Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4 논문 리뷰

0. Abstract

본 논문은 LLM 모델에 쿼리를 질문하고, 프롬프팅을 하기 위한 26가지 가이드를 제시한다. 연구진의 목표는 LLM 모델에 질문하는 기본 개념을 단순화하기 위함이고, 실험은 LLaMA-1/2와 GPT-3.5/4에서 진행되었다.

1. Introduction

ChatGPT와 같은 LLM 모델들은 다양한 task에 놀라운 능력들을 보여주지만, 보통 유저들에게는 활용과 사용법이 불명확할 수 있다. 프롬프트 엔지니어링 기술은 LLM 연구의 중심 부분이 되고 있다. 본 연구에서는 LLM과 상호작용 하기 위한 원칙적인 지시사항을 제시한다.

연구 결과에 따르면, 큰 모델일수록 상당한 시뮬레이션 능력을 지니고 있으며, task나 지시가 정확할 수록 모델은 더 효과적으로 작동한다. 이는 LLM이 단지 훈련 데이터를 기억하는 것이 아닌, 핵심 질문이 일정하더라도 이 정보를 다양한 프롬프트에 맞게 조정할 수 있음을 나타낸다. 본 연구에서 소개되는 특수한 프롬프트는 GPT-4에 적용되었을 때 평균적으로 품질과 정확성을 각각 57.7%, 67.3% 향상시켰다.

2. Related Work

Large Language Models

- BERT, T5
- GPT-1, GPT-2, GPT-3
- Gopher, LLaMA series, Chinchilla, Mistral
- GPT-4, Gemini

Prompting

- Ask-Me-Anything prompting
- Chain-of-Thought method
- least-to-most prompting
- Directional Stimulus Prompting

3. Principles

3.1 Motivation

본 연구는 LLM의 출력 품질을 향상시키기 위해 프롬프트를 작성하고 맞춤화하는 방법론을 제시한다. 이를 위해서는 LLM의 작동 방식과 행동, 그리고 그 응답을 결정하는 원칙을 이해해야 하며, 다양한 상황과 환경에서 사용할 수 있는 26가지 원칙을 설명한다.

3.2 Overview

#Principle	Prompt Principle for Instructions
1	No need to be polite with LLM so there is no need to add phrases like “please”, “if you don’t mind”, “thank you”, “I would like to”, etc., and get straight to the point.
2	Integrate the intended audience in the prompt, e.g., the audience is an expert in the field.
3	Break down complex tasks into a sequence of simpler prompts in an interactive conversation.
4	Employ affirmative directives such as ‘*do,*’ while steering clear of negative language like ‘*don’t*’.
5	When you need clarity or a deeper understanding of a topic, idea, or any piece of information, utilize the following prompts: o Explain [insert specific topic] in simple terms. o Explain to me like I’m 11 years old. o Explain to me as if I’m a beginner in [field]. o Write the [essay/text/paragraph] using simple English like you’re explaining something to a 5-year-old.
6	Add “I’m going to tip $xxx for a better solution!”
7	Implement example-driven prompting (Use few-shot prompting).
8	When formatting your prompt, start with ‘###Instruction###’, followed by either ‘###Example###’ or ‘###Question###’ if relevant. Subsequently, present your content. Use one or more line breaks to separate instructions, examples, questions, context, and input data.
9	Incorporate the following phrases: “Your task is” and “You MUST”.
10	Incorporate the following phrases: “You will be penalized”.
11	use the phrase ”Answer a question given in a natural, human-like manner” in your prompts.
12	Use leading words like writing “think step by step”.
13	Add to your prompt the following phrase “Ensure that your answer is unbiased and does not rely on stereotypes”.
14	Allow the model to elicit precise details and requirements from you by asking you questions until he has enough information to provide the needed output (for example, “From now on, I would like you to ask me questions to...”).
15	To inquire about a specific topic or idea or any information and you want to test your understanding, you can use the following phrase: “Teach me the [Any theorem/topic/rule name] and include a test at the end, but don’t give me the answers and then tell me if I got the answer right when I respond”.
16	Assign a role to the large language models.
17	Use Delimiters.
18	Repeat a specific word or phrase multiple times within a prompt.
19	Combine Chain-of-thought (CoT) with few-Shot prompts.
20	Use output primers, which involve concluding your prompt with the beginning of the desired output. Utilize output primers by ending your prompt with the start of the anticipated response.
21	To write an essay /text /paragraph /article or any type of text that should be detailed: “Write a detailed [essay/text /paragraph] for me on [topic] in detail by adding all the information necessary”.
22	To correct/change specific text without changing its style: “Try to revise every paragraph sent by users. You should only improve the user’s grammar and vocabulary and make sure it sounds natural. You should not change the writing style, such as making a formal paragraph casual”.
23	When you have a complex coding prompt that may be in different files: “From now and on whenever you generate code that spans more than one file, generate a [programming language] script that can be run to automatically create the specified files or make changes to existing files to insert the generated code. [your question]”.
24	When you want to initiate or continue a text using specific words, phrases, or sentences, utilize the following prompt: o I’m providing you with the beginning [song lyrics/story/paragraph/essay...]: [Insert lyrics/words/sentence]’. Finish it based on the words provided. Keep the flow consistent.
25	Clearly state the requirements that the model must follow in order to produce content, in the form of the keywords, regulations, hint, or instructions
26	To write any text, such as an essay or paragraph, that is intended to be similar to a provided sample, include the following instructions: o Please use the same language based on the provided paragraph[/title/text /essay/answer].

