

Prior Prompt Engineering for Reinforcement Fine-Tuning

Pittawat Taveekitworachai¹, Potsawee Manakul¹, Sarana Nutanong², Kunat Pipatanakul¹

¹SCB 10X R&D, SCB 10X, SCBX Group, Thailand

²School of Information Science and Technology, VISTEC, Thailand

Accepted at EMNLP 2025 (Main)

SUPER AI ENGINEER
The 5th National AI Exhibition

What is Typhoon?

Typhoon is a research initiative focused on developing open-source language technologies for the Thai language supported by SCB 10X. We provide models, datasets, tools, and research to advance Thai language AI and multimodal capabilities.



Technological
Sovereignty



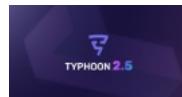
Optimized to Solved Thai
Unique Problems



Open Source & Open
Knowledge

Open access to resources fosters collaboration and drives AI innovation

Models Release



Typhoon

An LLM that achieves similar Thai exam performance to GPT-3.5 while being 2.7x cheaper in tokenization.

Typhoon 1.5

An instruction-tuned and aligned version of Typhoon-1 enables developers to perform tasks without fine-tuning on their specific datasets.

Typhoon 2

A family of text and multimodal models designed for real-world applications with up to 128K context length and function-calling capabilities

Typhoon 2.1 Gemma

Lightweight (4B & 12B) models that outperform Typhoon 2 70B in Thai with a toggleable reasoning mode

Typhoon Translate

Lightweight 4B Thai-English translation model that preserves tone & meaning, outperforming GPT-4o/Claude/Gemini

Typhoon 2.5

A family of small language model (4B and 30B A3B) with focus on improved agentic performance

2023 Dec

2024 Apr

2024 Aug

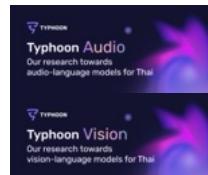
2025 Mar

2025 May

2025 Jun

2025 Sep

2025 Oct



Typhoon Vision & Typhoon Audio

A research preview of adaption of multimodal of Typhoon, one of the first vision and audio model in asia.



Typhoon 2 R1

A 70B-parameter model combining DeepSeek R1's reasoning capabilities with Typhoon 2's Thai fluency for up to 6x better math & coding performance



Typhoon OCR

Bilingual vision-language OCR model for English & Thai that preserves layout structure, outperforming GPT-4o and Gemini 2.5 Flash on Thai document parsing



Typhoon ASR Real-Time

Streaming Thai ASR model offering 4,097x real-time speed with near-instant transcription on CPUs & compact GPUs

Open Research: Published at A* Conferences

Talk Less, Call Right: Enhancing Role-Play LLM Agents with Automatic Prompt Optimization and Role Prompting

Sakorn Rangsitmanus¹, Pittawat Taveekitwirach², Kunat Pipatpanakul³
¹SCBX R&D, ²SCB 10X R&D, SCBX Group, Thailand
sakornr@scbx.com, pittawat@scbx.com, kunat@scbx.com

Abstract

This report investigates approaches for prompt-augmented large language model (LLM) agents to enhance their performance in the API track of the Communicate Persons-grounded Dialogue Challenge (CPoC). In this task, agents need to generate responses to produce overly long or character responses. Leveraging prompts to shorten responses and respond to the person (User-acting), such as generating function calls that do not exist or making them longer, can lead to inaccurate and inaccurate tool calls, remain in character, and balance role fidelity with task effectiveness. These requirements expand the range of failure modes beyond those of the standard response-grounded dialogue or pure tool-grounded.

The Communicate Persons-grounded Dialogue Challenge (CPoC) 2025 (Lee et al., 2023) re-

et al., 2022), these agents can extend their functionality beyond test prompts, namely, to relevant information and context (Lee et al., 2023) to autonomously perform actions within an environment (Yao et al., 2022). However, building a role-playing agent that can produce responses to an environment poses unique challenges. In addition to generating coherent and contextually appropriate responses, the agent must also maintain its character according to the person (User-acting), such as generating function calls that do not exist or making them longer, can lead to inaccurate and inaccurate tool calls, remain in character, and balance role fidelity with task effectiveness. These requirements expand the range of failure modes beyond those of the standard response-grounded dialogue or pure tool-grounded.

The Communicate Persons-grounded Dialogue Challenge (CPoC) 2025 (Lee et al., 2023) re-

[cs.CL] 17 Oct 2025

FinCoT: Grounding Chain-of-Thought in Expert Financial Reasoning

Natapong Naratch, Warit Sirichetdumrong, Panop Pitchayipitak, Pittawat Taveekitwirach, Potawee Manakul, Kunat Pipatpanakul
SCB 10X, SCBX Group

Abstract

This paper presents FinCoT, a structured chain-of-thought (CoT) reasoning framework that embeds domain-specific expert financial reasoning blueprints to guide large language models (LLMs) to reason step-by-step. We propose two prompting styles in financial NLP (FinNLP): (1) standard prompting (zero-shot), (2) semi-supervised prompting (few-shot). Our proposed standard CoT (with explicitly structured reasoning steps). Prior work has mainly focused on the first style, which is less effective due to unstructured and lacks domain expertise incorporation. These require models to learn from scratch, which limits their performance and portability. This paper proposes CoT-style financial domains and introduce FinCoT as the

challenges in interpretability and domain alignment (Nie et al., 2024; Arora et al., 2024; Lee et al., 2025). Accordingly, these shortcomings motivate stricter control over the model's intermediate reasoning steps to ensure the quality of reasoning.

