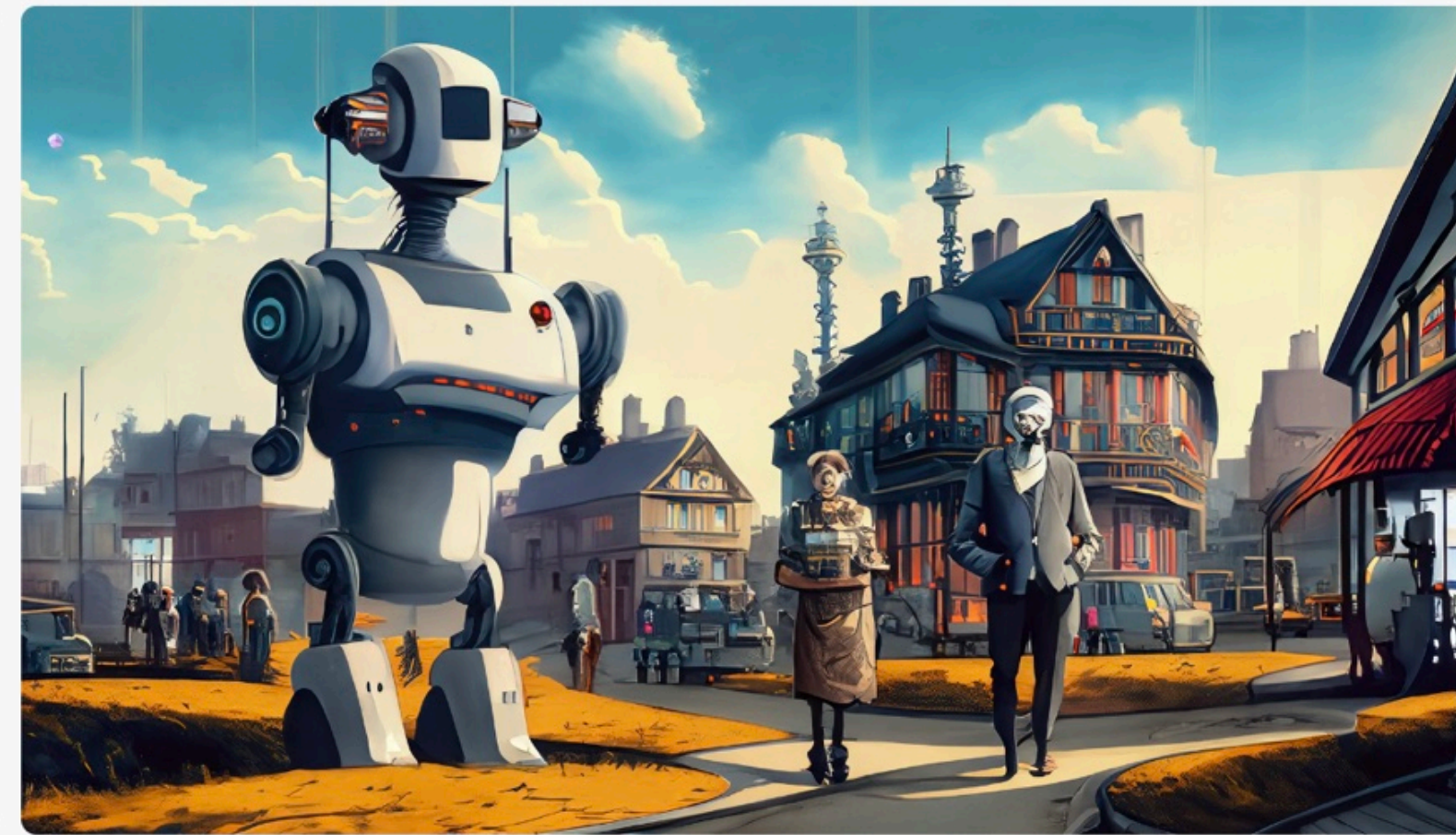


A Brief Introduction to

# Prompt Engineering

BY PETE - PITTAWAT TAVEEKITWORACHAI



A town with human and robots live together

Prompt

Clear style

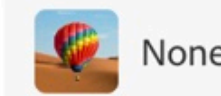
Art x

Refresh

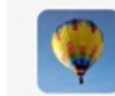
Aspect ratio

Widescreen (16:9)

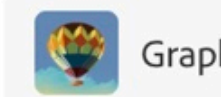
Content type



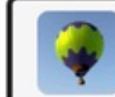
None



Photo



Graphic



Art

Styles

All

Popular

Movements

Themes

Techniques

Effects

Materials

Concepts

Popular



Digital art



Synthwave



Palette knife



Layered paper

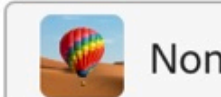


Neon



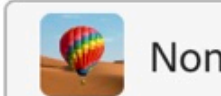
Chaotic

Color and tone



None

Lighting



None

---

# Agenda

- LLM
- Prompt and Prompt Engineering
- Future Outlook

---

# **(Large) Language Model**

---

---

# What is Language Model (LM)?

- A model that takes a words sequence as an input and predict a **next word** as an output (**text completion**)

- Example

Computer consists of two main parts, hardware and \_\_\_\_\_

---

# What is Language Model (LM)?

- A model that takes a words sequence as an input and predict a **next word** as an output (**text completion**)
- Example

Computer consists of two main parts, hardware and **software.**

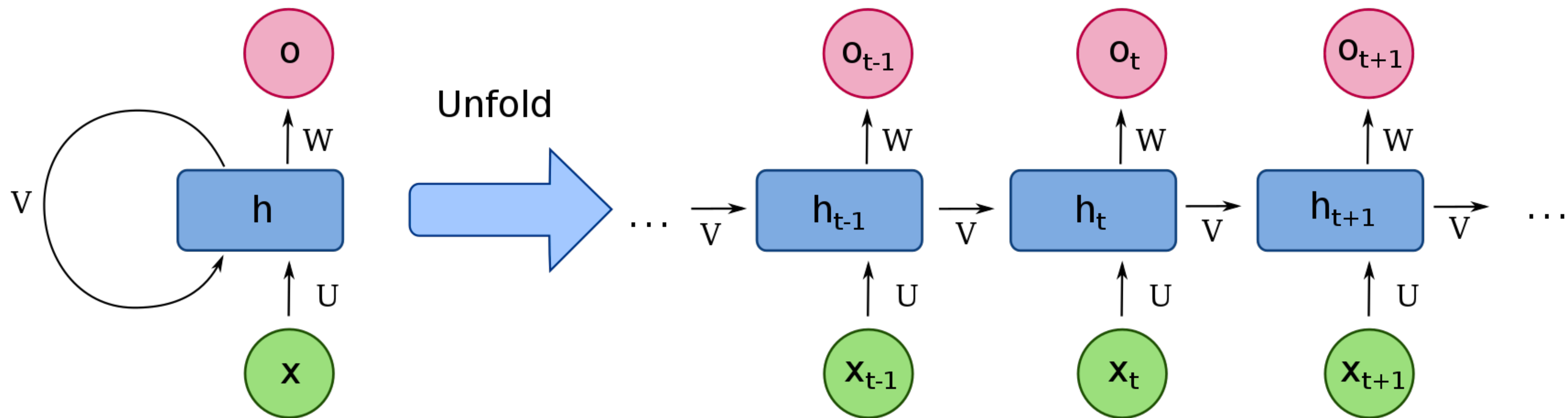
---

# Why LM?

- Pre-text task (self-supervised)
- Fine-tune it for downstream tasks, e.g., sentiment analysis, toxicity classification, summarization

