

Prior Prompt Engineering for Reinforcement Fine-Tuning

Pittawat Taveekitworachai¹, Potsawee Manakul¹, Sarana Nutanong², Kunat Pipatanaku¹

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

²*School 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 an **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.

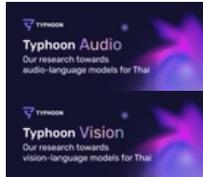
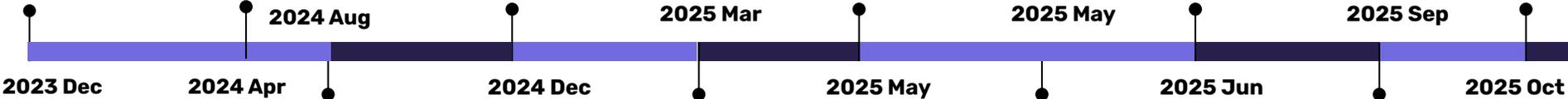
Typhoon 1.5
An instruction-tuned and aligned version of Typhoon-1 enables developers to perform tasks without fine-tuning on their specific datasets.

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



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

Sakura Ruangtornasak¹, Pittawat Tavakhiorwachai¹, Kanat Pijitwanakul¹
¹SCBX R&D, ²SCB IOX R&D, SCBX Group, Thailand
 sakura.r@scbx.com, pittawat@scbx.com, kanat@scbx.com

Abstract

This report investigates approaches for prompting a role-play LLM agent to act as a role-playing dialogue agent in the API task of the Communication Process-based Dialogue Challenge (CPDC) 2025. In this setting, dialogue agents often produce overly long in-character responses (over-speaking) while failing to do tasks effectively according to the persona (under-acting), such as providing function calls that do not exist or making unnecessary tool calls before answering. We explore four prompting approaches and address these issues: 1) basic role prompting, 2) improved role prompting, 3) automatic prompt optimization (APO), and 4) role-based role prompting. The role-based role prompting

et al., 2025), these agents can extend their functionality beyond text generation, retrieving relevant information on demand (Liu et al., 2023) and autonomously performing actions within an environment (Zhou et al., 2023). However, building a role-playing agent that can act autonomously within an environment poses unique challenges. In addition to generating coherent and contextually appropriate dialogue, such agents must make dynamic and accurate tool calls, remain in character, and balance role fidelity with task effectiveness. These requirements expand the range of factors modeled beyond those in traditional persona-grounded dialogue or pure text-generation.

The Communication Process-based Dialogue Challenge (CPDC) 2025¹ (Zhou et al., 2025) pro-

[2] [cs.CL] 12 Oct 2025

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

Natapon Nitarach, Warit Sirichotdomrong, Panop Pichayathorn,
 Pittawat Tavakhiorwachai, Pitsavee Manakul, Kanat Pijitwanakul
 SCB IOX, SCBX Group

Abstract

This paper presents FinCoT, a structured chain-of-thought (CoT) prompting framework that embeds domain-specific expert financial reasoning into prompts to guide large language models (LLMs) in financial tasks. We identify three main prompting styles in financial NLP (FinNLPs): (1) standard prompting (zero-shot), (2) structured CoT (with explicit structured reasoning steps), and (3) structured CoT (with explicit structured reasoning steps). Prior work has mostly focused on the first two, while structured CoT remains underexplored and lacks domain expertise incorporation. Therefore, we evaluate all three prompting approaches across ten CPA-style financial domains and introduce FinCoT as the

challenges in interpretability and domain alignment (Ota et al., 2024; Arora et al., 2024; Lee et al., 2025). Accordingly, these shortcomings motivate a stricter control over a model’s intermediate reasoning path, which we explore via prompt design. Prompting guides LLM reasoning without extra training. Methods such as Chain-of-Thought (Wei et al., 2023), Code Prompting (Hu et al., 2023), Plan-and-Solve (Wang et al., 2023), and Self-Reflection (Rosen and Green, 2023) encourage stepwise thinking but remain domain-agnostic. In finance, this can lead to omissions in validation checks or confusion between basis points and percentages. Yet real-world analysis involves well-defined workflows—valuation, discounting, portf-

Published as a workshop paper at Open Science for Foundation Models, ICLR 2025

[2] [cs.CL] 27 Mar 2025

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

Pitsavee Manakul¹, Gaungshi Sun², Warit Sirichotdomrong¹, Kasima Tharapititchai¹, Kanat Pijitwanakul¹