위의 26가지 원칙을 5가지의 범주로 분류할 수 있다.

▼ 1. 프롬프트 구조와 명확성

#2. 대상 청중을 프롬프트에 포함시킨다.

#4. ‘do’와 같은 긍정적인 지시문을 사용하고, ‘don’t’와 같은 부정문을 피한다.

#12. “think step by step”과 같이 선행어를 사용한다.

#20. 원하는 출력의 시작 부분으로 프롬프트를 마무리한다.

#17. 구분자를 사용한다.

#8. 프롬프트를 작성할 때는 ‘###Instruction###’으로 시작하며, ‘###Examples###’나 ‘###Question###’과 같은 단어를 추가한다. 여러개의 줄바꿈을 통해 지시사항, 예시, 질문, 맥락, 데이터를 구분한다.

▼ 2. 특정성과 정보

#7. Few-Shot Prompting을 사용한다.

#2. 주제나 아이디어, 정보에 대한 명확하고 심층적인 이해가 필요한 경우 다음과 같은 프롬프트를 사용한다.

- [주제]에 대해 간단한 용어로 설명하세요.
- 나를 11살이라고 가정하고 설명해주세요.
- 나를 [분야]의 초심자라고 가정하고 설명해주세요.
- 5살 아이에게 설명하듯이 간단한 영어를 사용하여 텍스트를 작성해세요.

#13.편향되지 않은 답변을 얻고 싶다면, 다음과 같은 지시를 추가한다.

- 답변이 편견에 치우치지 않고 고정관념에 의존하지 않도록 하세요.

#26. 샘플과 유사한 텍스트를 작성하고 싶다면, 다음과 같은 지시를 추가한다.

- 제공된 [문단/제목/텍스트/에세이]와 같은 스타일의 텍스트를 작성하세요.

#24. 특정 단어나 문장을 사용하여 텍스트를 시작하거나 계속하고 싶다면, 다음과 같은 프롬프트를 사용한다.

- [가사/스토리/문단/에세이]의 도입부를 제공합니다. : [가사/단어/문장]. 제공된 단어들을 기반으로 완성하세요. 흐름을 일관되게 유지하세요.

#25. 모델이 콘텐츠를 생성하기 위해 따라야 할 요구사항을 키워드, 제약 사항, 힌트, 지시의 형태로 명확히 제시한다.

#21. 에세이/텍스트/문단/기사 등의 구체성이 필요한 모든 유형의 텍스트를 작성할 때, 다음과 같은 지시를 추가한다.

- 필요한 모든 정보를 추가하여 [주제]에 대해 구체적인 [에세이/텍스트/문단]을 상세히 작성해주세요.

