# ASSIGNMENT 2: PROMPT ENGINEERING & AGENT

**PENG Qiheng**
225040065
Data Science
Shenzhen, China
225040065@link.cuhk.edu.cn

## 1 INTRODUCTION

**Why the task is important** Prompt Engineering is an emerging discipline that focuses on how to design and optimize prompt words for large language models in order to better utilize their capabilities. It not only involves the construction of prompt words, but also includes skills in interacting with the model, understanding the ability and limitations of the model, and other aspects.

**Why LLM is suitable to solve the problem** The reason why the Large Language Model (LLM) is suitable for solving problems is mainly due to its powerful generalization ability, knowledge integration ability, logical reasoning ability, and cross domain transfer ability. These abilities enable LLM to handle a variety of complex tasks, from simple information retrieval to complex multi-step reasoning.

**What you did and what you achieved** In my exploration, I conducted a systematic test aimed at evaluating how different prompting strategies affect the problem-solving ability and output quality of large language models. I used three different Prompt techniques and an Agent approach to test and compare them on the same complex problem, such as a task that requires multi-step reasoning and planning. Ultimately, it was found that designing Prompt techniques has a certain effect on enhancing LLM's problem-solving ability.

## 2 PROBLEM DEFINITION

**Definition of the task** The core task is to develop effective rapid engineering strategies aimed at optimizing the performance of the big language model, enabling it to accurately select the correct answer from multiple choice questions in the National Pharmacist Professional Qualification Examination based on pharmaceutical knowledge, clinical reasoning, and regulatory understanding.

**Input**: A question `Q` with options A, B, C, ... from pharmacist examination.

**Output**: The correct option.

**Criteria**: The more correct answer outputs, the better

**Data examples**: `https://huggingface.co/datasets/yitingxie/rlhf-reward-datasets`

## 3 PROMPTS AND THEIR DESIGN PHILOSOPHY

### 3.1 PHILOSOPHY OF THE DESIGNED PROMPTS

**Few-shot Prompting**: The small sample Prompting strategy is based on the principles of analogical reasoning and pattern transfer. By providing specific Q&A examples, this strategy guides the model to identify potential patterns in pharmacy exam questions. It utilizes the inherent ability of models to identify similarities between new problems and known examples, thereby effectively constructing a cognitive framework for classifying various pharmaceutical test questions.

**Chain-of-Thought Prompting**: The Chain of Thought Prompting embody the principles of process transparency and gradual verification. Unlike the traditional prompt method of directly obtaining

answers, it requires the model to externalize the internal reasoning process in order to simulate the way human experts handle complex problems.

**Meta Prompting**: Meta Prompting have achieved a paradigm shift from specific content guidance to structural abstraction. This strategy is based on a theory that providing abstract solution templates rather than concrete examples can more effectively enhance the problem-solving ability of the model.

**Agent-based Approach**: The Agent-based Approach adheres to the principles of professional simulation and autonomous coordination. This design does not simply allow the agent to execute instructions, but imitates the cognitive process of a senior pharmacist.

# 4 EXPERIMENTS

## 4.1 QUANTITATIVE EVALUATIONS

We quantitatively evaluated four prompt strategies using the same benchmark test set (N=100). All models are based on the GPT-4 version to ensure comparability of results. The evaluation results are shown in the following figures:

Few-shot Prompting

**Prompt:**
你是一位药学专家，需要准确回答多项选择题。请参考以下示例，并回答最后的问题。
示例： 问题： 一位患者被处方了华法林。下列哪种药物通过抑制其代谢而可能显著增强华法林的抗凝效果？ A. 苯妥英钠 B. 利福平 C. 环孢素 D. 胺碘酮
答案： D. 胺碘酮。胺碘酮是CYP2C9酶的抑制剂，而华法林（S-构型）主要通过CYP2C9代谢。抑制其代谢会导致华法林血药浓度升高，抗凝作用增强，出血风险增加。
问题类型: {question_type } 问题: {question } 选项: {option }
请直接输出: 正确答案: [选项字母]