¹SCB IOX, SCBX Group, Thailand
²Department of Engineering, University of Cambridge, United Kingdom
 [pitsavee, warit, kasima, kanat]@scbx.com, gsj3@cam.ac.uk

Abstract

Audio language models process audio inputs using textual prompts for task speech recognition and audio captioning. Although built on multilingual pre-trained components, most are trained primarily on English, limiting their usability for other languages. This paper evaluates audio language models on Thai, a low-resource language, and finds that they lack emergent cross-lingual abilities despite their multilingual foundations. To address this, we explore data mixtures that optimize audio language models for both a target language and English while integrating audio comprehension and speech instruction-following into a unified model. Our experiments provide insights into improving instruction-following in low-resource languages by balancing language-specific and multilingual training data. The proposed model, Typhoon-Audio, significantly outperforms 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 Audio Model, The Language Model, Low-resource Language Training

a low-resource language using Thai. As a case study, we provide a recipe to enhance the low-resource language ability while maintaining the English performance. Second, we integrate pre-trained audio-understanding and speech instruction-following capabilities into one unified model.

2. Related Work

Audio Language Models. GAZONDO improves three primary components: an LLM based on Vicuna (Li, 2023) on a custom-built Thai Whisper-large-v2 (2) and BERT5 (3) for audio events. The representation from Whisper and BERT5 are concatenated and passed through an adapter (4) trained on the audio representing instructions in the input to the LLM. Trainable models are in (5) (Pre-train) and (6) (Lora) weights of the LLM. Training data consists of speech recognition, instruction, audio captioning, or spoken Q&A. Similarly, sun et al. (2024) uses Whisper-large-v2 (model), and LLM as based on Qwen (1). No instruction models (adapter) is employed. However, a variety of special tokens are incorporated. The pre-training data involves training the audio models while freezing the LLM.

[2] [cs.CL] 23 May 2025

Unlearning vs. Obfuscation: Are We Truly Removing Knowledge?

Gaungshi Sun¹, Pitsavee Manakul¹, Xiao Zhai², Mark Gale²

¹Department of Engineering, University of Cambridge
²Department of Informatics, King’s College London
 SCB IOX, SCBX Group
 [gsj34, xzj76]@cam.ac.uk, x.zhai@kcl.ac.uk, p.manakul@kcl.ac.uk

Abstract

Unlearning has emerged as a critical capability for large language models (LLMs) to support data privacy, regulatory compliance, and ethical AI deployment. Recent techniques rely on LLMs’ ability to obfuscate information by injecting incorrect or irrelevant information to suppress knowledge. Such methods effectively erode knowledge addition rather than true removal, often leaving models vulnerable to probing. In this paper, we formally distinguish unlearning from obfuscation and introduce a problem-based evaluation framework to assess whether existing approaches genuinely remove targeted information. Moreover, we propose EMP-UC, a novel unlearning method that fuses the model’s predictive distribution over automatically generated multiple-choice questions using RL-

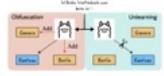


Figure 1: Illustration of obfuscation and unlearning. Obfuscation involves injecting incorrect or irrelevant information into the model’s knowledge base. Unlearning aims to remove targeted information from the model’s knowledge base. The diagram shows the flow from Obfuscation to Unlearning through the Model and Knowledge components.

ward signals based on the correctness of generated answers during training. Current studies have explored various components of the RFT pipeline, including objective functions and training algorithms (Liu et al., 2025; Yu et al., 2025; Yu et al., 2025; Yu et al., 2025). Data domains and curricula (Liu et al., 2025; Wei et al., 2025; Yu et al., 2025; Hu et al., 2025), reward functions and shaping (Yeo et al., 2025; Yu et al., 2025; Hu et al., 2025), and the influence of coherent behaviors across different LLM families and model sizes (Liu et al., 2025; Zeng et al., 2024; Guadix et al., 2025). However, despite these improvements for various components of the RFT pipeline, one critical aspect remains understudied: the design of the prompt.