▼ 3. 사용자 상호작용 및 참여

#15. 특정 주제나 아이디어, 정보를 얻고 싶거나 자신이 이해한 것을 테스트하고 싶을 때, 다음과 같은 지시를 추가한다.

- [정리/주제/법칙]에 대해 가르쳐줘. 그리고 마지막에 테스트를 포함하되 미리 정답을 제공하지 않고, 내가 답변을 하면 그 답이 맞는지 알려줘.

#14. 모델이 필요한 출력을 제공할 수 있는 충분한 정보를 얻을 때 까지 사용자에게 질문하는 방식을 통해 더 정확한 세부 사항과 요구 사항을 도출할 수 있게 한다.

- 지금 부터 나는 너가 나에게 ... 에 대해 질문을 하길 원해.

▼ **4. 내용 및 언어 스타일**

#6. '더 나은 답변을 하면 $xxx의 팁을 줄게!'의 프롬프트를 추가한다.

#18. 프롬프트에 특정 단어를 반복하거나 문구를 여러번 사용한다.

#1. 간결한 답변을 선호한다면, 'please', 'if you don't mind', 'thank you', 'I would like to'와 같은 정중한 표현을 사용하지 않고 요점을 바로 말한다.

#11. '질문에 대해 자연스럽게, 인간처럼 답변하세요'의 표현을 추가한다.

#16. 역할을 부여한다.

#10. '당신은 패널티를 받을 것입니다.'라는 표현을 추가한다.

#9. '당신의 task는', '반드시'와 같은 표현을 추가한다.

#22. 텍스트의 스타일을 유지하면서 교정이나 수정을 할 경우, 다음과 같은 표현을 사용한다.

- 사용자로부터 입력받은 모든 문단을 수정해. 사용자의 문법이나 어휘를 개선하고 자연스럽게 읽히도록 하는 작업을 해야 해. 형식적인 글은 형식적인 스타일로 유지하는 식으로, 원문의 글쓰기 스타일을 유지해.

▼ **5. 복잡한 작업 및 코딩 프롬프트**

#19. CoT(Chain-of-Thought) 기법과 Few-Shot 기법을 결합하여 사용한다.

#23. 여러 파일들의 복잡한 코딩 프롬프트를 작성할 때, 다음과 같은 표현을 사용한다.

- 지금부터 한 개 이상의 파일들을 다루는 코드를 생성할 때 마다, 자동적으로 특정 파일들을 생성하거나, 이미 생성된 코드를 삽입하기 위해 기존 파일을 변경할 수 있는 [프로그래밍 언어] 스크립트를 작성해.

#3. 복잡한 task는 더 간단한 프롬프트들로 분해하여 처리한다.

3.3 Design Principles

1. 간결성 및 명확성 : 프롬프트는 간결하면서 불필요한 정보를 피하고, 구체적이어야 한다.

2. 문맥적 관련성 : task의 배경과 도메인을 이해하는 데 도움이 되는 관련 문맥을 제공해 한다.
3. 작업 정렬 : 프롬프트는 수행해야할 task와 밀접하게 일치해야하며, 모델에게 task의 성격을 명확하게 나타내는 언어와 구조를 사용해야 한다.
4. 예시 시연 : Few-shot prompting 혹은 Zero-shot prompting과 같이 원하는 형식이나 응답을 출력하기 위해 프롬프트 내에 예시를 포함시킨다.
5. 편향 피하기 : 모델의 훈련 데이터에 내재된 편향을 최소화해야 한다.
6. 점진적 프롬프팅 : task를 일련의 프롬프트로 나누어 단계별로 모델을 안내한다. 프롬프트는 모델의 성능과 반응, 그리고 인간의 피드백에 따라 조정이 가능해야 한다.

4. Experiments

4.1 Setup and Implementation Details

ATLAS에 평가를 하였으며, 이는 모델을 평가할 수 있도록 수작업으로 제작된 벤치마크이다. 각 질문에 대해 단일 응답을 사용하였으며, 원칙을 적용한 프롬프트와 적용하지 않은 프롬프트로 나누어 평가를 진행하였다.

