

Prompt Engineering

in Enhancing reasoning ability

Jingfeng Yang et al., 2023

Content

1. Intro to Prompts Engineering
2. Why Prompt ?
3. Elements of a Prompt
4. Advanced Techniques
5. Conclusion
6. Further work
7. Reference

Intro to Prompts Engineering

- **Prompts** involve instructions and context passed to a language model to achieve a desired task
- **Prompt engineering** is the practice of developing and optimizing prompts to efficiently use language models (LMs) for a variety of applications
 - Prompt engineering is a useful skill for AI engineers and researchers to improve and efficiently **use** language models

Why Prompt ?

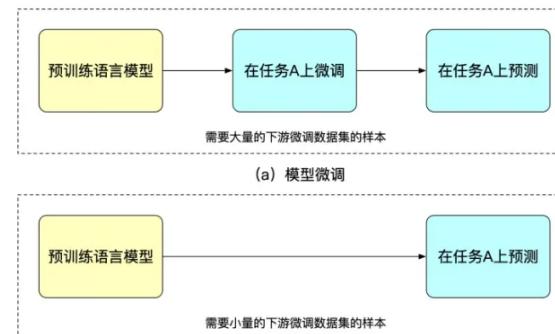
Before : Fine-tunning LLMs for every different task.

- Pro: Quite accurate in specific task but with less generalization in multitask.
- con: Computing resources and time consuming.

Recently : Prompt to adapt to downstream tasks.

Prompt can enables all kinds of innovative applications on top of LLMs.

- Pro: Better generalization, much less computing consuming,
- con: Not as accurate as finetuned LLM in specific field.



Elements of a Prompt

A prompt is composed with the following components:

- Instruction
- Context
- Input data
- Output indicator

Basic prompt

User

Classify the text into neutral, negative or positive.

Text: I think the food was okay.

Sentiment:

Bot

Sentiment: Neutral.

Advanced Techniques

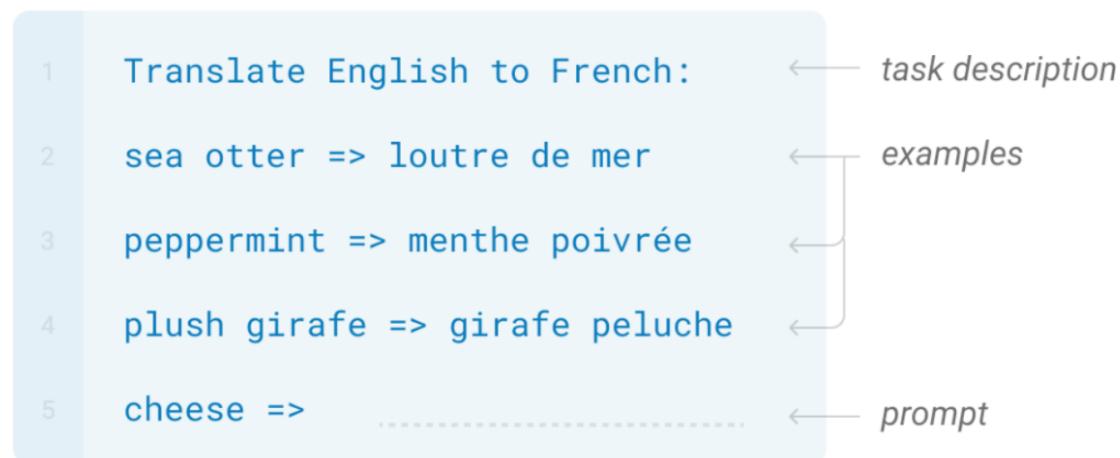
Many advanced prompting techniques have been designed to improve performance on complex tasks

- Few-shot Prompting
- Chain-of-thought Prompting(CoT)
- Zero-Shot CoT
- APE
- Self-consistency
- Generated Knowledge Prompting
- Progressive-Hint Prompting(PHP)

Few-shot prompts

Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.



Chain-of-thought(CoT)

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27.

Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

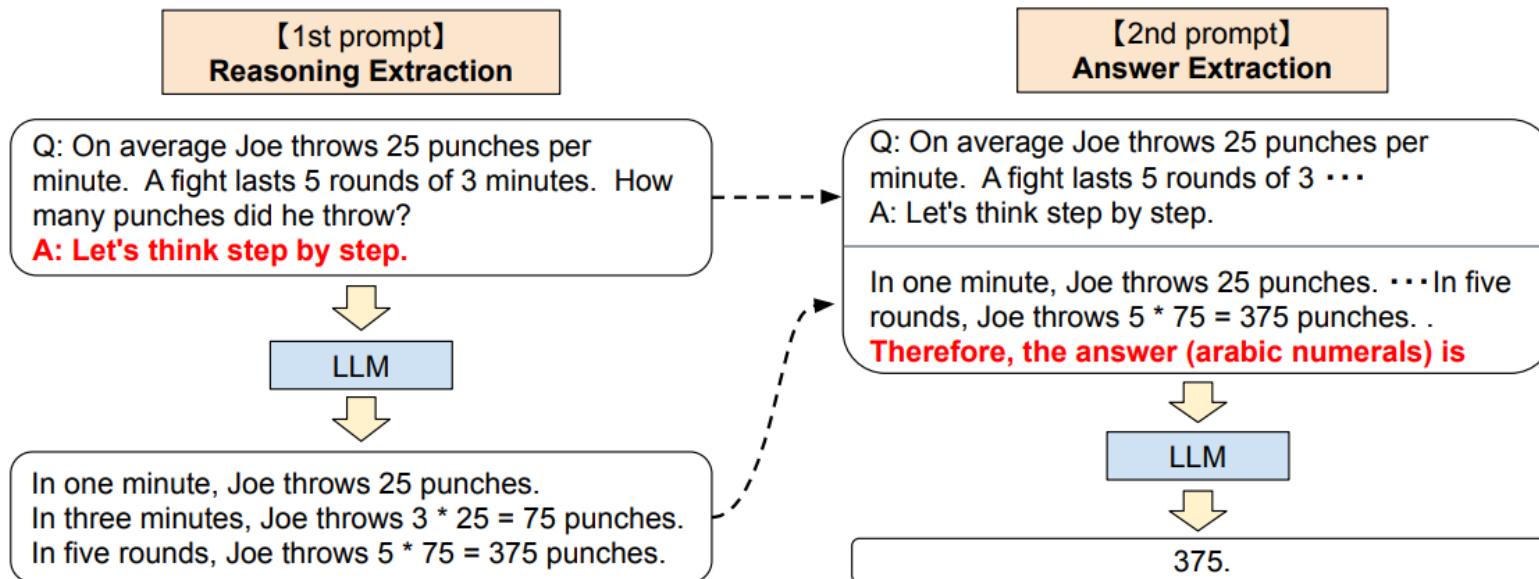
Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9.

Zero-Shot CoT

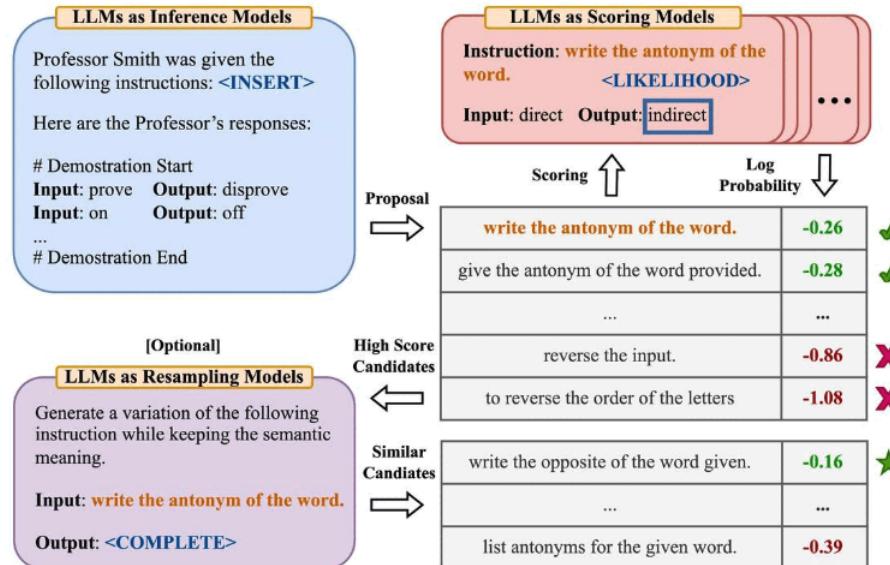
One recent idea that came out more recently is the idea of zero-shot CoT (Kojima et al. 2022) that essentially involves adding "Let's think step by step" to the original prompt.



APE

Zhou et al., (2022) propose automatic prompt engineer (APE) a framework for automatic instruction generation and selection.

Better zero-shot CoT prompt: "**Let's work this out in a step by step way to be sure we have the right answer.**"



Self-consistency

Self-consistency

Q: If there are 3 cars in the parking lot and 2 more cars arrive, how many cars are in the parking lot?

A: There are 3 cars in the parking lot already. 2 more arrive. Now there are $3 + 2 = 5$ cars. The answer is 5.

...
Q: Janet's ducks lay 16 eggs per day. She eats three for breakfast every morning and bakes muffins for her friends every day with four. She sells the remainder for \$2 per egg. How much does she make every day?

