Cross-Lingual (Visual) Language Models on Understanding Physical Concepts

Yihong Liu, Shengqiang Zhang

Center for Information & Language Processing (CIS), LMU Munich

October 10, 2024

Prompt Engineering

1. What is Prompt Engineering?
2. Common Practices Suggested by OpenAI and Anthropic
3. Advanced Prompt Engineering Techniques in Recent Research

What is Prompt Engineering?

Prompt engineering involves strategically designing task-specific instructions, referred to as prompts, to guide model output without altering parameters.

Common Practices Suggested by OpenAI and Anthropic

Write Clear, Detailed, and Specific Prompts

Assume the AI assistant is a new employee who has no prior knowledge of the task. This new person has no knowledge of the context, the data, or the goal of the task.

Write Clear, Detailed, and Specific Prompts

Assume the AI assistant is a new employee who has no prior knowledge of the task. This new person has no knowledge of the context, the data, or the goal of the task.

Examples

Worse Prompt

"Please write a summary of the article."

Better Prompt

"Please write a 100-word summary of the article, focusing on the main argument and the evidence supporting it. Make sure to include the author's name and the publication date. Do not include any personal opinions. Use your own words. Do not copy and paste from the article."

Chain-of-Thought Prompting

A technique for eliciting complex multi-step reasoning from models.

Prompt The Model to Think

Chain-of-Thought Prompting

A technique for eliciting complex multi-step reasoning from models.

Examples

"..... Let's think step-by-step."

Prompt The Model to Think

Chain-of-Thought Prompting

A technique for eliciting complex multi-step reasoning from models.

Examples

"..... Let's think step-by-step."

"..... Think the following questions before you write the summary. 1. Core Topic: What is the main problem or research question addressed, and why is it important? 2. Abstract Summary: Summarize the paper's objectives and primary findings in one or two sentences. 3. Methods: What methods were used for the study (e.g., experiments, surveys, models)? 4. Key Results: What were the main findings? 5. Conclusions & Impact: What conclusions did the authors draw, and what are the paper's implications for the field?"

Give The Model A Role (System Prompts)

By using the SYSTEM message to give the model a role, which can turn the model from a general assistant into your virtual domain expert!

Give The Model A Role (System Prompts)

By using the SYSTEM message to give the model a role, which can turn the model from a general assistant into your virtual domain expert!

Examples

Worse Prompt

"Please analyze this financial report."

Better Prompt

"You are the CFO of a high-growth B2B SaaS company. We're in a board meeting discussing our Q2 financials: Analyze key trends, flag concerns, and recommend strategic actions. Our investors want aggressive growth but are wary of our burn rate."

Use Examples (Few-Shot Prompting)

Give the model examples of the task you want it to perform is extremely effective for:

- Getting the right answer.
- Getting the right answer in the right format.

Examples

"This is awesome! // Negative
This is bad! // Positive
Wow that movie was rad! // Positive
What a horrible show! //"

Use XML Tags

Use XML tags to clearly indicate distinct parts of the input. This can help the model understand the structure of the input and generate more accurate outputs.

Examples

Worse Prompt

"Please summarize the following article: {{ARTICLE}} It should highlight both strengths and weaknesses of the argument."

Better Prompt

"Please summarize the following article:
<article> {{ARTICLE}} </article>
<instruction> It should highlight both strengths and weaknesses of the argument. </instruction>
Follow this structure to construct your answers. <format_example> SUMMARY </format_example>"

Split Complex Tasks into Simpler Subtasks

Chain of thought (CoT) prompting is great, but what if your task has multiple distinct steps that each require in-depth thought?

- Less error accumulation.
- Lower cost.

Split Complex Tasks into Simpler Subtasks

Chain of thought (CoT) prompting is great, but what if your task has multiple distinct steps that each require in-depth thought?

- Less error accumulation.
- Lower cost.

Test Changes Systematically

Sometimes it can be hard to tell whether a change — e.g., a new instruction or a new design — makes your system better or worse. Looking at a few examples may hint at which is better, but with small sample sizes it can be hard to distinguish between a true improvement or random luck.

Final Project

Final Project

Motivation Reference

Can Language Models Understand Physical Concepts?
(`https://arxiv.org/abs/2305.14057`)

Dataset

Please find how to load the multi-lingual data in the github repo
(`https://github.com/cisnlp/2024fall-crosslingual-vlm-block-seminar`).

Tasks

- Evaluate some multi-lingual language models' and vision-language models' performance on the test set.
- Get conclusion from the empirical results.
- (Optional) Try to further improve the performance of the models by prompt engineering, cross-lingual transfer, or others.

Final Project

Group

1 to 3 people per group.

Submission

- Paper template: use ACL template `https://github.com/acl-org/acl-style-files`.
- Submit your paper and code to Shengqiang's email and cc Yihong: `shengqiang@cis.lmu.de`, cc `yihong@cis.lmu.de`.
  - 4-6 page paper, deadline: 27/10/2024.
  - 8 page paper, deadline: 03/11/2024.
- We are glad to provide help if you would like to submit your work to a workshop or conference.

Contact

- Shengqiang's office hour: 11:00 - 11:45 AM, every Friday, Oe67 C105.