To scope our study, we separate a prompt used

[2] [cs.CL] 10 Sep 2025

Published as a workshop paper at Open Science for Foundation Models, ICLR 2025

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

Kanat Pijitwanakul, Pittawat Tavakhiorwachai,
 Pitsavee Manakul, and Kasima Tharapititchai
 SCB IOX R&D
 SCBX Group
 Bangkok, Thailand
 [kanat, pittawat, pitsavee, kasima]@scbx.com

ABSTRACT

This paper introduces data selection and model merging methodologies aimed at incorporating advanced reasoning capabilities such as those of Deepseek R1 into language-specific large language models (LLMs), with a particular focus on the Thai LLM. Our goal is to enhance the reasoning capabilities of language-specific LLMs while maintaining their target language abilities. Deepseek R1 excels in reasoning but primarily benefits high-resource languages such as English and Chinese. However, low-resource languages remain underexplored due to the dominance of English-centric training data and model optimizations, which limit performance in these languages. This limitation results in unreliable code-switching language-specific effectiveness on tasks in low-resource languages. Meanwhile, local and regional LLM initiatives have attempted to bridge this gap by developing language-specific LLMs that focus on improving local linguistic fidelity. We demonstrate that, with only publicly available datasets and a computational budget of \$200, it is possible to enhance the reasoning capabilities of language-specific LLMs to match the level of Deepseek R1, without compromising their performance on tar-

TYPHOON T1: AN OPEN THAI REASONING MODEL

Pittawat Tavakhiorwachai, Pitsavee Manakul,
 Kasima Tharapititchai, and Kanat Pijitwanakul
 SCB IOX R&D
 SCBX Group
 [pittawat, pitsavee, kasima, kanat]@scbx.com

ABSTRACT

This paper introduces Typhoon T1, an open 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 performance on complex tasks. However, details on developing such models are limited, especially for reasoning models that can generate text in a low-resource language. Typhoon T1 presents an open effort that dives into the details of developing a reasoning model in a more cost-effective way by leveraging supervised fine-tuning using open datasets, instead of reinforcement learning. This paper shares the details about synthetic data generation and training, as well as an dataset and model weights. Additionally, we provide insights gained from developing a reasoning model that generalizes across domains and is capable of 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.

1 INTRODUCTION

[2] [cs.CL] 27 Mar 2025

Large Language Models Are Powerful!



ช่วยแต่งเรื่องเกี่ยวกับยักษ์ในวรรณคดีไทยให้หน่อย



แน่นอนเลยครับ

ในยุคสมัยที่ป่าไม้กว้างใหญ่ล้อมรอบหมู่บ้านเล็กๆ ริมฝั่งแม่น้ำแคว ชาวบ้านเล่าขานกันว่า ทุกคืนเดือนเพ็ญ มีเสียงกลองดังก้องกังวานจากที่สูงบนเขาผาแดง บางครั้งก็เห็นแสงสีทองสาดส่องลงมาที่พื้นดิน ราวกับใครสักคนยืนอยู่บนยอดเขา มองลงมาด้วยสายตาอันลึกซึ้ง ชาวบ้านกลัว เด็กๆ ต่างหลบเข้าไปในกระท่อม

...

And It's Even More Useful With a Good Prompt



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



ในยามราตรีอันเงียบสงัด ใต้แสงจันทร์แผดเผาผืนป่าลึกที่ปกคลุมด้วยหมอกบางๆ ขุนเขาสูงตระหง่านทางกรอบ ท้องฟ้าราวกับกำแพงทองคำ ณ ริมฝั่งลำธารใสสะอาด ผู้เฒ่าชื่อว่า พระยาพิชัย ซึ่งครั้งหนึ่งเคยเป็นนายพลผู้กล้าหาญ ได้นั่งพักอยู่ใต้ต้นศรีตรัง พร้อมกับถือคันเบ็ดไม้ไผ่ สายตาจ้องมองผิวน้ำที่สะท้อนดวงจันทร์ราวกับหยดน้ำค้างที่สั่นไหว ทันใดนั้น เสียงดังก้องกังวานดังมาจากเบื้องล่างของหุบเขา—เสียงเท้ายักษ์ที่กระทบพื้นดินจนดินสั่นไหว แล้วปรากฏร่างยักษ์ผู้มีขนาดใหญ่โตราวภูเขา ผิวสีเขียวคล้ำเหมือนใบไม้แห้ง ดวงตาเป็นประกายแดงเหมือนถ่านไฟร้อน แต่กลับมีแวตาทิศาคร่าโศก ราวกับมีภาระอันหนักอึ้งซ่อนอยู่ในหัวใจ

...