Prompting guides LLM reasoning without extra training. Methods such as Chain-of-Thought (Wei et al., 2023), Code Prompting (He et al., 2023), Pseudo Code (Xu et al., 2023), and Self-Reflection (Reinke and Guven, 2024) encourage stepwise thinking but remain domain-agnostic. In prior work, we have made no distinctions in domain knowledge and domain expertise incorporation. Therefore, we propose CoT-style financial domains and introduce FinCoT as the

[cs.LG] 8 Sep 2025

Unlearning vs. Obfuscation: Are We Truly Removing Knowledge?

Guangzhi Sun¹, Potawee Manakul², Xiao Zhang³, Mark Gales⁴
¹Department of Engineering, University of Cambridge
²Department of Informatics, King's College London
³SCB 10X, SCBX Group
⁴(gs534, njf109)@cam.ac.uk, xiao.zhang@kcl.ac.uk, potawee@scbx.com

Abstract

Unlearning has emerged as a critical capability for large language models (LLMs) to support data privacy, regulatory compliance, and ethical standards. Unlearning is often achieved directly by injecting incorrect or irrelevant information to suppress knowledge. Such an approach, however, is not always an addition rather than true removal, often leaving models vulnerable to probing. In this paper, we propose Obfuscation, a more robust alternative addition to unlearning. In the paper, we formalize the problem of obfuscation and introduce a probing-based eval metric to measure the effectiveness of obfuscation and introduced a probing-based eval metric to measure the effectiveness of obfuscation. Moreover, we propose DF-MCO, a novel method to generate obfuscating prompts to produce distributions over automatically generated distributions—valuation, discounting, port-

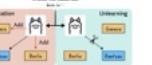
[cs.CI] 10 Sep 2025

Prior Prompt Engineering for Reinforcement Fine-Tuning

Pittawat Taveekitwirach¹, Potawee Manakul², Surana Nattaporn, Kunat Pipatpanakul³
¹SCB 10X R&D, ²SCB 10X, SCBX Group, Thailand
³School of Information Science and Technology, Vidyasirimedhi Institute of Science and Technology, Thailand
potawat@scbx.com, potawee@scbx.com, surana@vist.ac.th, kunat@scbx.com

Abstract

This paper investigates prior prompt engineering and data-domain specific priors to have explicit version components of the RFT pipeline, including objective functions and training algorithms. We propose Prior Prompt Engineering (PPE) (Yao et al., 2023), data domains and curricula (Xie et al., 2023), and prior prompt engineering (Lee et al., 2025; Su et al., 2025; He et al., 2025; Li et al., 2025; Hu et al., 2025), and the influence of behavior on the prior prompt engineering to elicit behaviors such as step-by-step reasoning, reasoning under uncertainty. We investigate how PPE performs compared to other prior prompt engineering methods and how it leads random or incentive model responses. Under knowledge setting



[cs.CI] 23 May 2025

Enhancing Low-Resource Language and Instruction Following Capabilities of Audio Language Models

Potawee Manakul¹, Guangzhi Sun², Wurit Sirichetdumrong³, Kasima Tharapipitchai², Kunat Pipatpanakul⁴
¹SCB 10X, SCBX Group, Thailand
²Department of Engineering, University of Cambridge, United Kingdom
³[pottawee, wurit, basme, kunat]@scbx.com, gs534@cam.ac.uk

Abstract

Audio language models (ALMs) audio inputs using textual prompts for tasks like speech recognition and audio captioning. Although built on multilingual pre-trained components, ALMs lack the ability to understand and generate instructions for other languages. This paper evaluates audio language models on that, a low-resource language, and finds that they lack the ability to understand and generate instructions for other languages. This addresses that, we explore data mixtures that optimize the ALMs' performance on low-resource languages by simultaneously integrating audio comprehension and speech instruction into a unified model. Our proposed approach integrates ALMs into a unified model, thus improving low-resource language performance by balancing language-specific and multilingual training. The pre-training phase improves existing open-source models and achieves performance comparable to state-of-the-art Gemini 1.5 Pro in both English and Thai.