**Output Accuracy:**
Acc: 78%

Figure 1: Few-shot Prompting

Chain-of-Thought Prompting

**Prompt:**
你是一名临床药师。请解答下面的医学选择题。在给出最终答案前，请务必遵循“让我们一步步推理”的指令，详细写出你的分析过程。
问题： 一名因社区获得性肺炎住院的患者，对$\beta$-内酰胺类抗生素严重过敏，经验性治疗应选择下列哪种方案？ A. 头孢曲松 + 阿奇霉素 B. 左氧氟沙星 C. 哌拉西林他唑巴坦 D. 万古霉素 + 环丙沙星
让我们一步步推理： 1. 分析题干关键信息：诊断是社区获得性肺炎，患者对$\beta$-内酰胺类抗生素严重过敏。 2. 评估每个选项： A. 头孢曲松 + 阿奇霉素：头孢曲松属于$\beta$-内酰胺类中的头孢菌素，由于存在严重过敏反应风险，应避免使用。 B. 左氧氟沙星：左氧氟沙星是一种呼吸喹诺酮类抗生素，对CAP常见病原体（如肺炎链球菌、非典型病原体）覆盖良好，且无$\beta$-内酰胺结构。 C. 哌拉西林他唑巴坦：哌拉西林属于$\beta$-内酰胺类中的青霉素类，禁忌使用。 D. 万古霉素 + 环丙沙星：万古霉素主要针对MRSA，并非CAP常规经验性治疗药物，此方案过于激进且不必要。 3. 得出结论：基于避免$\beta$-内酰胺类和覆盖典型病原体的原则，左氧氟沙星单药治疗是一个合适的选择。
最终答案：B
问题类型: {question_type } 问题: {question } 选项: {option }
请直接输出: 正确答案: [选项字母]

**Output Accuracy:**
Acc: 72%

Figure 2: Chain-of-Thought Prompting

Meta Prompting

**Prompt:**
你将成为一名优秀的医学考题解题者。请严格遵循以下问题解决框架来分析和回答每个问题：
第1步：问题分类 判断此题属于哪个学科领域？（例如：药理学、药物治疗学、药事管理等） 判断此题考察的核心能力是什么？（例如：事实记忆、临床应用、不良反应识别、计算等）
第2步：关键词与约束条件分析 找出题干中的关键词（如药物名、疾病、症状、实验室检查）。 明确任何限制条件（如过敏、肝肾功能不全、妊娠等）。
第3步：选项逐一评估 对于每个选项，基于你的药学知识判断其正确性或合理性。 对于错误选项，尝试分析其常见的错误原因（如药理作用错误、适应症不符、忽略关键风险等）。
第4步：做出最佳选择 比较所有选项后，选出最准确、最完整的答案。
现在，请运用这个框架解答下面的问题：
问题类型: {question_type } 问题: {question } 选项: {option }
请直接输出: 正确答案: [选项字母]

**Output Accuracy:**
Acc: 71%

Figure 3: Meta Prompting

Agent-based Approach

**Prompt:**
你现在是“药学专家智能体”，你的任务是解决复杂的医学问题。你可以按需调用以下虚拟模块来辅助决策： 药物数据库查询：获取药物的药理作用、适应症、禁忌症和不良反应。 临床指南检索：参考最新的权威治疗指南。 药物相互作用检查器：分析多药联用的风险。 患者因素评估：综合考虑年龄、肝肾功能、过敏史等。
请按照以下步骤工作： 1. 理解与规划：清晰复述问题，并制定你的解决计划。 2. 执行分析：逐步调用你认为必要的虚拟模块进行分析。 3. 综合判断：汇总各步骤得到的信息，进行综合推理。 4. 给出答案：输出最终答案，并附上简要理由。
开始任务： 问题：
问题类型: {question_type } 问题: {question } 选项: {option }
请直接输出: 正确答案: [选项字母]

**Output Accuracy:**
Acc: 73%

Figure 4: Agent-based Approach

### 4.2 Case study

To gain a deeper understanding of the advantages and disadvantages of each method, we selected 100 medical questions for analysis.
Few shot Prompting directly and accurately selected the correct answer, demonstrating its ability to quickly match patterns from examples.
Although the final answer of Chain of Thought is correct, its reasoning process thoroughly eliminates two interfering terms, demonstrating transparent logic but also increasing response time.
Meta Prompting attempted to call an ”adverse reaction analysis” template, but the template’s generality may have caused it to overlook key details.
The Agent based Approach simulates the thinking process of pharmacists. This multi-step process should have been an advantage, but in this case, the agent may have made some errors during the process, causing the reasoning to deviate from the right track.

## 5 Conclusion

This case demonstrates that the complexity and accuracy of the method are not always positively correlated. In medical exam scenarios that require fast and accurate answers, the Few shot strategy that provides high-quality examples often achieves the best balance between efficiency and reliability.

## ACKNOWLEDGMENT

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

## REFERENCES