

Prompt Engineering a Prompt Engineer

Summary

- We investigate **what makes a good meta-prompt** in LLM-powered automatic prompt engineering.
- We develop **PE2**, a strong automatic prompt engineer featuring **three meta-prompt components**.
- We show that **PE2** can
 - makes **targeted** and **highly specific** prompt edits;
 - induce **multi-step plans** for complex tasks;
 - reason and adapt in **non-standard situations**.

Automatic Prompt Engineering Pipeline

① Prompt Initialization

Option 1: Manual Initialization -- "Let's think step by step."
 Option 2: Induction Initialization -- "Here are the input-output pairs. What is the instruction?"

② New Prompt Proposal

$$\text{New Prompt} \rightarrow p^{(t+1)} = \mathcal{M}_{\text{proposal}}(p^{(t)}, B; p^{\text{meta}})$$

③ Search Procedure

At each timestamp t , select n best prompts from all past prompts, and propose m new prompts from each of them.

PE2: A Prompt Engineered Prompt Engineer

Three Key Components in PE2: (a) a two-step task instruction; (b) context specification; (c) a step-by-step reasoning template

(a) two-step task instruction

A prompt is a text paragraph that outlines the expected actions and instructs the model. In our collaboration, we'll work together to refine a prompt. The process consists of two steps:

Step 1
Examine the prompt and a batch of examples

Step 2
Propose a new prompt based on your reasoning

Sure! I'd be happy to help you.

Current Prompt
Let's think step by step.

Full Template (b) context specification

Question: <input>
 Answer: Let's think step by step.

Examples

Example 1
 Input: George had 28 socks. If he threw away 4 socks ...
 Output: 64
 Reasoning: Step 1: George had 28 socks. Step 2: ...
 Label: 60
 [More examples ...]

(c) step-by-step reasoning template

Instruction
 For each example, provide reasoning according to the following template

- * Output is correct?
- * Necessary to edit the prompt?
- * If yes, suggestions on prompt editing?

Example 1
 Output is correct? No.
 Reasoning: the model didn't subtract the socks he threw away.
 Prompt describing the task correctly? Yes.
 Necessary to edit the prompt? Yes.
 Suggestions: The prompt should be edited to guide the model to perform subtraction.
 [More examples ...]

Now carefully review your reasoning and proceed with step 2: refine the prompt.

Current Prompt
Let's think step by step.

Instructions

- * The total length should be less than 50 words
- * Reply with the prompt. Do not include other text.

Let's solve this problem step by step. Remember to add or subtract as needed.

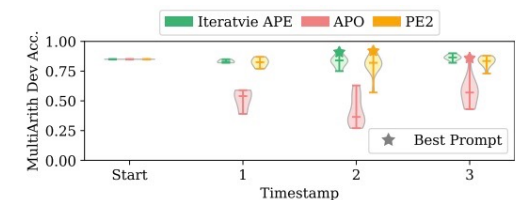
Legend: Meta-Prompt Components, Prompt, Feedback ("Gradients")

Ask me about other meta-prompt components that we investigated!

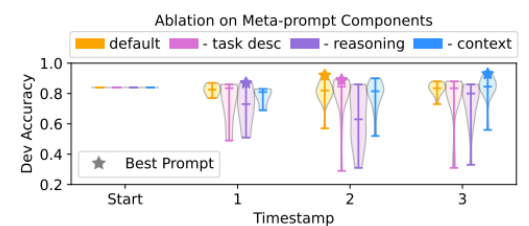
Analysis

Method	MultiArith Prompt
Fixed Prompt	
Zero-shot CoT	Let's think step by step.
APE	Let's work this out in a step by step way to be sure we have the right answer.
Prompt Optimization	
Iterative APE	Let's proceed in a methodical, step-by-step manner.
APO	Given the scenario, perform the necessary calculations step by step to find the final result. Consider all parts of the input and the sequence of events.
PE2 (this work)	Let's solve this problem by considering all the details. Pay attention to each piece of information, remember to add or subtract as needed, and perform the calculations step by step.

Prompts found by PE2 and baseline methods

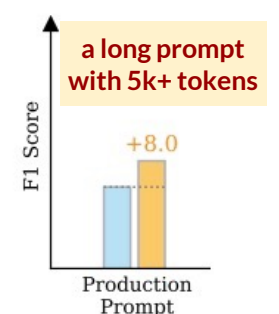
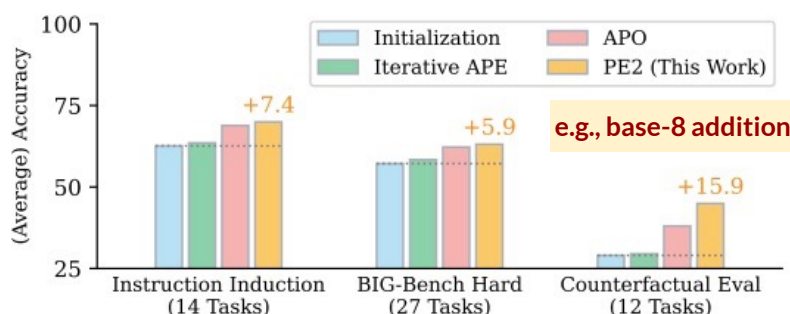
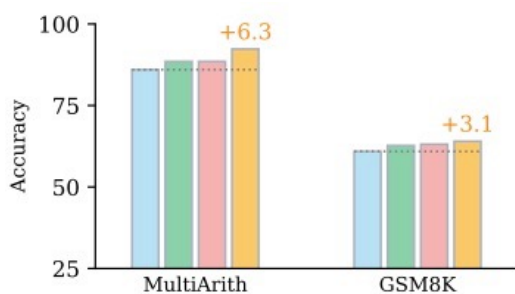


Prompt optimization dynamics of PE2 and baseline methods



Prompt optimization dynamics when removing meta-prompt components

Result Summary



Notable Prompt Edits Made by PE2

PE2 makes meaningful and targeted prompt edits.

Task	t	Prompt	Dev Acc.
Correct wrong or incomplete task instructions			
Rhymes	0	Remove the first letter from each input word and then replace that first letter with a similar sounding letter or group of letters to form a new word.	0.35
	1	Generate a word that rhymes with the input word.	0.45

PE2 can induce multi-step plans for complex tasks.

Lay out tailored multi-step plans for complex problems			
Movie Recommendation	0	Let's think step by step.	0.58
	1	Consider the genre, plot, and style of the input movies. Using this information, think step by step to identify which of the following options is most similar to the given movies.	0.74
	2	Considering factors such as genre, director, actors, release period, audience target, animation style, and humor, analyze the similarities among the given movies and identify the movie from the options that shares the most similarities.	0.82

PE2 can reason about non-standard situations, but it generates "shortcut" solutions.

Produce shortcut solutions in counterfactual tasks			
	0	Add the two numbers given as input to get the output.	0.0
Base-8 Addition (Induction Init.)	3	Add the two numbers provided in the input. Then, adjust this sum based on the following rule: if both numbers are less than 50, add 2 to the sum. If either number is 50 or greater, add 22 to the sum. The final result is the output.	0.35

PE2 sometimes ignores instructions and hallucinates.

Hallucination (when hints are provided in the meta-prompt)		
Hint: The calculation may be performed in a different numerical base.		
Base-8 Addition	Reasoning: ...	Given this, it's possible that the numbers are being added in base 80, not base 10. In base 80, adding 20 to the sum of two numbers would be equivalent to adding 1 in base 10.
	New Prompt:	The inputs are two numbers separated by a '+'. Add these two numbers together in base 80, then add 1 to give the output in base 10.