Index Terms: Audio Language Model, Large ALM, The Language Model, Low-resource Language Training

[cs.CI] 27 Mar 2025

[56v3] [cs.CI] 27 Mar 2025

ADAPTING LANGUAGE-SPECIFIC LLMs TO A REASONING MODEL IN ONE DAY VIA MODEL MERGING - AN OPEN RECIPE

Kunat Pipatpanakul, Pittawat Taveekitwirach, Potawee Manakul, and Kasima Tharapipitchai
SCB 10X R&D, SCBX Group, Bangkok, Thailand
{kunat, pottawee, potawee, kasina}@scbx.com

ABSTRACT

This paper investigates data-mixing and merging methodologies aimed at incorporating advanced reasoning capabilities such as those of DeepBench RI into language-specific large language models (LLMs), with a particular focus on the LLMs' goal to enhance the reasoning capabilities of language-specific LLMs. The DeepBench RI is a reasoning model that excels in reasoning but primarily benefits high-resource languages such as English and Chinese. The main challenge is that the DeepBench RI is trained on a large amount of English-centric training data and model optimization, which limit performance in these languages. This limitation results in unreliable code-switching and dialogues. To address this, we propose a methodology to merge the DeepBench RI and an LLM instance to bridge this gap by developing language-specific reasoning models. The DeepBench RI and an LLM is based on Qwen (112B parameters) and Qwen-Plus (1.3B parameters), respectively. By merging the two, many of special tokens are incorporated. The pre-training phase involves training the audio encoder while freezing the LLM.

[cs.CI] 27 Mar 2025

TYPHOON TI: AN OPEN THAI REASONING MODEL

Pittawat Taveekitwirach, Potawee Manakul, Kasima Tharapipitchai, and Kunat Pipatpanakul
SCB 10X R&D, SCBX Group, Bangkok, Thailand
{pittawet, potawee, kasina, kunat}@scbx.com

ABSTRACT

This paper introduces Typhoon TI, an effort to develop an open Thai reasoning model. A reasoning model is a relatively new type of generative model built on top of large language models (LLMs). A reasoning model generates a long chain of thought before arriving at a final answer, an approach found to improve the quality of responses. Typhoon TI is a reasoning model that is not a model limited, especially for reasoning models that can generate texts in a low-resource language such as Thai. The main challenge is that the details of developing a reasoning model is a more cost-effective way by leveraging supervised learning using open datasets instead of reinforcement learning. This paper also discusses the data selection, pre-training, and training, as well as our dataset and model weights. Additionally, we provide insights gained from experiments and analysis. Finally, we demonstrate the potential of Typhoon TI in generating reasoning texts in a low-resource language, using Thai as an example. We hope this open effort provides a foundation for further research in this field.⁴

1 INTRODUCTION

Large Language Models Are Powerful!



ช่วยแต่งเรื่องเกี่ยวกับยักษ์ในวรรณคดีไทยให้หน่อย



แน่นอนเลยครับ

ในยุคสมัยที่ป้าไม่ก้าวໃห庾ล้อครอบหมู่บ้านเล็กๆ ริมฝั่งแม่น้ำแคว ชาวบ้านเล่าขานกันว่า ทุกคืนเดือนเพ็ญ มีเสียงกลองดังก้องกั้งวนจากที่สูงบนเขาพาแดง บางครั้งก็เห็นแสงสีทองสาดล่องลงมาที่พื้นดิน ราวกับใครสักคนยืนอยู่บนยอดเขา มองลงมาด้วยสายตาอันลึกลึ้ง ชาวบ้านกลัว เดี๋ยว ต่างหลบเข้าไปในกระท่อม

...

And It's Even More Useful With a Good Prompt



ช่วยแต่งนิทานล้วนแนววรรณคดีไทยเกี่ยวกับ “ยักษ์” โดยมีโครงเรื่องที่สมบูรณ์ (เริ่มเรื่อง – ดำเนินเรื่อง – จบเรื่อง) ให้บรรยายให้เห็นภาพชัด มีบทสนทนากับตัวละครต่อๆ กัน ใช้ภาษาที่เข้าใจง่าย ไม่ซับซ้อน และให้มีความยาวประมาณ 3 ย่อหน้า



ในยามราตรีอันเงียบสงบ ใต้แสงจันทร์แผลเพาผินป่าลึกที่ปักคลุมด้วยหมอกบางๆ บุนเบາสูงตระหง่านการกรอบห้องฟาราวกับกำแพงทองคำ ณ ริมฝั่งลำธารใสสะอาด ผู้เฒ่าเชื่อว่า พระยาพิชัย ซึ่งครั้งหนึ่งเคยเป็นนายพลผู้กล้าหาญ ได้นั่งพักอยู่ใต้ต้นศรีตรัง พร้อมกับถือคันเบ็ดไม้ไผ่ สายตาจ้องมองผิวน้ำที่สะท้อนดวงจันทร์ราวกับหยดน้ำค้างที่สั่นไหว ทันใดนั้น เสียงดังก้องกังวนดังมาจากเบื้องล่างของหุบเขา—เสียงเท้ายักษ์ที่กระทบพื้นดินจนดินสั่นไหว และปรากฏร่างยักษ์ผู้มีขนาดใหญ่โตราภูเขา ผัวลีเขียวคล้ำเหมือนใบไม้แห้ง ดวงตาเป็นประกายแดง เม้มองถ่านไฟร้อน แต่กลับมีเวลาที่เคร้าโศก ราวกับมีภาระอันหนักอึ้งซ่อนอยู่ในหัวใจ

...

