ASSIGNMENT 1: PROMPT ENGINEERING

Dai Xunlian 225040015
The Chinese University of Hong Kong, Shenzhen
Shenzhen, China
225040015@link.cuhk.edu.cn

1 INTRODUCTION

Why the task is important Prompt engineering is crucial because it enables cost-effective use of language models by optimizing prompts without expensive retraining, and it allows flexible exploration of LLM capabilities, enhancing their performance on tasks ranging from simple queries to complex AI agent systems.

Why LLM is suitable to solve the problem LLMs are well-suited due to their inherent flexibility in adapting to diverse tasks through natural language prompts, coupled with the controllability of their outputs, which allows for precise guidance and iterative refinement via prompt design.

What you did and what you achieved In this assignment, we investigate the effectiveness of prompt engineering techniques on two medical domain tasks: pharmacist licensure exam question answering and preference-based AI feedback generation. We systematically explore multiple prompt optimization strategies, including language alignment, contextual enrichment with domain-specific metadata, few-shot learning demonstrations, and structured evaluation frameworks. Our work examines how these techniques can enhance LLM performance on specialized medical tasks without model retraining. Through comparative experiments on both tasks, we provide empirical evidence on the practical value of strategic prompt design in domain-specific applications and offer insights into the key factors that contribute to prompt effectiveness.

2 PROBLEM DEFINITION: LLMS AS A KNOWLEDGEABLE DOCTOR

2.1 DEFINITION OF THE TASK

Design and optimize few prompts to improve the accuracy of large language models (LLMs) on a pharmacist licensure exam dataset, ensuring reliable outputs for effective exam preparation.

2.2 PROMPT ENGINEER METHODOLOGY

original prompt A prompt with the question Q and corresponding options O.

3 shots prompt : 1) A prompt with the quesition Q, question type T and options. 2) With 3-shot hints. 3) Transform the prompt language to Chinese due to the question source of dataset

Extra information prompt : A prompt with the quesition Q, quesiton source S, question type T, options and prefix answer string from analysis[1]. 2) Transform the prompt language to Chinese due to the question source of dataset

CoT : A prompt with the quesition Q, question type T and options. 2) Using the ”deepseek-reasoner” model for CoT 3) Transform the prompt language to Chinese due to the question source of dataset

[1]The correct answer in analysis is replaced and acts as a hint for the model answering

2.3 Agent Methodology

Knowledge workflow : 1) The former model is asked to explain the concept of options to be the knowledge. 2) The later model is used to answer the question with the knowledge.

Output: The option answer

Criteria: Whether it as same as the ground truth after **progress** of the answer string.

3 Results Exhibition

See Figures for the advanced example prompt and Table 1 for the final result.

3.1 Prompt Exhibition

Original prompt

System Prompt:
Given a question and two answers. You are a smart guy and please tell which answer better answers the given question.

Figure 1: Original prompt

Prompt with 3 shots

System Prompt:
你是一位医学专家。请回答给定的医学问题。仅输出正确的答案选项。

few shot tempalte:
– 示例 –
问题：{question}
问题类型：{question type}
选项：{options}
分析：{analysis}
答案：{answer}

Figure 2: Prompt with 3 shots

Prompt with extra information(analysis hint)

System Prompt:
你是一位医学专家。请根据提供的医学内容、问题来源和题目类型回答以下问题。
按照指定的题目类型回答医学问题并仅输出正确的答案选项。

Extra information template:
– 示例 –
问题：{question}
问题类型：{question type}
选项：{options}
回答：{analy}

Figure 3: prompt with extra information(analysis hint)

Answer with CoT

System Prompt:
你是一位医学专家。请回答给定的医学问题。仅输出正确的答案选项。
{question} {options}

Inference Model: DeepSeek-reasoner

Figure 4: Answer with CoT

Prompt with Agent Workflow

Teacher Model Prompt: 你是一位医学教育专家。请根据学生提供的问题和选项概念，**简明**地给学生一些知识点概念提示，注意不要告诉学生答案

Student Model prompt: 你是一位医学生，按照指定的题目类型回答医学问题，做大钱，请仔细阅读给定的题目提示，仅输出正确的答案选项。

Figure 5: An example prompt for AI feedback.

3.2 Experiment Results

	Description	Accuracy
Baseline	Basic prompt	86.96
3-shot	Randomly select 3 sample from other questions set	92.75
Extra Information	Clean the final result in the analysis and act as a hint for answer	<u>94.20</u>
CoT	Use DeepSeek-reasoner model for CoT answer	**95.65**
Knowledge Workflow	Teacher model explains the question and student model answers the question	88.41

Table 1: Prompt Settings and Result

4 Conclusion

In this assignment, we systematically investigated various prompt engineering techniques for medical domain question answering, demonstrating substantial improvements over the baseline approach. Our experiments reveal that strategic prompt design can significantly enhance LLM performance on specialized medical tasks.

The baseline approach achieved 86.96% accuracy, establishing a foundation for comparison. Introducing 3-shot demonstrations improved accuracy to 92.75%, showing a 5.79 percentage point gain, which highlights the effectiveness of few-shot learning in providing the model with answer patterns. By incorporating extra information including question sources, types, and analysis hints, accuracy further increased to 94.20%, demonstrating that domain-specific contextual metadata helps guide the model's reasoning process. The CoT methodology approach with DeepSeek-reasoner achieved the highest accuracy at 95.65%, indicating that explicit reasoning steps are particularly valuable for complex medical questions requiring multi-step inference. Interestingly, the knowledge workflow agent methodology, which decomposed the task into knowledge retrieval and answer generation stages, achieved 88.41% accuracy, a modest improvement over baseline but underperforming compared to simpler prompt engineering techniques, suggesting that the two-stage architecture may introduce error propagation or coordination overhead.

However, these improvements come with trade-offs. Few-shot and extra information prompts increase token consumption, while CoT reasoning significantly elevates inference latency and computational costs. The agent workflow methodology further introduces additional API calls and architectural complexity, escalating operational expenses.

ACKNOWLEDGMENT

This is the first asignment for CSC 6201/CIE 6021, see details in `https://llm-course.github.io/`.

REFERENCES