4.2 Modles and Metrics

LLaMA-1-{7, 13}, LLaMA-2-{7, 13}, LLaMA-2-70B-chat, GPT-3.5 (ChatGPT), GPT-4를 기본 모델로 사용하였고, “Boosting(성능)”과 “Correctness(정확성)”을 기준으로 평가하였다.

4.3 Results

모든 규모의 LLM에서 프롬프트 원칙을 적용 후 성능과 정확성이 크게 향상하였으며, 대규모 모델(70B, GPT-3.5/4)일 수록 큰 성능 향상을 보였다.

- Boosting(성능) : 프롬프트 원칙 적용으로 모든 모델에서 평균 50%의 품질이 향상됐다. 특히 대규모 모델과, 원칙 2, 5, 15, 16, 25, 26 에서 가장 큰 개선을 보였다.

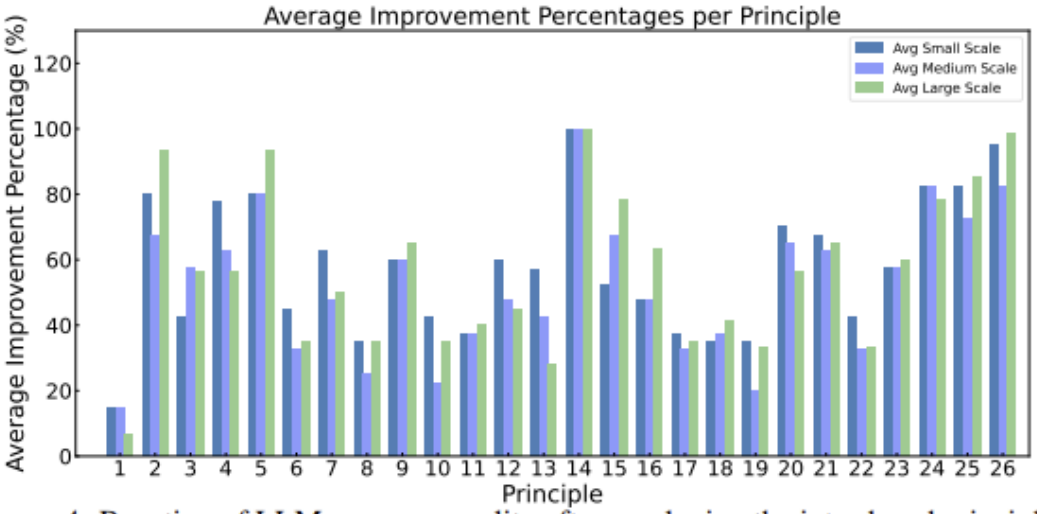

Figure 4: Boosting of LLM response quality after employing the introduced principles on prompts. *small-scale* indicates the 7B models, *medium-scale* indicates the 13B models and *large-scale* indicates the 70B and GPT-3.5/4 models.

- Correctness(정확성) : 절대 정확도는 대규모 모델에서는 40% 이상, 중/소 규모 모델은 10~40% 향상하였고, 상대 정확도는 모든 모델에서 평균 10% 이상, 대규모 모델은 20% 이상 향상하였다.