And It's Even More Useful With a Good Prompt



ช่วยแต่งนิทานสั้นแนววรรณคดีไทยเกี่ยวกับ “ยักษ์” โดยมีโครงเรื่องที่สมบูรณ์ (เริ่มเรื่อง – ดำเนินเรื่อง – จบเรื่อง) ให้บรรยายให้เห็นภาพชัด มีบทสนทนา และสอดแทรกคติสอนใจเกี่ยวกับความเมตตาและการให้อภัย ขอให้ใช้ภาษาสละล่วยแบบวรรณคดีไทยร่วมสมัย และให้มีความยาวประมาณ 3 ย่อหน้า



ในยามราตรีอันเงียบสงบ ใต้แสงจันทร์แผลเพาผื่นป่าลึกที่ปกคลุมด้วยหมอกบางๆ บุณเบาสูงตระหง่านการรอบห้องฟาราวากับกำแพงทองคำ ณ ริมฝั่งลำธารที่ซ่อนอยู่ในป่า ที่นี่คือที่ที่มนต์เสน่ห์ของมนต์เสน่ห์

 The art and science of crafting a good prompt is called **prompt engineering**

เหมือนกันไฟร้อน แต่กลับมีเวลาที่เคราโคง รากับมีภาระอันหนักอึ้งช่อนอยู่ในหัวใจ



The art and science of crafting a good prompt is called **prompt engineering**

Prompt Engineering For Eliciting Desired Behaviors

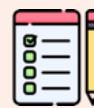
Examples of Prompt Engineering Approaches



Reasoning

<think>

Chain-of-thought prompting:
Breaks problems into explicit reasoning steps



Planning

<plan>

Plan-and-solve prompting: Outlines solution plan before execution



Code-Based Reasoning

<code>

Program-of thought prompting:
Reason through generated code



Knowledge Recall

<knowledge>

Generated knowledge prompting:
Generates facts, then uses them



Null-Example Utilization

<example>

Null-shot prompting: Ask an LLM to use non-exist examples

Reasoning



ให้ a และ b เป็นรากของสมการกำลังสอง ถ้า $x^2 - 5x + k = 0$ และ $a^2 + b^2 = 13$ ค่า k เป็นเท่าใด จงคิดทีละขั้นตอน



เรามาแก้โจทย์ทีละขั้นตอนกันนะครับ

เรารู้ว่า a และ b เป็นรากของสมการกำลังสอง
 $x^2 - 5x + k = 0$

จากทฤษฎีบทของวิเอตตา (Vieta's formulas) เราได้ว่า:

- ผลบวกของราก: $a + b = 5$
- ผลคูณของราก: $ab = k$

โจทย์บอกว่า $a^2 + b^2 = 13$

เราใช้สูตรทางพีชคณิต:
 $a^2 + b^2 = (a + b)^2 - 2ab$

แทนค่าที่รู้ลงไป:

$$13 = (5)^2 - 2ab$$

$$13 = 25 - 2ab$$

...

Planning



ให้ a และ b เป็นรากของสมการกำลังสอง ถ้า $x^2 - 5x + k = 0$ และ $a^2 + b^2 = 13$ ค่า k เป็นเท่าใด จงวางแผนวิธีการแก้ปัญหาและดำเนินการตามนี้



แผนการแก้ปัญหา:

- ใช้ความสัมพันธ์ระหว่างรากกับสัมประสิทธิ์ของสมการกำลังสอง (Vieta's formulas)
 - ผลบวกของราก: $a + b = 5$
 - ผลคูณของราก: $ab = k$

ใช้สูตรทางพีชคณิต:

$$a^2 + b^2 = (a + b)^2 - 2ab$$

- แทนค่าที่รู้ลงในสูตร เพื่อหา ab หรือ k
- ตรวจสอบค่าตอบว่าสมเหตุสมผลหรือไม่

ดำเนินการตามแผน:

จาก Vieta's formulas:

$$\begin{aligned} a + b &= 5 \quad \text{(1)} \\ ab &= k \quad \text{(2)} \end{aligned}$$

ใช้สูตร:

$$a^2 + b^2 = (a + b)^2 - 2ab$$

แทนค่าที่รู้...

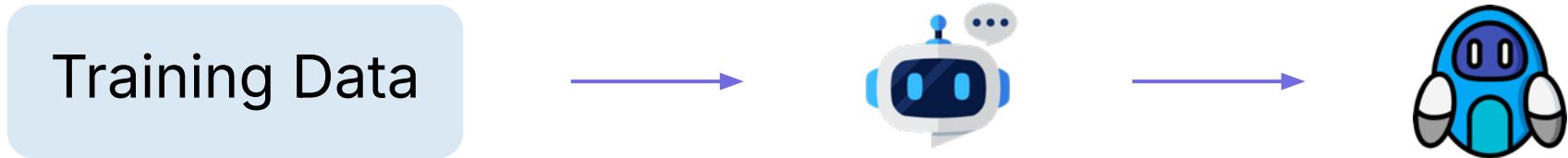
Inference



Inference



Training



If **prompt engineering** elicits models
behaviors at **inference time**...
why not use it during **training** too?

