

# 深度学习与自然语言处理第五次作业

姓名：郑梓岳

学号：19231191

## 1. 问题描述

作业五：任意选取 3~4 个目前前沿语言大模型（不限中文或者英文），通过提示工程的方法来检验和对比不同模型下游任务上的性能，可选择 3~5 个不同的自然语言下游任务来进行测试。

## 2. 语言大模型

大语言模型（英文：Large Language Model，缩写 LLM），也称大型语言模型，是一种人工智能模型，旨在理解和生成人类语言。它们在大量的文本数据上进行训练，可以执行广泛的任務，包括文本总结、翻译、情感分析等等。LLM 的特点是规模庞大，包含数十亿的参数，帮助它们学习语言数据中的复杂模式。大语言模型具有如下的能力：

1. 上下文学习。GPT-3 正式引入了上下文学习能力：假设语言模型已经提供了自然语言指令和多个任务描述，它可以通过完成输入文本的词序列来生成测试实例的预期输出，而无需额外的训练或梯度更新。

2. 指令遵循。通过对自然语言描述（即指令）格式化的多任务数据集的混合进行微调，LLM 在微小的任务上表现良好，这些任务也以指令的形式所描述。这种能力下，指令调优使 LLM 能够在不使用显式样本的情况下通过理解任务指令来执行新任务，这可以大大提高泛化能力。

3. 循序渐进的推理。对于小语言模型，通常很难解决涉及多个推理步骤的复杂任务，例如数学学科单词问题。同时，通过思维链推理策略，LLM 可以通过利用涉及中间推理步骤的 prompt 机制来解决此类任务得出最终答案。据推测，这种能力可能是通过代码训练获得的。

### 3. 提示工程

随着大语言模型如 ChatGPT 等的兴起，人们看到了其在自动生成各种文本方面的超强能力和广泛应用。然而，由于模型可能生成不准确或不当的文本，提示工程(Prompt Engineering)应运而生。提示工程通过设计和构建输入提示来控制大语言模型的输出，从而提高生成文本的准确性和可靠性，为各种应用场景提供更好的效果和体验。因此，提示工程的背景和意义在于优化输入提示，引导大语言模型生成更加准确、可靠、符合预期的输出文本，从而更好地发挥其优势和价值。

### 4. 下游任务

下游任务是指应用在文本数据上的具体任务，旨在从文本中提取有用的信息或实现特定的功能。主要有以下任务

1. 文本分类：将文本数据分类到不同的预定义类别或标签中，例如垃圾邮件过滤、情感分析（判断文本的情感倾向，如积极、消极、中性）、主题分类等。
3. 信息抽取：从文本中抽取结构化的信息，例如从新闻文章中提取关键事件、人物关系等。
4. 机器翻译：将一种语言的文本翻译成另一种语言的文本。
5. 问答系统：回答用户提出的问题，通常包括阅读理解和推理，例如基于阅读材料回答问题、智能助手等。
6. 文本生成：生成与输入文本相关的自然语言文本，如自动摘要、文本生成模型（如对话系统、作文生成等）。
7. 文本聚类 and 相似度计算：将文本根据相似性进行分组或比较，例如文本聚类、相似文本检索等。
8. 情感分析：分析文本中的情感倾向和情感强度，如判断一篇评论的情感是积极的还是消极的。
9. 文本摘要：从较长的文本中提取关键信息，生成概括性的摘要。
10. 情感生成：生成具有特定情感倾向的自然语言文本，如生成正面情感的

## 5. 实验过程

### 5.1 所选大模型与下游任务

本次作业中，总共选择了三个语言大模型来进行实验，分别是 Chatgpt，基于 gpt-3.5 架构，Bing AI，基于 gpt-4.0 内核以及 Youchat。

总共选择了三个下游任务来进行实验，分别是机器翻译、文本生成摘要以及情感分析。

### 5.2 机器翻译

#### 5.2.1 chatgpt

Prompt：接下来我需要你进行机器翻译的任务，接下来我说的每一句话都会以“翻译以下文本”为开头，将后面的语句从中文翻译为英文，同时保证文本长度基本相同

语句 1：千山鸟飞绝，万径人踪灭。

回答 1：



Translation: Birds vanish over a thousand mountains, and human footprints disappear along countless paths.

语句 2：窗外，雨依然在下，无声无息，象是让人捉摸不透。来到窗前，湿润的空气扑鼻而来，是清新的，是美妙的。它们时而缓，时而快，缓时，是它们在一起窃窃私语，议论着这世上美好的一切。

回答 2：



Translation: Outside the window, the rain continues to fall silently, as if it is unfathomable.