# How to Train Your Dragon 🐉 LM?

- In the past, we use **Recurrent Neural Network (RNN)**





---

# How to Train Your Dragon LM?

- In the past, we use **Recurrent Neural Network (RNN)**
- One huge drawback: It is **sequential!**

---

# The Rise of Transformers

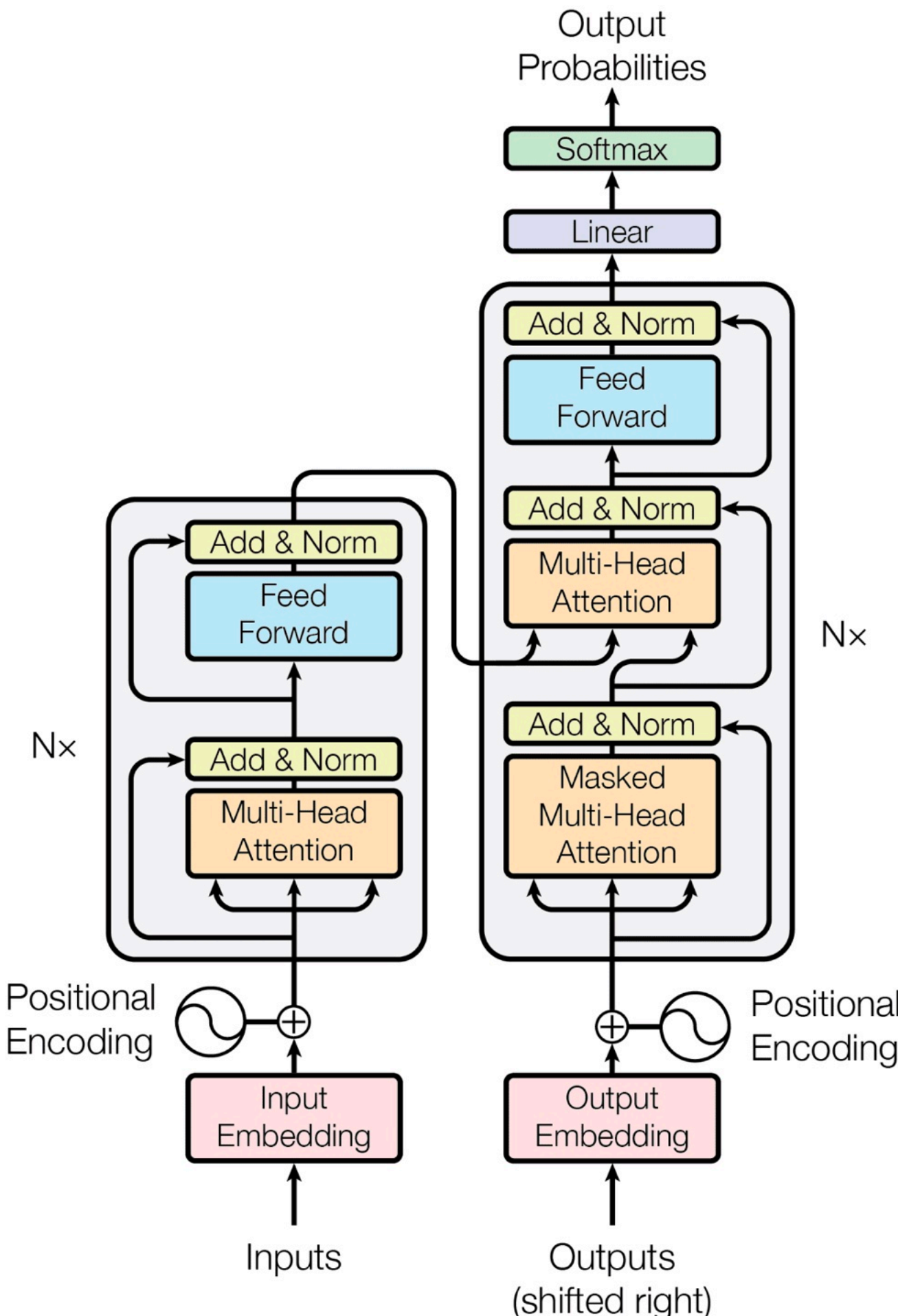