And It's Even More Useful With a Good Prompt



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



ในยามราตรีอันเงียบสงัด ใต้แสงจันทร์แผดเผาผืนป่าลึกที่ปกคลุมด้วยหมอกบางๆ ขุนเขาสูงตระหง่านทางกรอบ
 ท้องฟ้าราวกับกำแพงทองคำ ณ ริมฝั่งลำธารใสสะอาด ณ ช่องน้ำตื้นๆ บนผืนดินที่ขรุขระ มีคนเดินผ่านไปมา
 หาญ ได้นั่งพักอยู่ใต้ต้นศรีตรัง พร้อมกับถ้วยน้ำค้างที่สั้นไหว หันไฉนนั้น เสียงดังก้องกังวาน
 สั้นไหว แล้วปรากฏร่างยักษ์ผู้มีขนาดใหญ่โต
 เหมือนถ่านไฟร้อน แต่กลับมีแววตาที่เศร้าโศก ราวกับมีภาระอันหนักอึ้งซ่อนอยู่ในหัวใจ
 ...



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-existent examples

Reasoning



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



เรามาแก้โจทย์ที่ละขั้นตอนกันนะครับ

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

จากทฤษฎีบทของวิเอตา (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 จงวางแผนวิธีการแก้ปัญหาและดำเนินการตามนั้น



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
- ตรวจสอบคำตอบว่าสมเหตุสมผลหรือไม่

ดำเนินการตามแผน:

จาก Vieta's formulas:

$$a + b = 5 \quad \text{\texttt{(1)}}$$

$$ab = k \quad \text{\texttt{(2)}}$$

ใช้สูตร:

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

แทนค่าที่รู้...

