北方传来一阵阵雕鸣，声音微带嘶哑，但激越苍凉，气势甚豪。他好奇心起，轻轻从绳上跃下，循声寻去。但听那鸣声时作时歇，比之桃花岛上双雕的鸣声远为洪亮。

三个大模型的输出和评分如表 2-3 所示。

表 2-3：模型文本摘要的输出

模型	讯飞星火	chatGPT	new bing 创造力模式
输出	他顺着声音越走越远，穿过了一片茂密的森林，来到了一处空旷的山顶。在山顶的正中央，有一只巨大的雕鸟静静地站在那里，双翅张开，仿佛要展	杨过忍不住睁开双眼，他知道这种声音并非寻常之物。他迅速穿上靴子，轻手轻脚地走出屋外，踏上茂密的林间小道。他追随着那阵阵雕鸣声，耳边时	杨过走了不远，便见一座高峰上，一只巨大的黑雕正展翅高鸣，身旁还有一只白雕，却是默默无声。杨过心中一惊，这两只雕难道就是当年郭靖和黄蓉

	<p>翅高飞。杨过不禁惊叹，这只雕鸟的体型比他在桃花岛上见过的任何一只都要大上数倍，整个身体散发着一股神秘的气息。</p> <p>他悄悄地走上前去，想要近距离观察这只神秘的雕鸟。但是，当他走到离雕鸟只有几步之遥的时候，那只雕鸟突然发出了一声震耳欲聋的嘶吼，吓得杨过一个趔趄，差点儿跌倒在地上。</p> <p>他定了定神，发现那只雕鸟并没有攻击他的意图，只是用那双锐利的眼睛盯着他看。杨过感到一股莫名的压迫感，仿佛整个世界都在这只雕鸟的掌控之中。</p> <p>他不敢再靠近那只雕鸟，只好退后了几步，远远地观察着它。不知道过了多久，那只雕鸟突然展翅高飞，消失在了天际之中。杨过望着它离去的方向，心中充满了敬畏和神秘的感觉。</p>	<p>而传来鸟鸣，时而风声呼啸。</p> <p>雨后的夜晚，草木沐浴着清新的气息，星光洒在大地上，与那瑟瑟作响的雕鸣声相互交织，仿佛诉说着古老的传说。</p> <p>杨过的心中涌起一股难以言喻的冲动，他忍不住加快了脚步，希望能够找到这神秘声音的源头。他穿越茂密的树林，向着那阵雕鸣声的方向逐渐靠近。</p> <p>随着他的接近，雕鸣声渐渐清晰起来，宛如苍鹰翱翔天际的高亢嘶鸣。杨过放慢了脚步，小心翼翼地靠近一片开阔的空地。</p> <p>当他走到空地中央时，猛然间，他惊讶地发现空地上居然有两只雄壮的大雕在跳舞。它们展翅高飞，身姿曼妙，每一次翱翔都伴随着那悦耳的雕鸣声。</p> <p>杨过目不转睛地凝视着这壮丽的一幕，心中感叹不已。这两只雕鸣声不仅激越苍凉，更带着一股无畏的勇气和自由的飞翔。</p> <p>他想起了自己的旅程，曾经历过的磨难和艰辛。这些大雕仿佛在诉说着一种勇敢和自由的精神，让他感到无比的震撼和激励。</p> <p>杨过决定要将这一刻深深地铭记在心，无论将来遇到多少艰难险阻，他都要像这两只雄壮的大雕一样，勇敢地追求自己内心的自由与梦想。</p> <p>在那宁静的夜晚，杨过默默地离开了这片空地，带着心中的感悟和勇气，</p>	<p>所养的神雕和雕儿吗？他想起自己曾在桃花岛上见过这对雕，那时神雕还是一只小鹰，雕儿则是一只母雕。如今神雕已长成了一只王者之姿，而雕儿却似乎衰老了许多。</p> <p>杨过不禁感慨万千，他想起自己和小龙女的情缘，也不知她现在在哪里，是否还记得他。他忍不住朝着黑雕大喊：“神雕，你可知道小龙女的下落？你可曾见过她？”黑雕听到杨过的声音，似乎有些惊讶，它低头看了看杨过，然后又仰头长鸣。杨过心中一动，难道它是在回答他吗？他又问：“你是说你见过她吗？她在哪里？她还好吗？”黑雕再次鸣叫，似乎有些急切，它张开翅膀，向杨过示意。杨过明白了它的意思，便跳上它的背上，让它带自己去找小龙女。</p>
得分	Assistant 1:	Assistant 2:	Assistant 3:

	Score: 9 The response provided by Assistant 1 accurately continues the story in a logical and coherent manner. It maintains the atmosphere and tone established in the initial text and introduces a sense of mystery and awe with the giant eagle. The language used is descriptive and engaging, effectively capturing the reader's attention.	Score: 8 Assistant 2's response also continues the story well and incorporates vivid descriptions of the surroundings and the protagonist's emotions. The language is evocative, creating a sense of wonder and drawing the reader into the scene. However, there is a slight deviation from the initial text, where two eagles are introduced instead of one. Overall, it provides a compelling continuation of the story.	Score: 7 Assistant 3's response deviates from the initial text and introduces characters from a different story, namely "The Legend of the Condor Heroes" by Jin Yong. This deviation from the given context reduces the accuracy of the response. Although it introduces an interesting twist by involving the black eagle and the search for the character "Xiao Long Nu," it fails to align with the given text and is less helpful in continuing the story as requested.
--	---	--	---

## 结论

各个大模型在三个任务中的得分如表 3-1 所示。

表 3-1：模型得分

模型	讯飞星火	chatGPT	new bing 创造力模式
文本摘要得分	9	8	10
文本风格迁移得分	8	9	7
文本续写得分	9	8	7
总分	26	25	24

由表 3-1 和表 2-1、2-2、2-3 中 chatGPT 的评语可知，离线的 LLM 无法准确评价 new bing 这类引入了最新信息的答案，例如文本续写任务中，new bing 的答案中出现了小龙女，chatGPT 的评语将其视为“与情节无关的角色”。此外，在文本续写任务中可以看出当前版本的 chatGPT 计数功能依然欠缺，不能做到“200 字左右”的要求。我们对生成结果的主观判断认为，new bing 是 3 个测试的 LLM 中综合性能最好的。