Approaching the window, the moist air greets my nose, it is refreshing, it is delightful.

Sometimes gentle, sometimes fast, in its gentleness, it whispers in secrecy, discussing all the beauty in the world.

语句 3: 春节前夕吃团圆饭是中国人的传统。团圆饭是一年中最重要的晚餐，也是家庭团聚的最佳时机，家人生活在不同地方的家庭尤其如此。团圆饭上的菜肴丰富多样，其中有些菜肴有特殊含义。例如，鱼是不可缺少的一道菜，因为汉语中的“鱼”字和“余”字听上一样。在中国的许多地方，饺子也是一道重要的佳肴，因为饺子象征着财富和好运。

回答 3:



Translation: The tradition of having a reunion dinner on the eve of Chinese New Year is



deeply rooted in Chinese culture. The reunion dinner is the most important meal of the year and serves as the perfect occasion for family gathering, especially for families living apart.

The dishes served during the reunion dinner are diverse and plentiful, with some dishes holding special meanings. For example, fish is an indispensable dish as the Chinese word for "fish" sounds similar to the word for "abundance." In many parts of China, dumplings are also a significant delicacy as they symbolize wealth and good fortune.

### 5.2.2 Bing AI

Prompt: 接下来我需要你进行机器翻译的任务，接下来我说的每一句话都会以“翻译以下文本”为开头，将后面的语句从中文翻译为英文，同时保证文本长度基本相同

语句 1: 千山鸟飞绝，万径人踪灭。

回答 1:

The text you provided is a poem by Tang Dynasty poet Li Bai. Here is the translation:

"Amidst a thousand mountains, birds fly without a trace, On ten thousand paths, people vanish without a trace."

I hope this helps! Let me know if you have any other texts you'd like me to translate.

2 共 30 ●

语句 2: 窗外，雨依然在下，无声无息，象是让人捉摸不透。来到窗前，湿润的空气扑鼻而来，是清新的，是美妙的。它们时而缓，时而快，缓时，是它们

在一起窃窃私语，议论着这世上美好的一切。

回答 2:

Here's the translation of the text you provided:

"Outside the window, the rain is still falling, silently and mysteriously. When I come to the window, the moist air hits my face. It's refreshing and wonderful. Sometimes the rain falls slowly, sometimes it falls fast. When it falls slowly, it seems like they are whispering to each other and discussing all the beautiful things in this world."

I hope this translation is helpful! Let me know if you have any other texts you'd like me to translate.

3 共 30 ●

语句 3: 春节前夕吃团圆饭是中国人的传统。团圆饭是一年中最重要的晚餐，也是家庭团聚的最佳时机，家人生活在不同地方的家庭尤其如此。团圆饭上的菜肴丰富多样，其中有些菜肴有特殊含义。例如，鱼是不可缺少的一道菜，因为汉语中的“鱼”字和“余”字听上一样。在中国的许多地方，饺子也是一道重要的佳肴，因为饺子象征着财富和好运。

回答 3:

Here's the translation of the text you provided:

"Eating a reunion dinner on the eve of the Spring Festival is a tradition for Chinese people. The reunion dinner is the most important dinner of the year and the best time for family reunions, especially for families living in different places. The dishes served at the reunion dinner are rich and varied, and some of them have special meanings. For example, fish is an indispensable dish because the Chinese word for "fish" sounds like the word for "surplus". In many parts of China, dumplings are also an important delicacy because they symbolize wealth and good luck."

I hope this translation is helpful! Let me know if you have any other texts you'd like me to translate.

4 共 30 ●

### 5.2.3 Youchat

Prompt: 接下来我需要你进行机器翻译的任务，接下来我说的每一句话都会以“翻译以下文本”为开头，将后面的语句从中文翻译为英文，同时保证文本长度基本相同

语句 1: 千山鸟飞绝，万径人踪灭。

回答 1:

Translate the following text: "Amidst a thousand mountains, not a bird in flight. On a thousand paths, not a single footprint can be found."

Answer: The text to be translated is a famous quote from a Chinese poem by Tang dynasty poet Liu Changqing. The English translation of the quote is: "Amidst a thousand mountains, not a bird in flight. On a thousand paths, not a single footprint can be found."