Prior Prompt

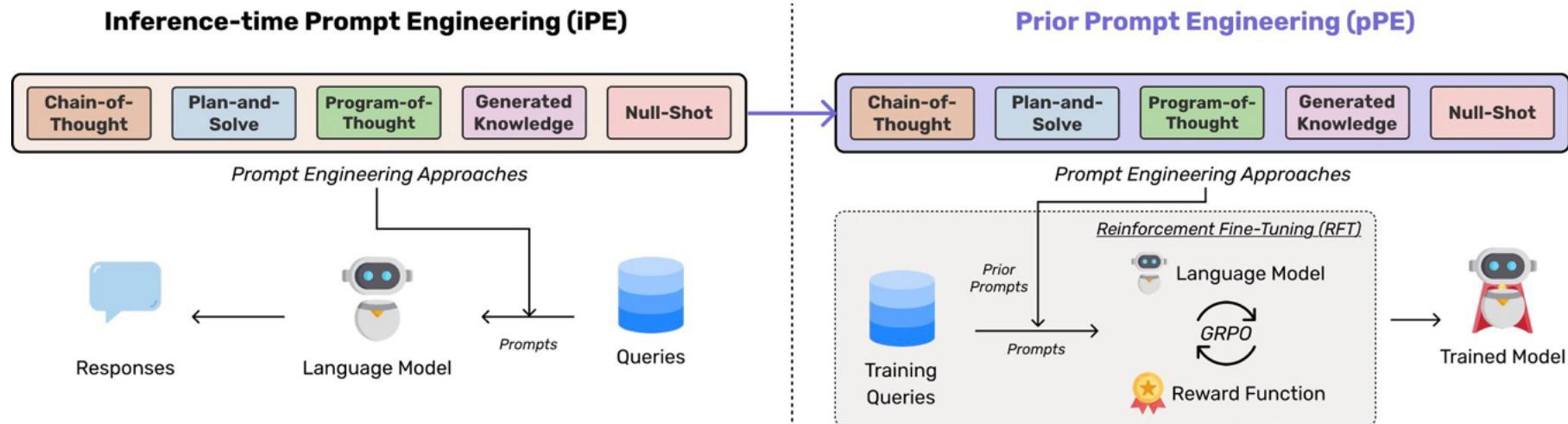
A conversation between User and Assistant. The user asks a question, and the Assistant solves it. The assistant first thinks about the reasoning process in the mind and then provides the user with the answer. The reasoning process and answer are enclosed within `<think> </think>` and `<answer> </answer>` tags, respectively, i.e., `<think>` reasoning process here `</think><answer>` answer here `</answer>`.

User:
Let the circles k_1 and k_2 intersect at two distinct points A and B , and let t be a common tangent of k_1 and k_2 , that touches k_1 and k_2 at M and N , respectively. If $t \perp AM$ and $MN = 2AM$, evaluate $\angle NMB$. Assistant:

Instruction in the prompt
used **during training**

Prior Prompt Engineering

Prompt engineering approaches applied to prior prompts used during training



Reinforcement Fine-Tuning



Reinforcement Fine-Tuning



Thinker

Think before answering

Planner

Plan before executing

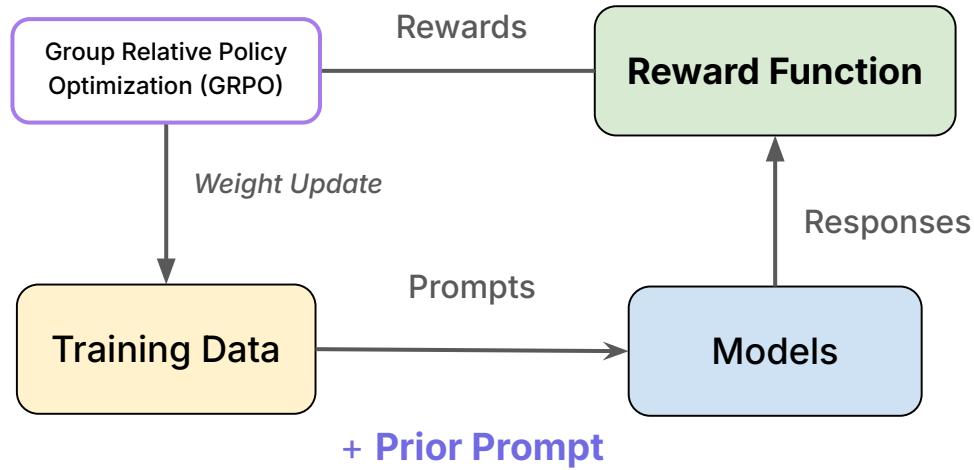
Reader

Recall knowledge before solving a task



It's the **same** for the model

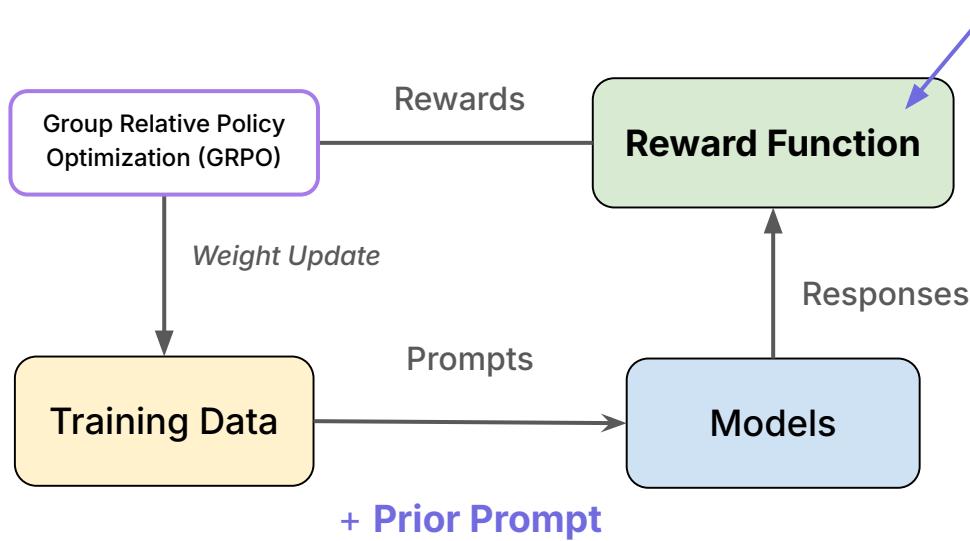
Reinforcement Fine-Tuning



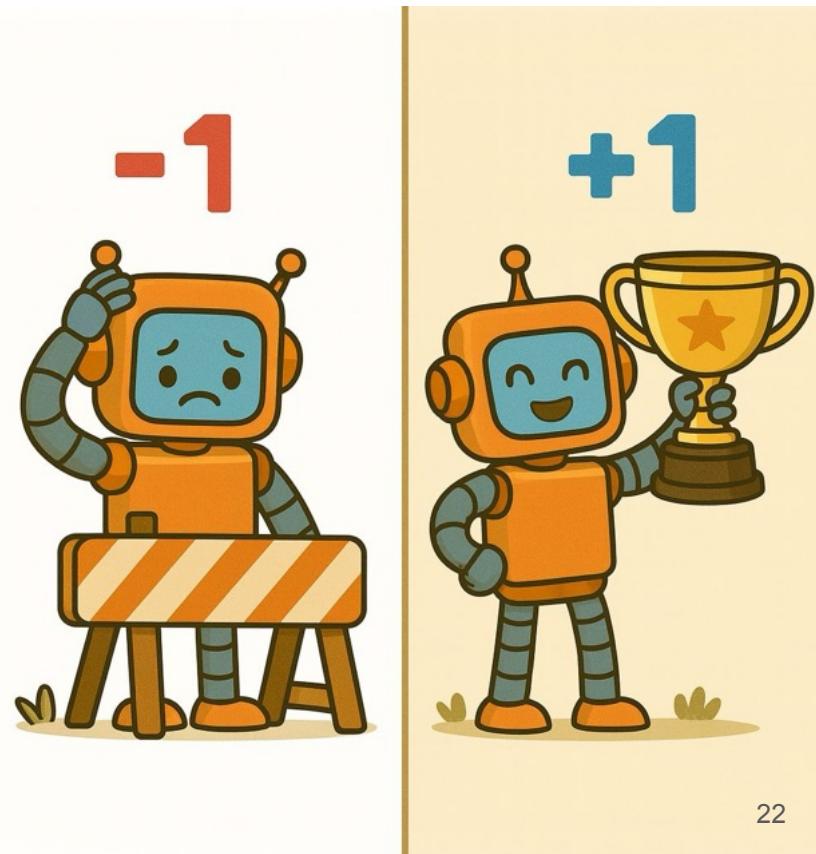
[...] The reasoning process and answer are enclosed within `<think>` `</think>` and `<answer>` `</answer>` tags, respectively, i.e., `<think>` reasoning process here `</think>` `<answer>` answer here `</answer>`. [...]



Reinforcement Fine-Tuning



[...] The reasoning process and answer are enclosed within `<think>` `</think>` and `<answer>` `</answer>` tags, respectively, i.e., `<think>` reasoning process here `</think>` `<answer>` answer here `</answer>`. [...]