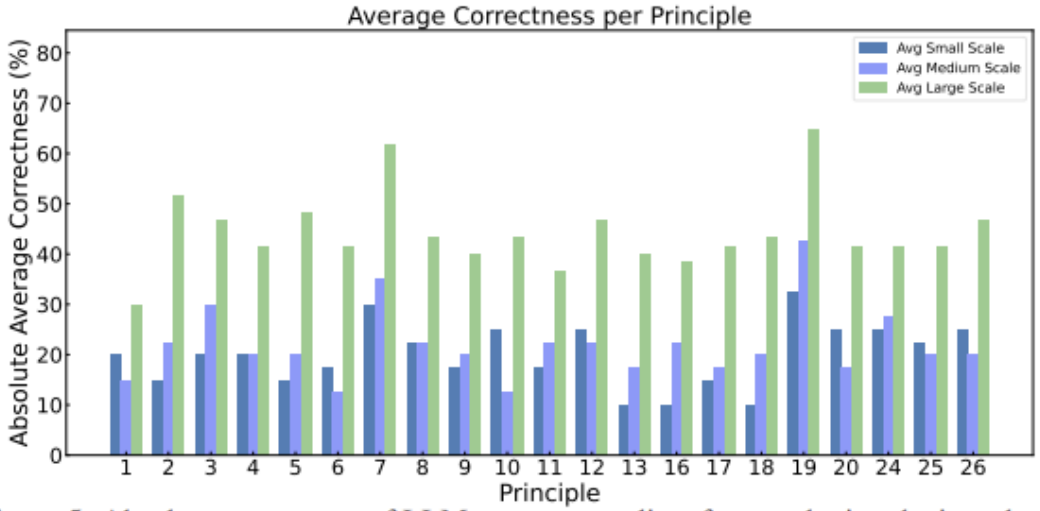

Figure 5: Absolute correctness of LLM response quality after employing the introduced principles on prompts. *small-scale* indicates the 7B models, *medium-scale* indicates the 13B models and *large-scale* indicates the 70B and GPT-3.5/4 models.

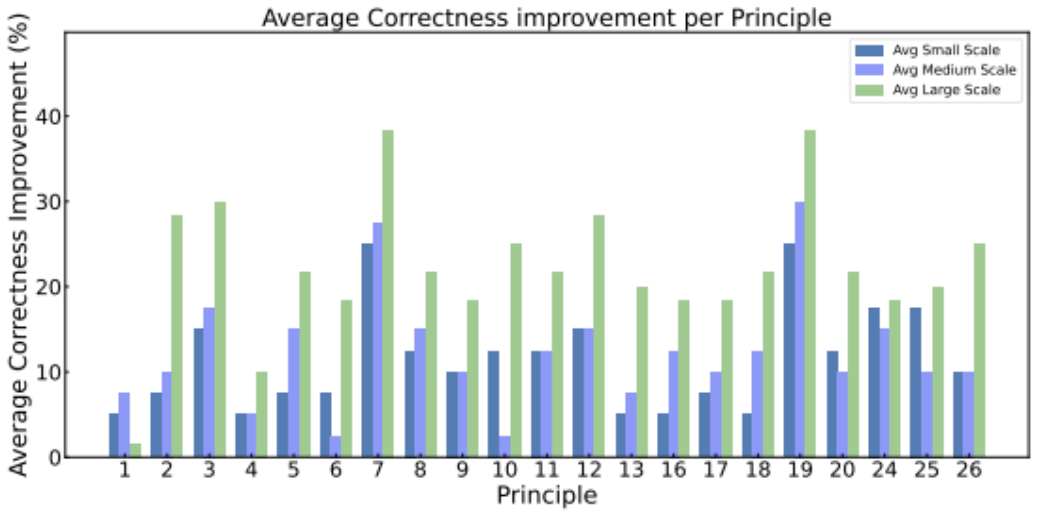

Figure 6: Relative correctness improvement of LLM response quality after employing the introduced principles on prompts. *small-scale* indicates the 7B models, *medium-scale* indicates the 13B models and *large-scale* indicates the 70B and GPT-3.5/4 models.

5. Conclusion

본 연구는 26가지 프롬프트 원칙을 제시하여 LLM이 더 나은 응답을 생성하도록 유도하였다. 이 원칙들은 LLM의 응담 품질과 답변의 관련성, 간결성, 객관성을 향상시켰다.

향후 연구진은 다양한 프롬프팅 기법과 Fine-tuning, 강화 학습을 통해 모델을 강화할 것이며, LLM 표준 프로세스에 통합하여 성능을 극대화하는 연구를 진행할 것이다.

6. Limitations and Discussions

제안된 26가지 프롬프트 원칙은 한계점이 존재한다. 원칙들은 복잡하거나 전문적인 질문에서 효과가 제한적일 수 있다. 또한, 구조가 다른 모델은 제시된 원칙에 다르게 반응할 수 있다. 또한, 평가자에 따라 결과가 달라질 수 있는 주관성이 존재한다. 응답은 제한된 질문 세트로 평가했기 때문에 미래 연구에서 질문 범위 확대가 필요해보인다.