---

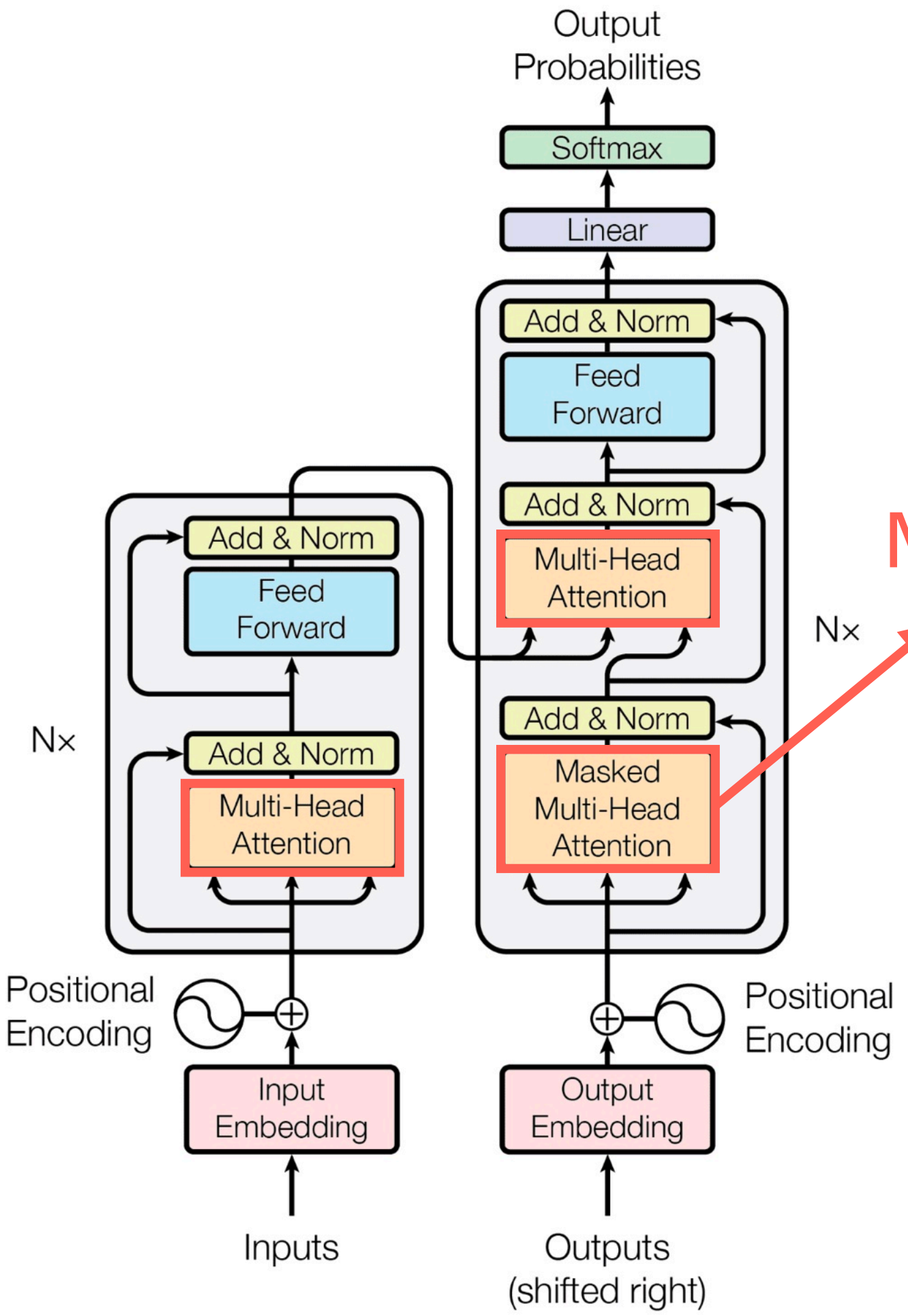
# The Rise of Transformers



# The Rise of Transformers



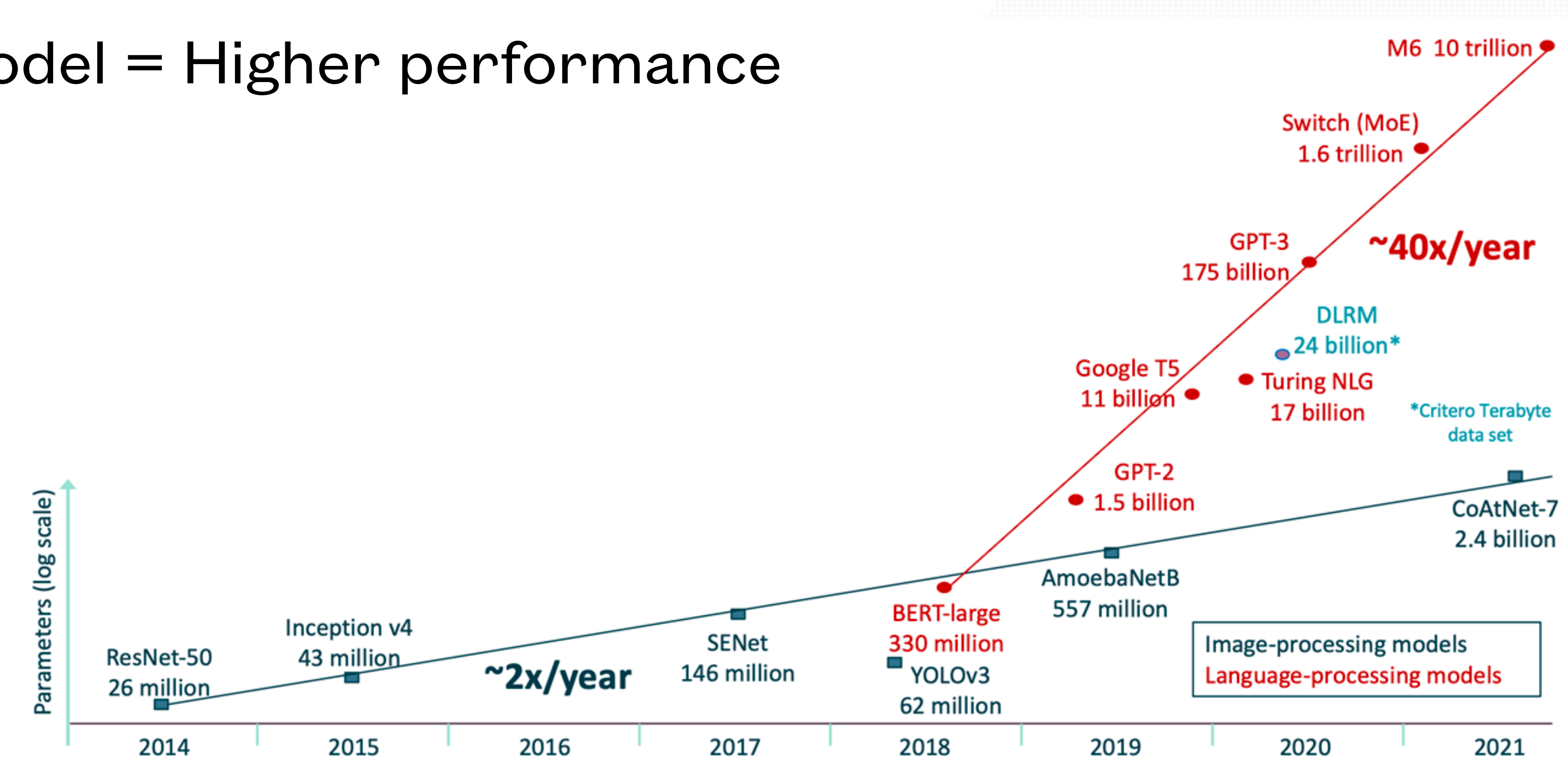
# The Rise of Transformers



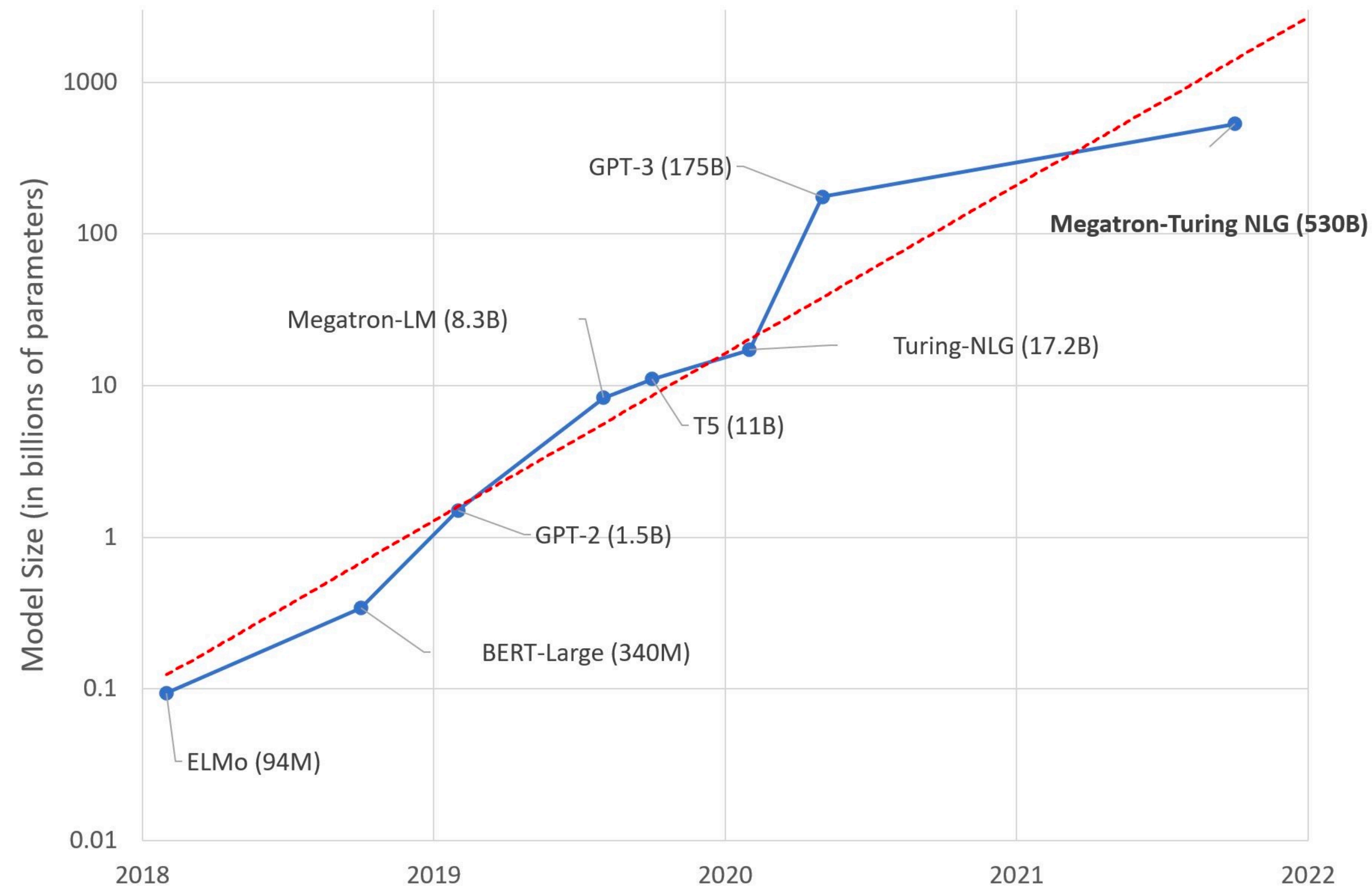
Multi-head Self-attention (MSA) enables **parallel computing**