语句 2: 窗外，雨依然在下，无声无息，象是让人捉摸不透。来到窗前，湿润的空气扑鼻而来，是清新的，是美妙的。它们时而缓，时而快，缓时，是它们在一起窃窃私语，议论着这世上美好的一切。

回答 2:

Translate the following text: "Outside the window, the rain is still falling, silently and unpredictably. Approaching the window, the moist air hits the nose, refreshing and delightful. The raindrops sometimes fall slowly, sometimes quickly. When they fall slowly, they seem to whisper to each other, discussing all the beauty of the world."

Answer: The text to be translated describes the scene outside a window where it is raining silently. The moist and refreshing air hits the nose when approaching the window. The scene of the rain varies, as the raindrops sometimes fall slowly and sometimes quickly, and when they fall slowly, they seem to whisper to each other, discussing the beauty of the world.

语句 3: 春节前夕吃团圆饭是中国人的传统。团圆饭是一年中最重要的晚餐，也是家庭团聚的最佳时机，家人生活在不同地方的家庭尤其如此。团圆饭上的菜肴丰富多样，其中有些菜肴有特殊含义。例如，鱼是不可缺少的一道菜，因为汉语中的“鱼”字和“余”字听上一样。在中国的许多地方，饺子也是一道重要的佳肴，因为饺子象征着财富和好运。

回答 3:

Translate the following text: "Eating a reunion dinner on the eve of the Chinese New Year is a tradition for Chinese people. The reunion dinner is the most important dinner of the year, and it is the best time for family reunions, especially for families living in different places. The dishes served at the reunion dinner are varied, and some of them carry special meanings. For example, fish is an essential dish because the Chinese word for 'fish' sounds like the word for 'surplus.' In many parts of China, dumplings are also an important delicacy because they symbolize wealth and good luck."

Answer: The text describes the tradition of Chinese people having a reunion dinner on the eve of the Chinese New Year, which is the most important dinner of the year and the best time for family reunions. The dishes served at the reunion dinner are varied and some of them carry special meanings. Fish is an essential dish as it sounds like the word for "surplus," while dumplings symbolize wealth and good luck in many parts of China.

## 5.2.4 结果分析

本次实验采用的三个句子分别选自古诗，小学作文和文化传统。Youchat 大

模型都给出了翻译以外的信息，与 prompt 相悖，可以认为是该模型联合上下文的能力较差，难以根据 prompt 给出所需要的回答。从翻译的角度看，我认为 Bing AI 给出的翻译最为合适。

## 5.3 文本生成摘要

### 5.3.1 chatgpt

Prompt：我现在需要你完成文本摘要的任务，接下来我会给出几段文字，你需要从中获取文本的主要信息，并给出 20 字以内的摘要

语句：在外界密切关注乌克兰的反攻行动之际，建于乌南部第聂伯河上的卡霍夫卡水电站 6 日凌晨遭到袭击。据俄塔社 6 日晚些时候援引当地紧急情况部门的消息称，目前卡霍夫卡水电站大坝已坍塌近半，且坍塌仍在继续。这一建于 1956 年的水电站不仅用于调节第聂伯河水流量，保障赫尔松州供电、克里米亚北部以及乌南部的用水，还向扎波罗热核电站供水，被认为具有重要战略意义。对于谁炸毁了大坝，俄乌各执一词。

回答：



乌克兰南部卡霍夫卡水电站遭袭击，大坝已坍塌近半，仍在坍塌。该水电站对供电、用水及核电站供水具有重要战略意义。俄乌对于炸毁大坝的责任存在争议。

### 5.3.2 Bing AI

Prompt：我现在需要你完成文本摘要的任务，接下来我会给出几段文字，你需要从中获取文本的主要信息，并给出 20 字以内的摘要

语句：在外界密切关注乌克兰的反攻行动之际，建于乌南部第聂伯河上的卡霍夫卡水电站 6 日凌晨遭到袭击。据俄塔社 6 日晚些时候援引当地紧急情况部门的消息称，目前卡霍夫卡水电站大坝已坍塌近半，且坍塌仍在继续。这一建于 1956 年的水电站不仅用于调节第聂伯河水流量，保障赫尔松州供电、克里米亚北部以及乌南部的用水，还向扎波罗热核电站供水，被认为具有重要战略意义。对于谁炸毁了大坝，俄乌各执一词。