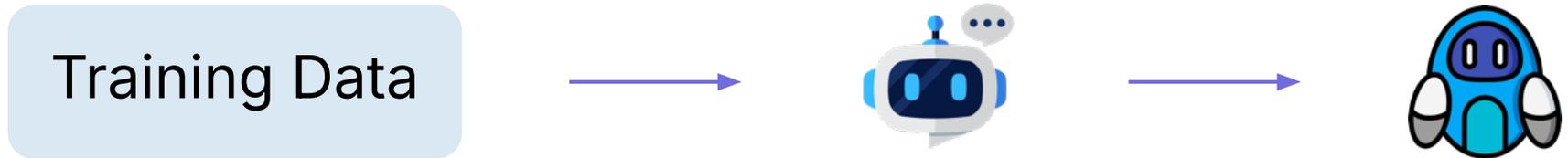
Inference



Inference



Training



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

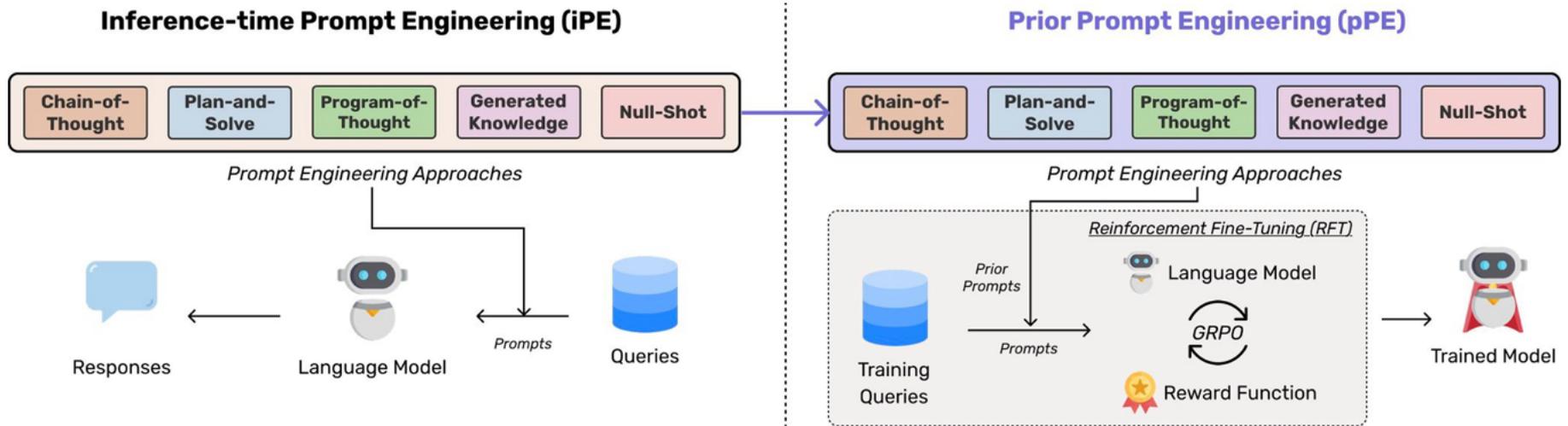
Prior Prompt

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

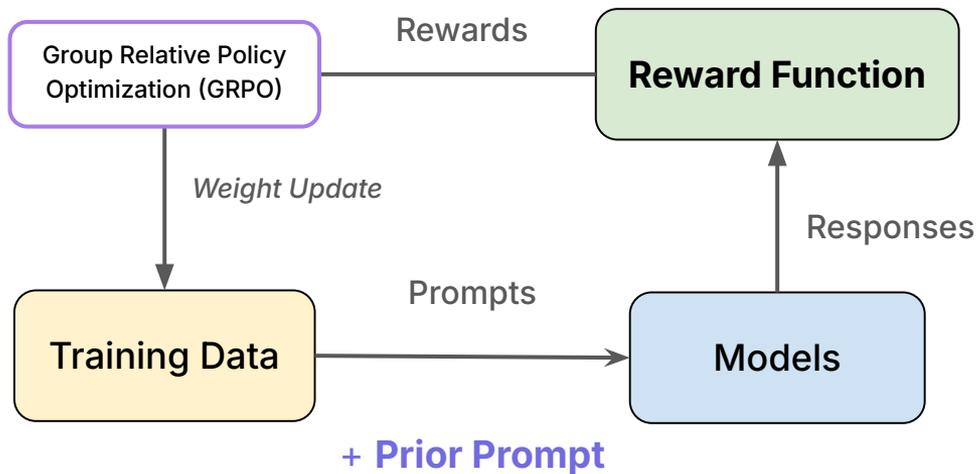
Reader

Recall knowledge before solving a task



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