# Why We Need Large LM (LLM)?

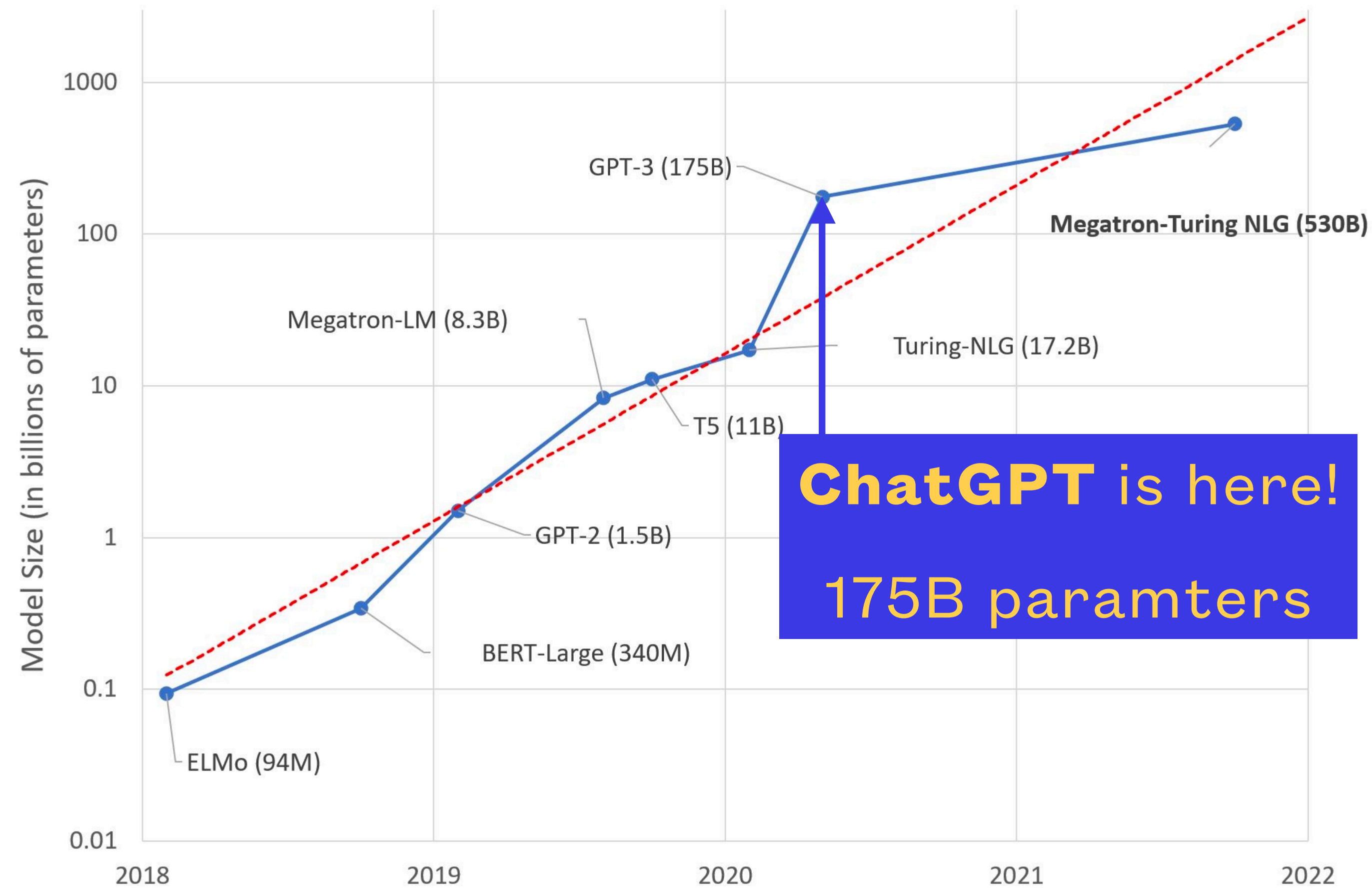
- Larger model = Higher performance



# Why We Need Large LM (LLM)?



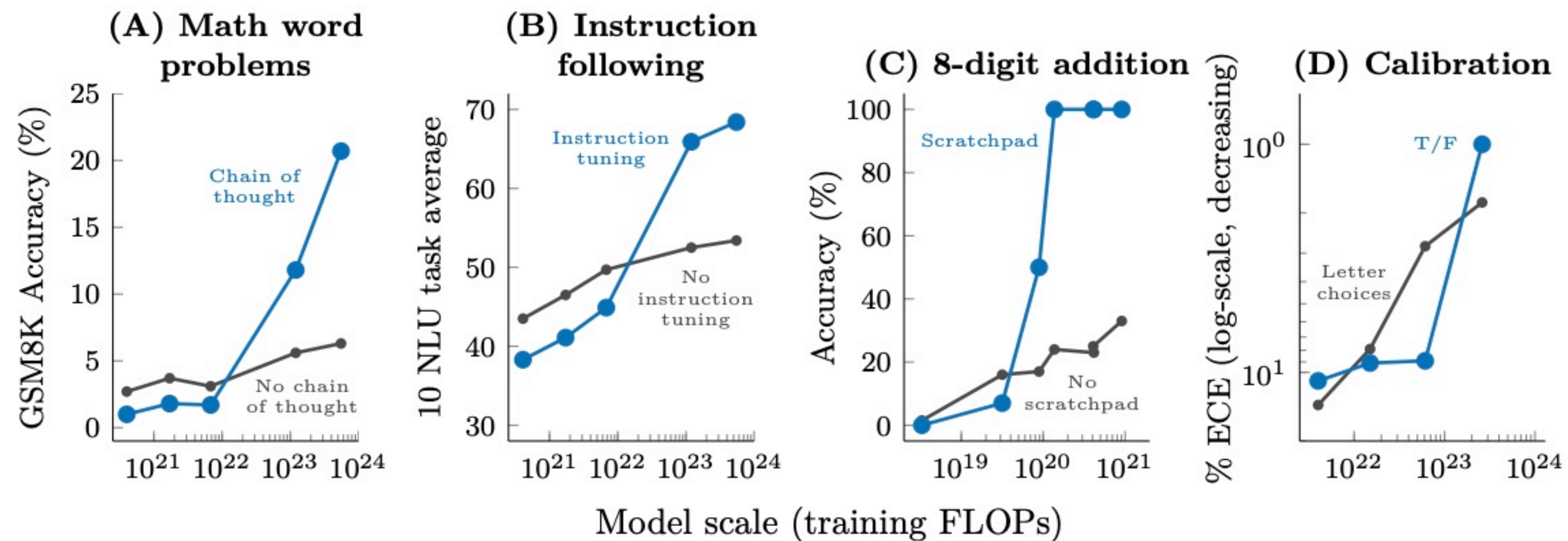
# Why We Need Large LM (LLM)?