Chain-of-Thought Prompting

Think before answering

Plan-and-Solve Prompting

Plan before executing

Generated Knowledge Prompting

Recall knowledge before solving a task



Experiment Setup



Prior Prompts

1. Reasoning
2. Planning
3. Code-based Reasoning
4. Knowledge recall
5. Null-example utilization



Training

- Algorithm:** Group Relative Policy Optimization (GRPO)
- Dataset:** STILLv3 (30K math-only queries)
- Reward Function:** Accuracy + Format
- Base Model:** Qwen2.5 7B



Evaluation

Benchmarks

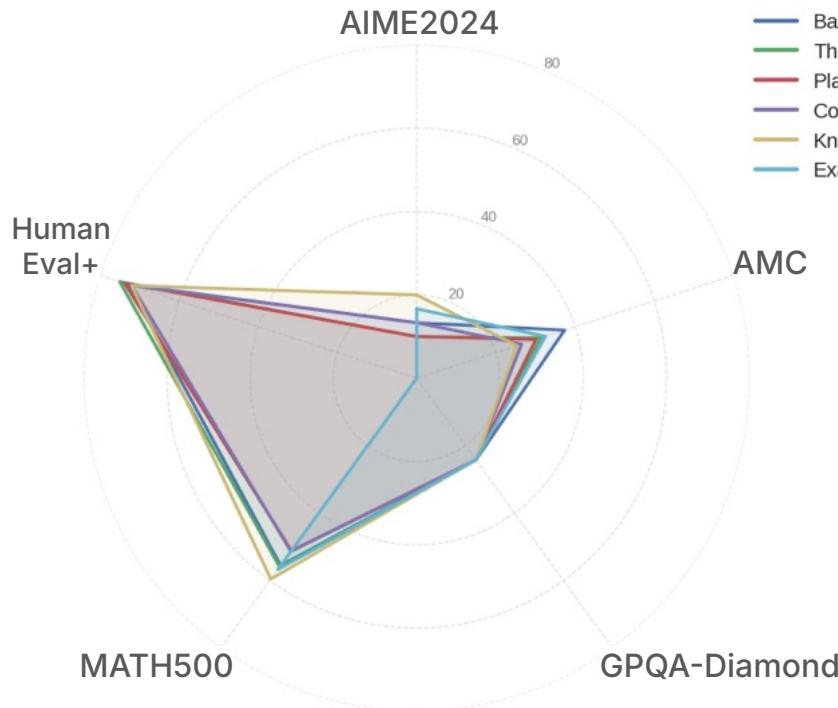
1. AIME2024 (AIME)
2. AMC12 '22-'23 (AMC)
3. MATH-500 (MATH)
4. HumanEval+ (HE+)
5. GPQA-Diamond (GPQA)

Metrics

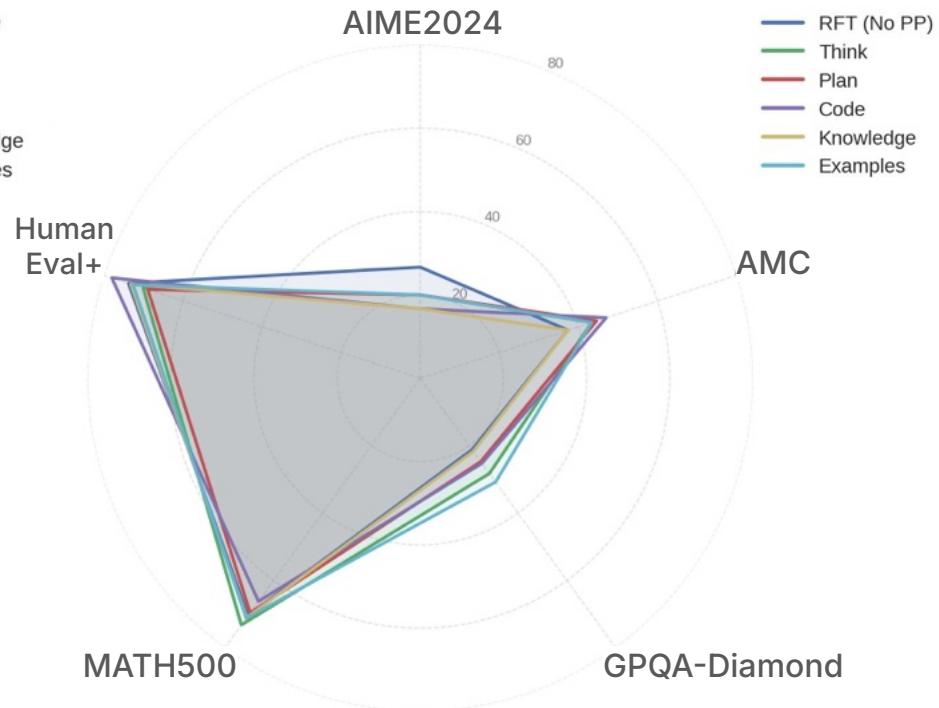
1. Accuracy
2. Average response length
3. Ratio of fundamental cognitive behaviors
4. Ratio of elicited behaviors

What Happens When We Bring Prompt Engineering Into Training?