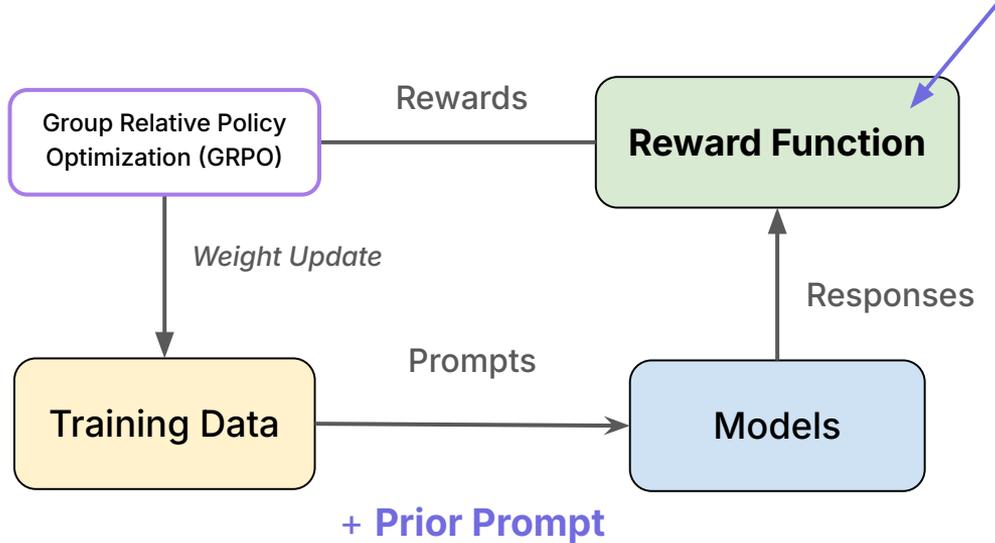
Reinforcement Fine-Tuning



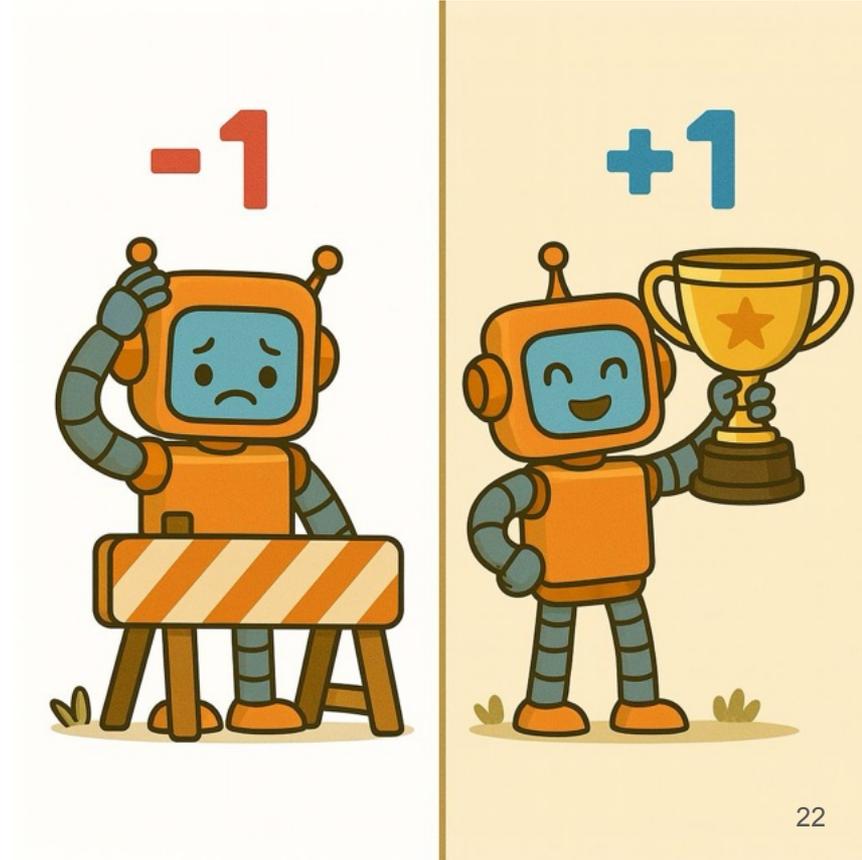
[...] The reasoning process and answer are enclosed within `<think>` `</think>` and `<answer>` `</answer>` tags, respectively, i.e., `<think>` reasoning process here `</think>` `<answer>` answer here `</answer>`. [...]



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)