# Emergent Abilities: LLM Exclusivity

- When the model get bigger to a certain point, it gains **emergent abilities**
- Can perform **untrained** tasks with high performance



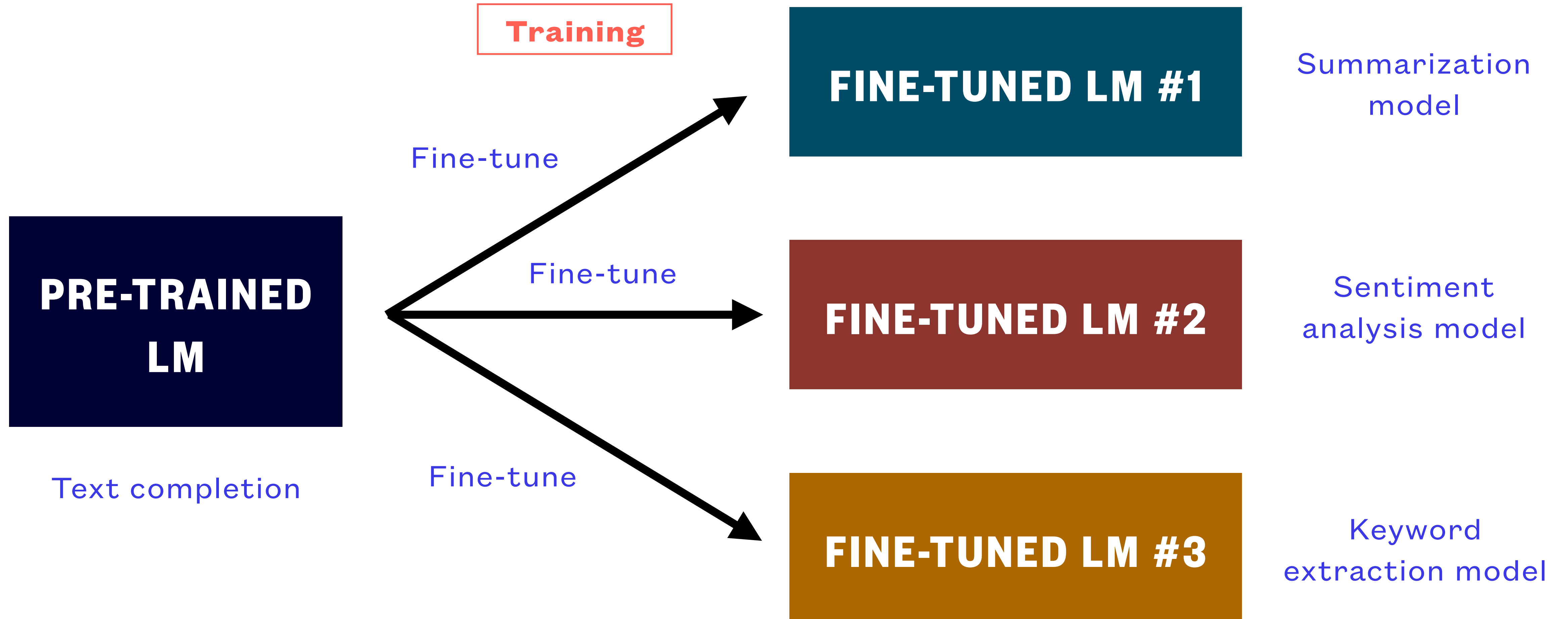
---

# LM vs LLM

**PRE-TRAINED  
LM**

Text completion

# LM vs LLM



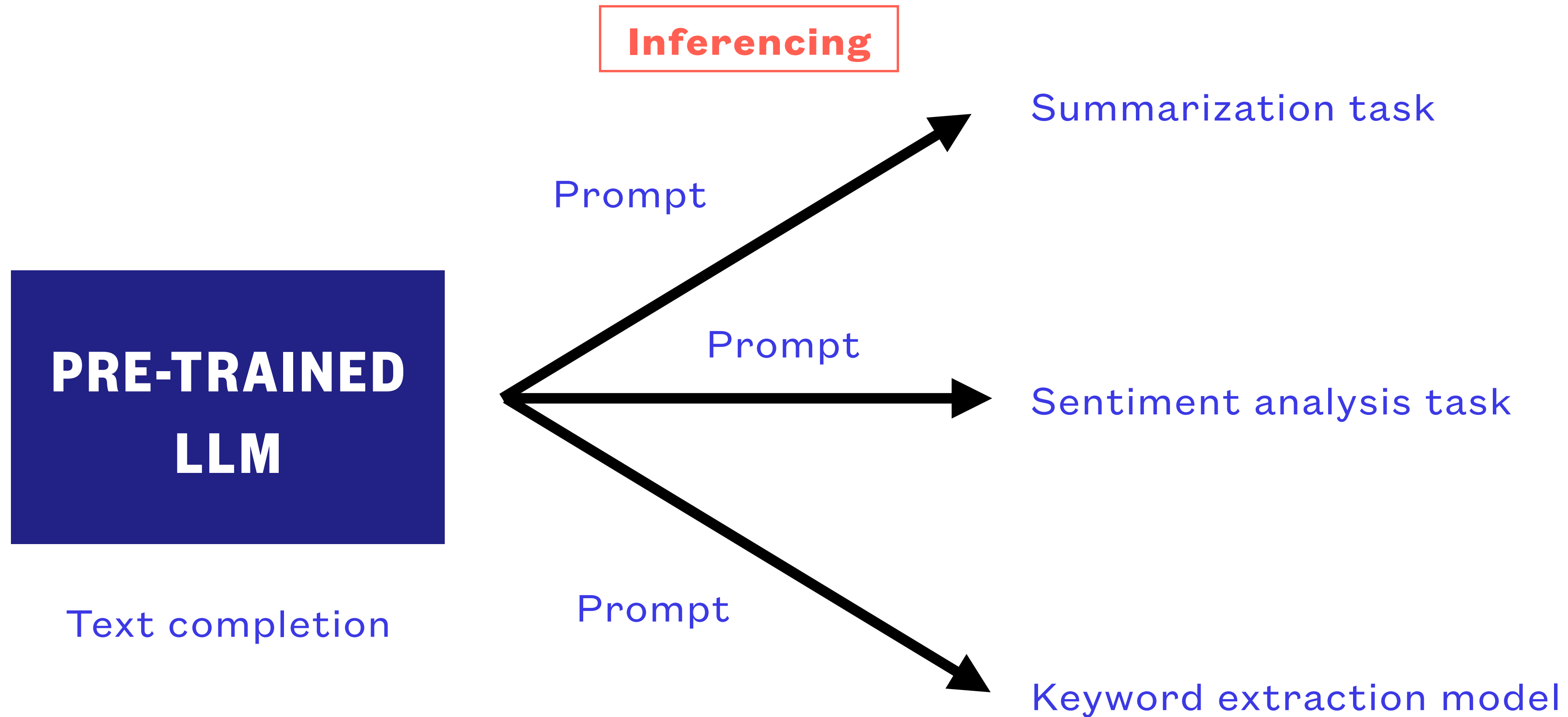
---

# LM vs LLM

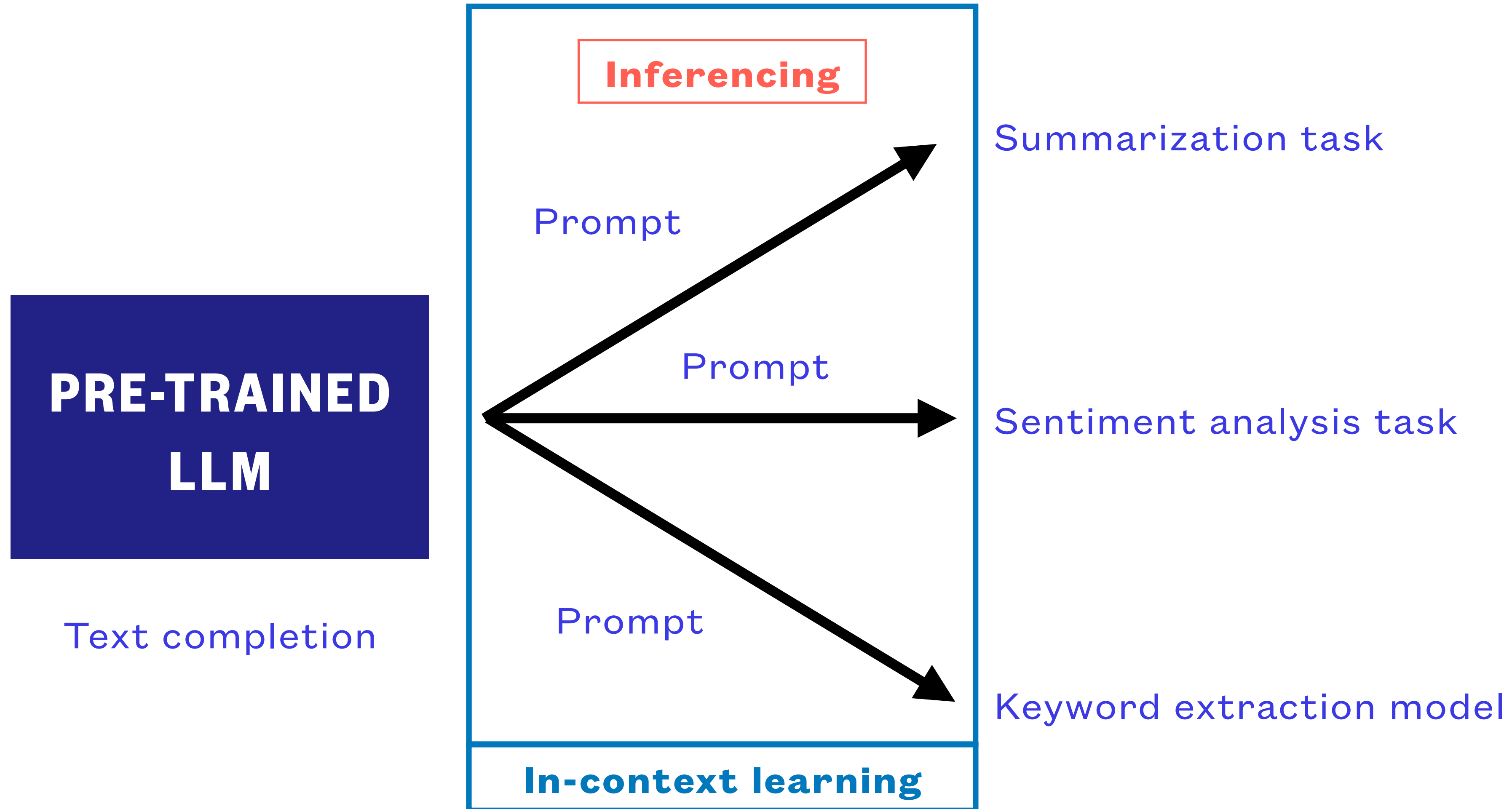
**PRE-TRAINED  
LLM**