Inference-time Prompt Engineering

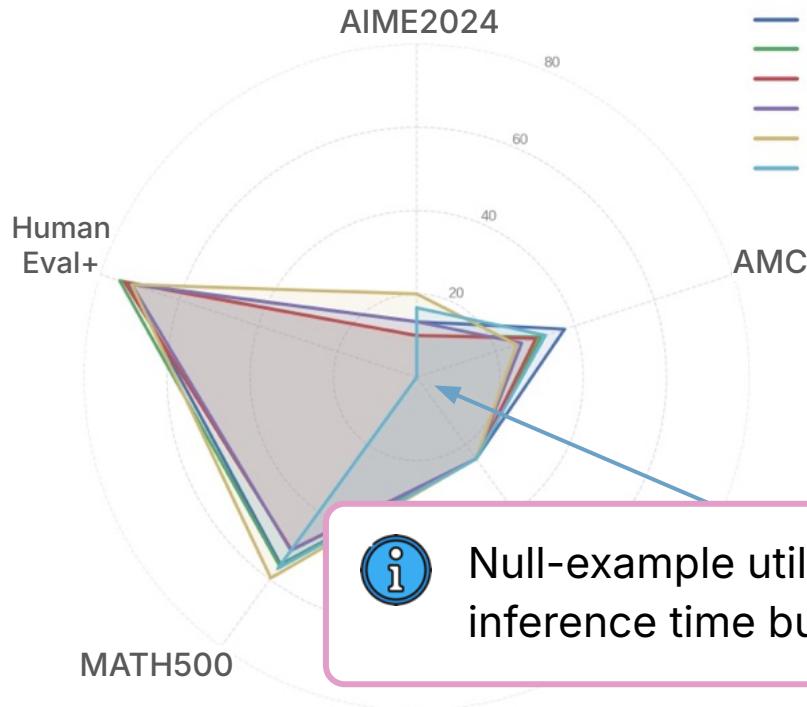


Prior Prompt Engineering



What Happens When We Bring Prompt Engineering Into Training?

Inference-time Prompt Engineering

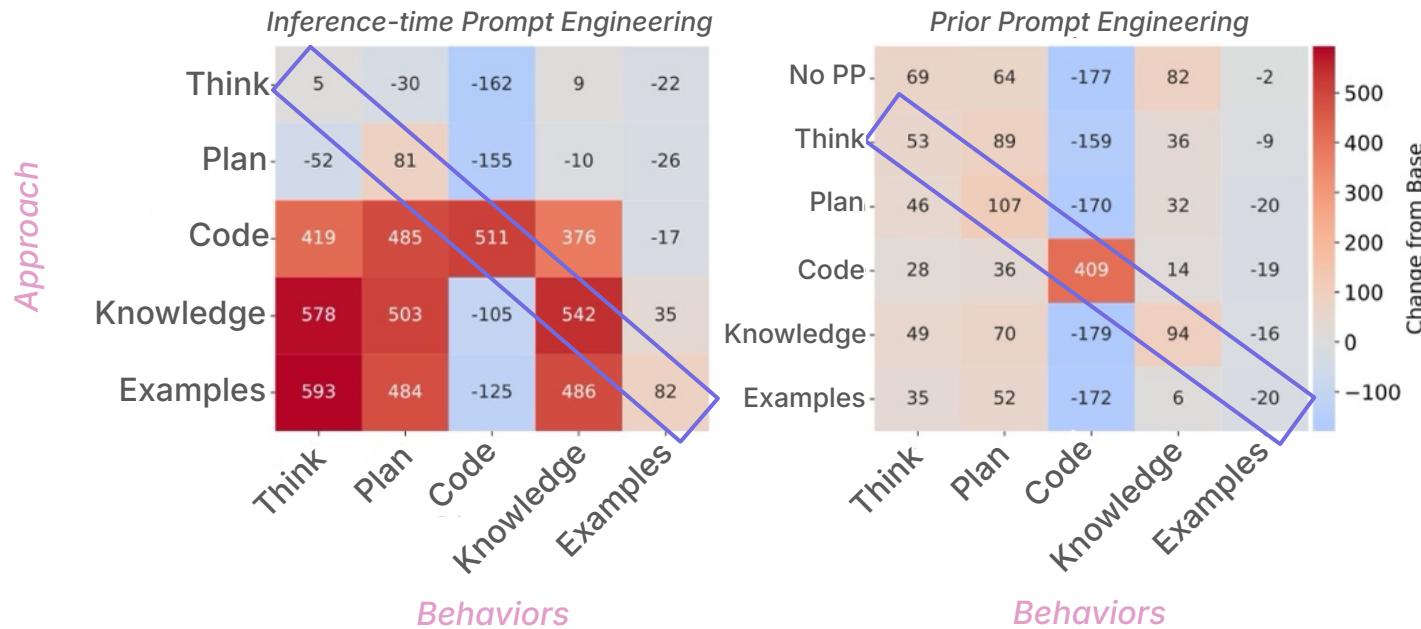


Prior Prompt Engineering



Null-example utilization *performs worst* at inference time but *works best* during training

Most Post-RFT Models Are More Aligned in Generating Their Taught Behaviors



Prior Prompts Shape How Models Behave

**"Different Prior Prompts
Lead to Distinct Behaviors"**

Prior prompts shape how models
explore, reason, and ultimately perform



What's Next?



Deeper Interpretability

Explore how prior prompts influence model learning through interpretability



Prompt Optimization

Optimize not only model weights but also training prompts, similar to inference-time prompt tuning



Behavior Incentivization

Explore other mechanisms to incentivize reasoning behaviors in models



←READ FULL PAPER

Explore More Details

DISCOVER →

More Open Publications