A:

Language model

Sample a diverse set of reasoning paths

She has $16 - 3 - 4 = 9$ eggs left. So she makes $\$2 * 9 = \18 per day.

The answer is \$18.

This means she sells the remainder for $\$2 * (16 - 4 - 3) = \26 per day.

The answer is \$26.

She eats 3 for breakfast, so she has $16 - 3 = 13$ left. Then she bakes muffins, so she has $13 - 4 = 9$ eggs left. So she has $9 \text{ eggs} * \$2 = \18 .

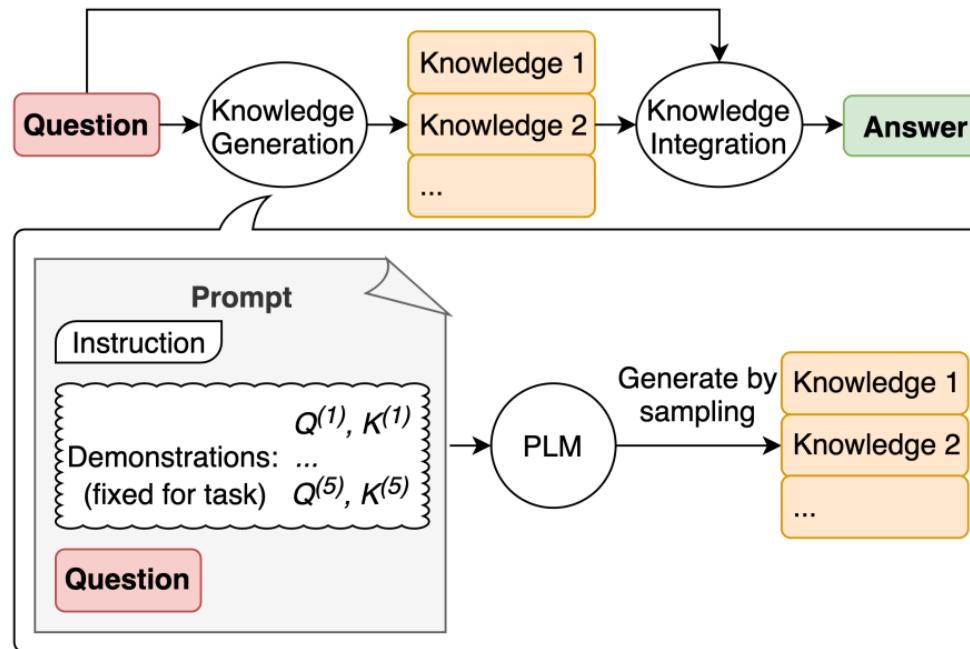
The answer is \$18.

Marginalize out reasoning paths to aggregate final answers

The answer is \$18.

Generated Knowledge Prompting

Incorporate knowledge(LLM generated) to help the model make more accurate predictions



Progressive-Hint Prompting(PHP)

Base Prompting

Q: There are 96 fourth-graders at Small Tree School. 43 of them are girls. On Friday, 5 fourth-grade girls and 4 fourth-grade boys were absent. How many fourth grade boys were at Small Tree School on Friday?



A: There were 96 fourth-graders at Small Tree School. 43 of them were girls. On Friday, 5 girls and 4 boys were absent. So there were 43 - 5 = 38 girls at school on Friday, and 96 - 38 = 58 boys at school on Friday. The answer is **58**.

Progressive-Hint Prompting

Q: There are 96 fourth-graders at Small Tree School. 43 of them are girls. On Friday, 5 fourth-grade girls and 4 fourth-grade boys were absent. How many fourth grade boys were at Small Tree School on Friday? **(Hint: The answer is near to 58).**



A: We know the Answer Hints: **58**. With the Answer Hints: **58**, we will answer the question. There were 96 fourth-graders at Small Tree School, and 43 of them were girls. On Friday, 5 girls and 4 boys were absent. So there were 43 - 5 = 38 girls and 96 - 43 = 53 boys at Small Tree School on Friday. After subtracting the absent students, there were 38 - 4 = 34 girls and 53 - 5 = 48 boys. The answer is **48**.

Progressive-Hint Prompting

Q: There are 96 fourth-graders at Small Tree School. 43 of them are girls. On Friday, 5 fourth-grade girls and 4 fourth-grade boys were absent. How many fourth grade boys were at Small Tree School on Friday? **(Hint: The answer is near to 58, 48).**



A: We know the Answer Hints: **58, 48**. With the Answer Hints: **58, 48**, we will answer the question. There were 96 fourth-graders, and 43 of them were girls. On Friday, 5 girls and 4 boys were absent. So there were 43 - 5 = 38 girls and 96 - 9 = 87 students in total. Of those 87 students, 38 were girls, so there must have been 87 - 38 = 49 boys. The answer is **49**.