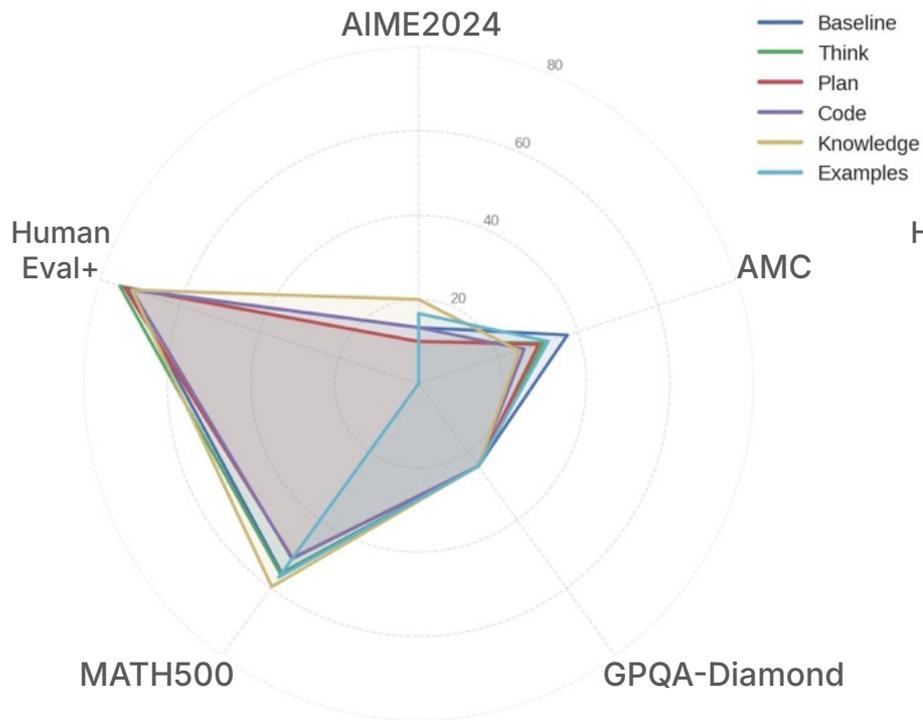


Metrics

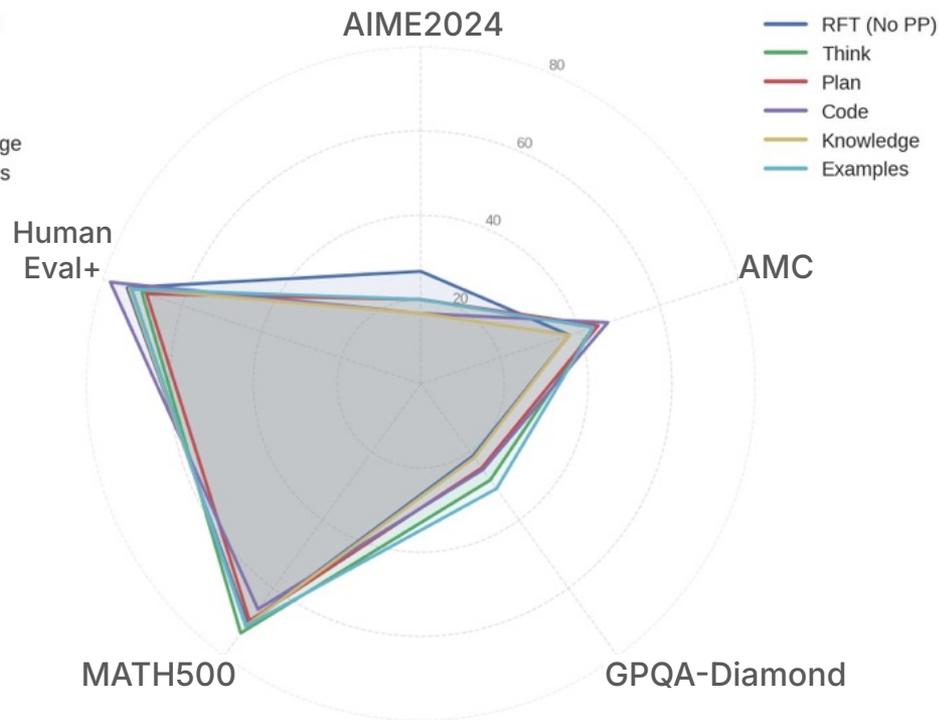
1.  Accuracy
2.  Average response length
3.  Ratio of fundamental cognitive behaviors
4.  Ratio of elicited behaviors

What Happens When We Bring Prompt Engineering Into Training?