Text completion

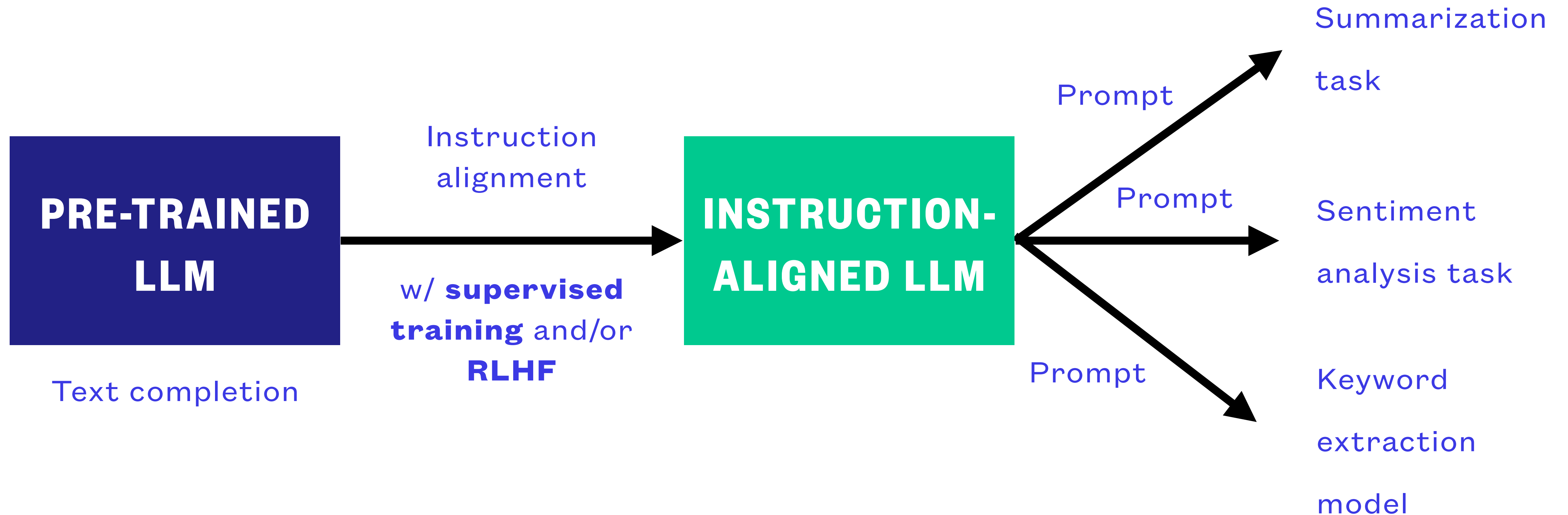
# LM vs LLM



# LM vs LLM



# Instruction-tuned LLM



---

# Prompt (Engineering)

---



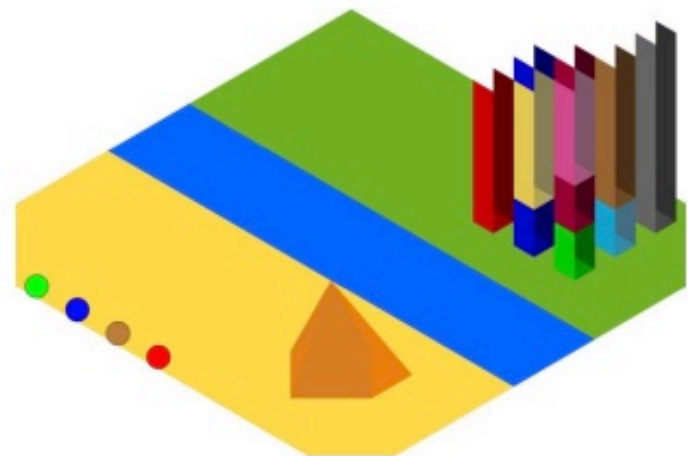
# Prompt

- An **input** given to a model
- A **condition** given to a model

## GPT-4 and Stable Diffusion

**Prompt:** A screenshot of a city-building game in 3D. The screenshot is showing a terrain