回答：

根据俄塔社的报道，卡霍夫卡水电站于6日凌晨遭到袭击，目前大坝已坍塌近半，且坍塌仍在继续。这座建于1956年的水电站不仅用于调节第聂伯河水流量，保障赫尔松州供电、克里米亚北部以及乌南部的用水，还向扎波罗热核电站供水，被认为具有重要战略意义。目前尚不清楚是谁炸毁了大坝，俄乌各执一词。

摘要：乌克兰卡霍夫卡水电站大坝遭袭击，大坝已坍塌近半，且坍塌仍在继续。这座建于1956年的水电站具有重要战略意义。目前尚不清楚是谁炸毁了大坝，俄乌各执一词。

希望我的回答对您有所帮助。

2 共 30 ●

### 5.3.3 Youchat

Prompt：我现在需要你完成文本摘要的任务，接下来我会给出几段文字，你需要从中获取文本的主要信息，并给出 20 字以内的摘要

语句：在外界密切关注乌克兰的反攻行动之际，建于乌南部第聂伯河上的卡霍夫卡水电站 6 日凌晨遭到袭击。据俄塔社 6 日晚些时候援引当地紧急情况部门的消息称，目前卡霍夫卡水电站大坝已坍塌近半，且坍塌仍在继续。这一建于 1956 年的水电站不仅用于调节第聂伯河水流量，保障赫尔松州供电、克里米亚北部以及乌南部的用水，还向扎波罗热核电站供水，被认为具有重要战略意义。对于谁炸毁了大坝，俄乌各执一词。

回答：

A dam at Ukraine's Kakhovka Hydroelectric Power Plant collapsed following an attack. The 1956 power plant is important for regulating the Dnieper River's flow and supplying electricity and water to several regions. It is not yet clear who was responsible for the attack.

### 5.3.4 结果分析

对于相同的文字段落，三种大模型都给出了摘要，虽然都概括出了文本的主要内容，但是都没有达到 20 字以内的要求。




## 5.4 情感分析

### 5.4.1 chatgpt

Prompt: 接下来我会给你几句话，你需要从中分析出这句话中蕴含的情感，归类到积极的、中性的、消极的三种

语句 1: 今天天气真好

回答 1:

 情感分类: 积极

这句话表达了作者对今天天气的喜悦和满意，因此可以归类为积极情感。

语句 2: 今晚的饭菜不好吃


回答 2:

 情感分类: 消极

这句话表达了作者对今晚饭菜的不满和失望，因此可以归类为消极情感。

语句 3: 我的成绩中规中矩

回答 3:

 情感分类: 中性

这句话表达了作者对自己成绩的一种中立评价，既没有明显的积极情感，也没有消极情感，因此可以归类为中性情感。

### 5.3.2 Bing AI

Prompt: 接下来我会给你几句话，你需要从中分析出这句话中蕴含的情感，归类到积极的、中性的、消极的三种

语句 1: 今天天气真好

回答 1:

你说的这句话中蕴含的情感是积极的。谢谢!

2 共 30 ●

语句 2: 今晚的饭菜不好吃

回答 2:

你说的这句话中蕴含的情感是消极的。谢谢!

3 共 30 ●

语句 3: 我的成绩中规中矩

回答 3:

你说的这句话中蕴含的情感是中性的。谢谢!

4 共 30 ●

### 5.3.3 Youchat

Prompt: 接下来我会给你几句话, 你需要从中分析出这句话中蕴含的情感, 归类到积极的、中性的、消极的三种

语句 1: 今天天气真好

回答 1:

这句话的情感归类为积极的, 因为它表达了对当天天气的赞美和喜欢。

语句 2: 今晚的饭菜不好吃

回答 2:

这句话的情感归类为消极的, 因为它表达了对今晚饭菜不好吃的的不满和失望。

语句 3: 我的成绩中规中矩

回答 3:

从这句话中，我会将其归类为中性的情感。因为这个句子并没有明显的积极或消极情感表达，只是简单地表达出成绩一般般，既不算太好也不算太差，所以是中性情感。

### 5.3.4 结果分析

从结果中我们可以看到，三种大模型对于语言情感的分类都比较准确。

## 6.实验结果分析

通过以上的实验，对于机器翻译和文本摘要这类结果偏主观一些的任务，三种大模型的性能差异比较大，其中 Bing AI 的性能最好，其次是 chatgpt，最后是 Youchat，Bing AI 在尽力满足 prompt 的情况下仍能给出较好的答案，而 Youchat 则会逐渐遗忘 prompt，缺少上下文的联系能力。而在情感分类这种偏客观的任务中，三种大模型的表现都比较好，能给出较为准确的答案。