Inference-time Prompt Engineering



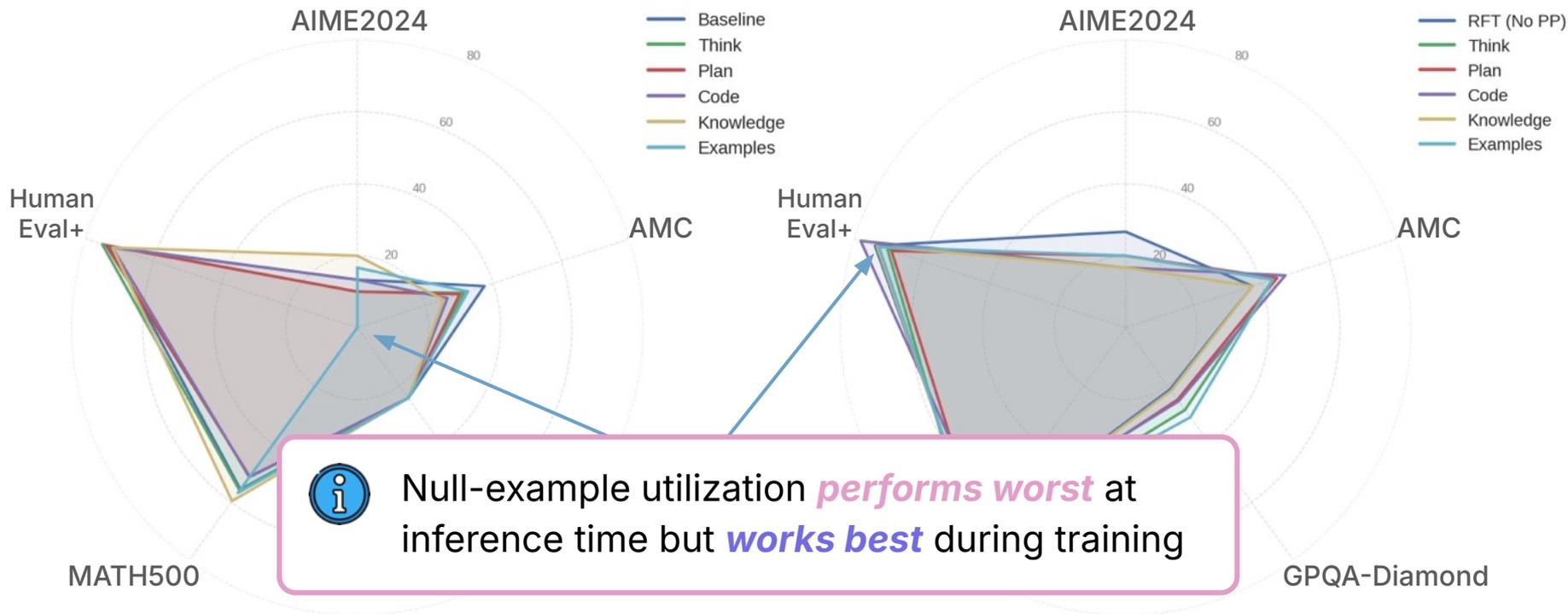
Prior Prompt Engineering



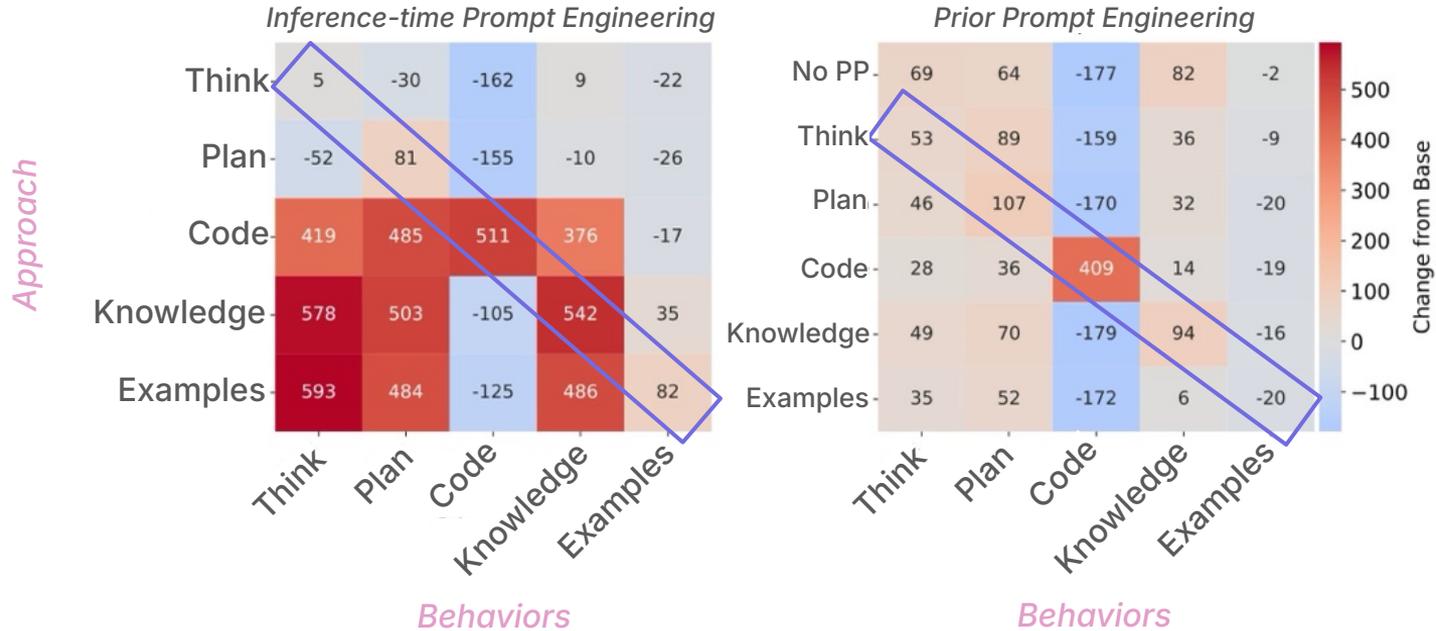
What Happens When We Bring Prompt Engineering Into Training?

Inference-time Prompt Engineering

Prior Prompt Engineering



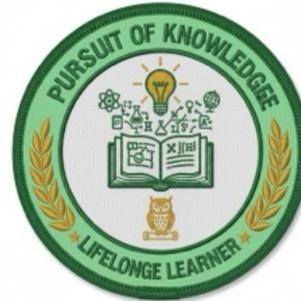
Most Post-RFT Models Are More Aligned in Generating Their Taught Behaviors



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