Progressive-Hint Prompting method combines the generated answers and questions for double-checking purposes.

MATH Dataset

Problems are tagged by difficulty from 1 to 5.

A CS PhD student attained approximately **40%** on **MATH**, while a three-time IMO gold medalist attained **90%**.

Metamath Theorem Proving

$n \in \mathbb{N} \wedge \frac{n+1}{2} \in \mathbb{N} \implies \exists m \in \mathbb{N} : n = 2m + 1$.
GPT-f's generated proof:

```
| - ((N e. NNO /\ ((N + 1) / 2) e. NNO) -> ((N - 1) / 2) e. NNO)
| - (N e. NNO -> N e. CC)
| - 1 e. CC
| - ((N e. CC /\ 1 e. CC) -> (N - 1) e. CC)
:
:
```

DeepMind Mathematics Dataset

Divide 1136975704 by -142121963.
A: -8
Let $k(u) = u^{**2}+u-4$. Find $k(0)$.
A: -4
Sort 2, 4, 0, 6.
A: 0, 2, 4, 6
Solve $4 - 4 - 4 = 188*m$ for m .
A: -1/47

MATH Dataset (Ours)

Problem: Tom has a red marble, a green marble, a blue marble, and three identical yellow marbles. How many different groups of two marbles can Tom choose?

Solution: There are two cases here: either Tom chooses two yellow marbles (1 result), or he chooses two marbles of different colors ($\binom{4}{2} = 6$ results). The total number of distinct pairs of marbles Tom can choose is $1 + 6 = \boxed{7}$.

Problem: The equation $x^2 + 2x = i$ has two complex solutions. Determine the product of their real parts.

Solution: Complete the square by adding 1 to each side. Then $(x + 1)^2 = 1 + i = e^{\frac{i\pi}{4}}\sqrt{2}$, so $x + 1 = \pm e^{\frac{i\pi}{8}}\sqrt[4]{2}$. The desired product is then $(-1 + \cos(\frac{\pi}{8})\sqrt[4]{2})(-1 - \cos(\frac{\pi}{8})\sqrt[4]{2}) = 1 - \cos^2(\frac{\pi}{8})\sqrt{2} = 1 - \frac{(1+\cos(\frac{\pi}{4}))}{2}\sqrt{2} = \boxed{\frac{1-\sqrt{2}}{2}}$.

Conclusion

- Prompt is quite effective in enhancing or activating LLM's reasoning ability.
- When prompting, write clear and specific instructions and give the model time to "think".
- Some of these techniques are orthogonal.

which means they can be put into practice all together without conflicts.

For example: CoT combined with PHP and Few Shot achieves SOTA in MATH dataset.

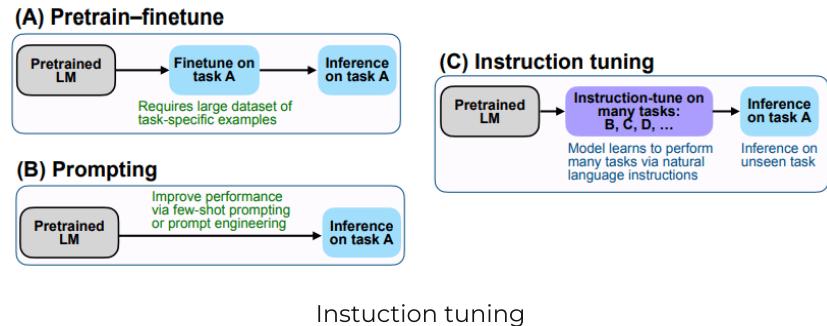
Further work

- Some other prompt techniques
 - Active-Prompt
 - ReAct Prompting
 - Multimodal CoT Prompting
 - ...
- Instruction tuning

finetuning + prompt learning = Instruction tuning

- How to Pretrain-finetune ?

Start to finetune some models like chatGLM.



Reference

1. [Learn Prompting](#)
2. [Prompt Engineering Guide](#)
3. [ChatGPT3-Free-Prompt-List](#)
4. [Language Models are Few-Shot Learners](#)
5. [Large Language Models are Zero-Shot Reasoners](#)
6. [Generated Knowledge Prompting for Commonsense Reasoning](#)
7. [Self-Consistency Improves Chain Of Thought Reasoning In Language Models](#)
8. [Large Language Models Are Human-Level Prompt Engineers](#)
9. [Progressive-Hint Prompting Improves Reasoning in Large Language Models](#)
10. [Measuring Mathematical Problem Solving With the MATH Dataset](#)