where there is a river from left to right, there is a desert with a pyramid below the river, and a city with many highrises above the river. The bottom of the screen has 4 buttons with the color green, blue, brown, and red respectively.

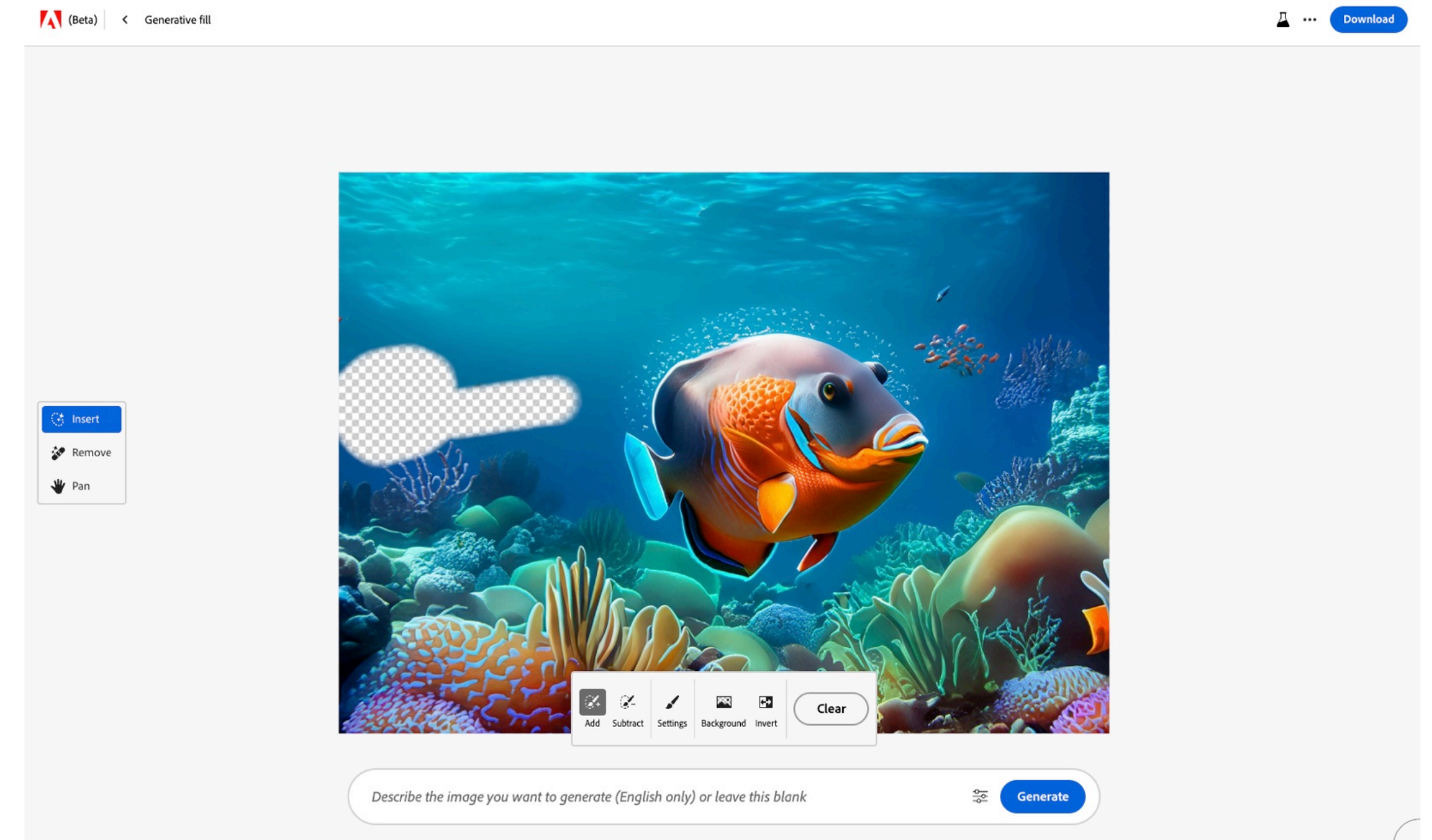
GPT-4:



Stable Diffusion v2.1:



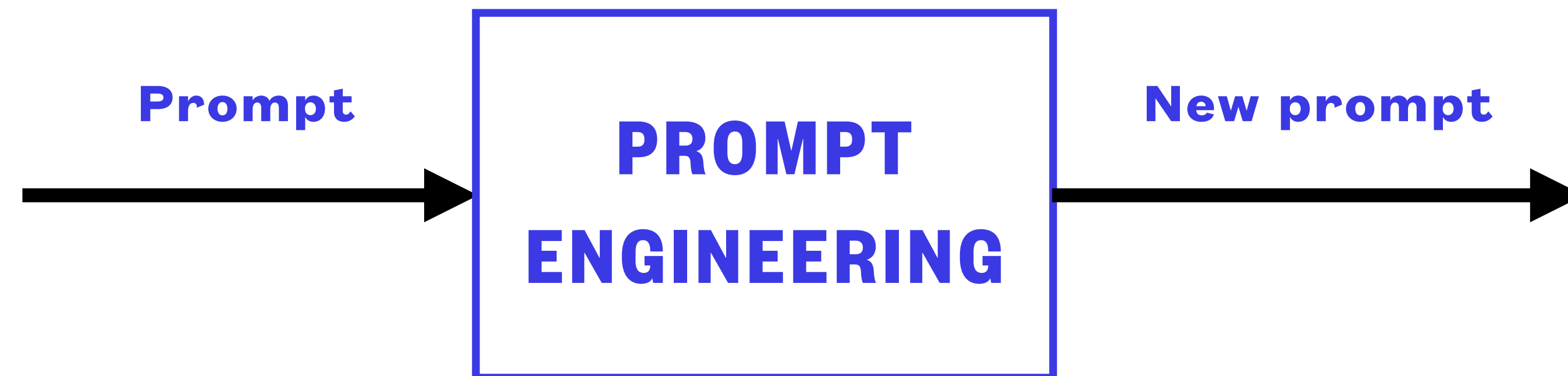
Stable Diffusion v2.1  
with GPT-4's sketch:



---

# Prompt Engineering (PE)

- A process of **composing/improving** a prompt given to a model to get desired outcomes



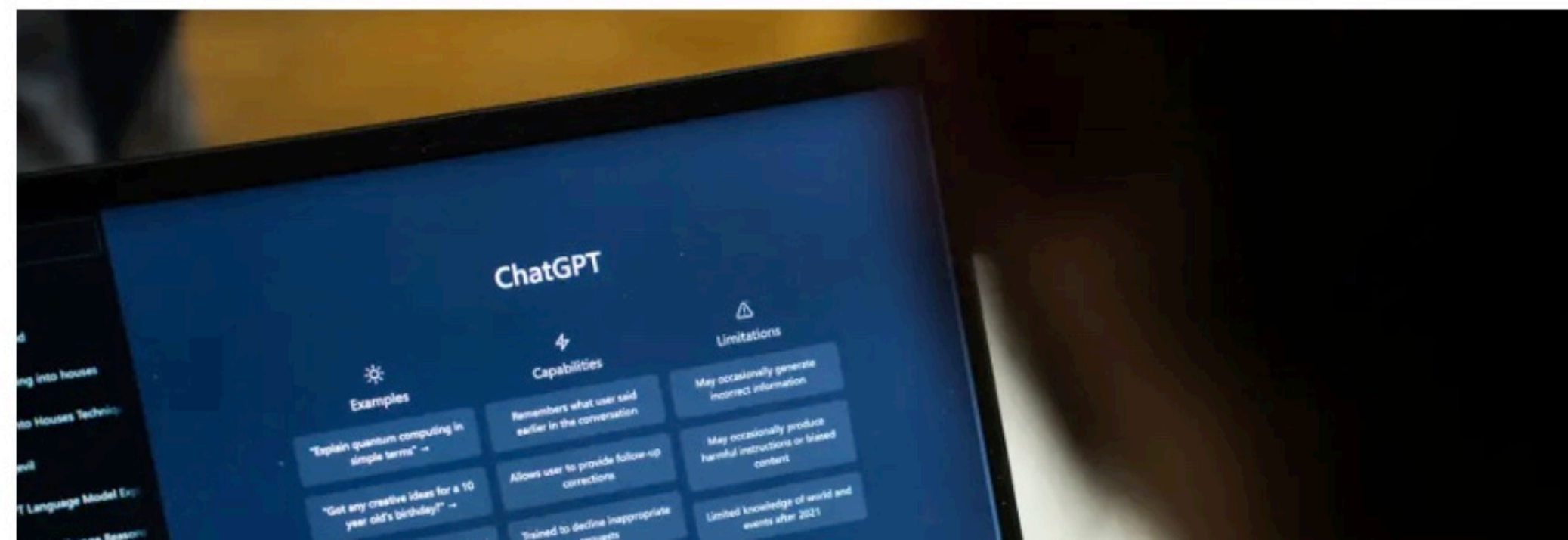
# PE is a New Google Skills

- Late 2000s: How to make a search query for Google to get desired results
- **Now:** How to make a prompt for a model to get desired results

HOME > TECH

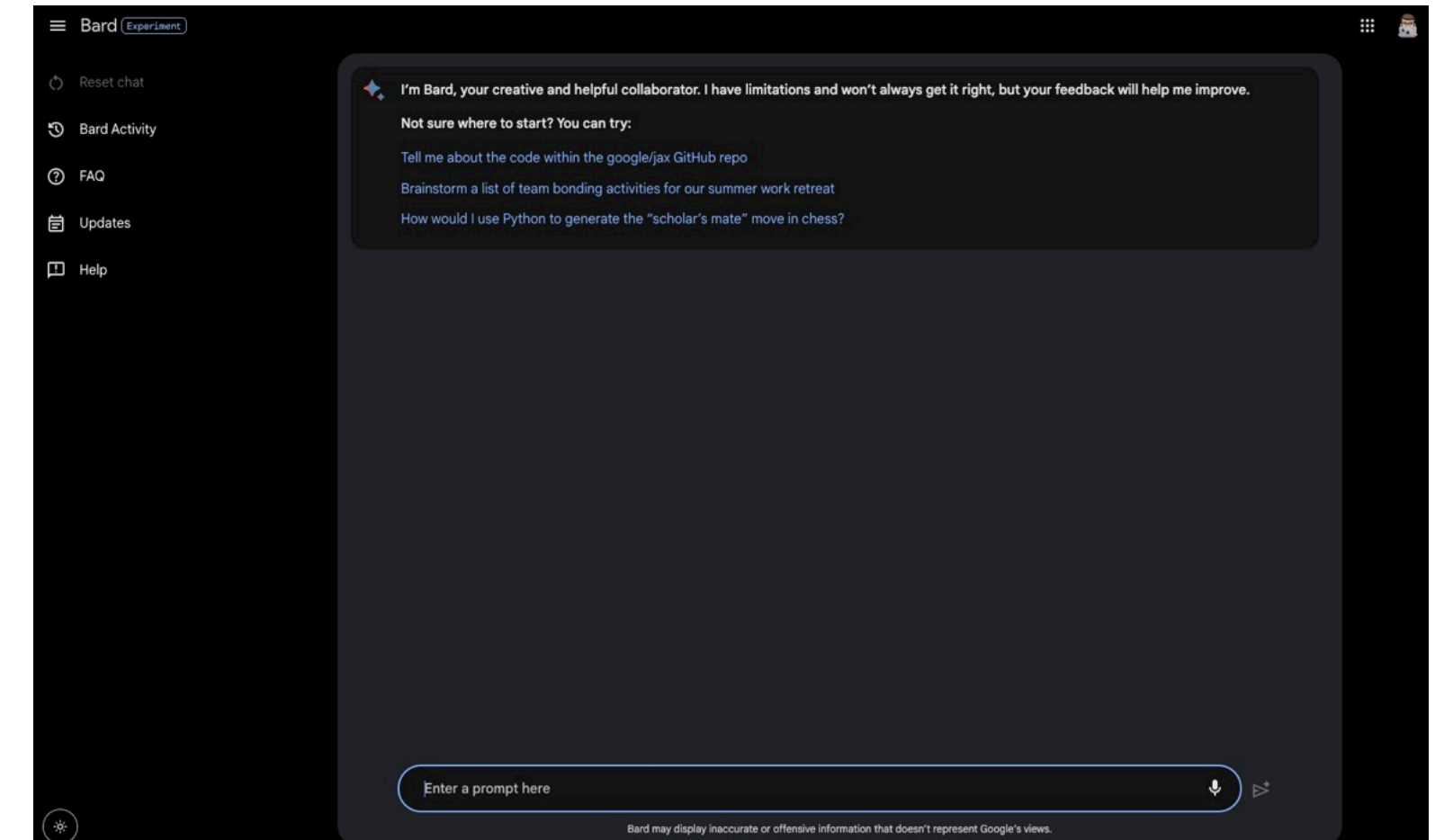
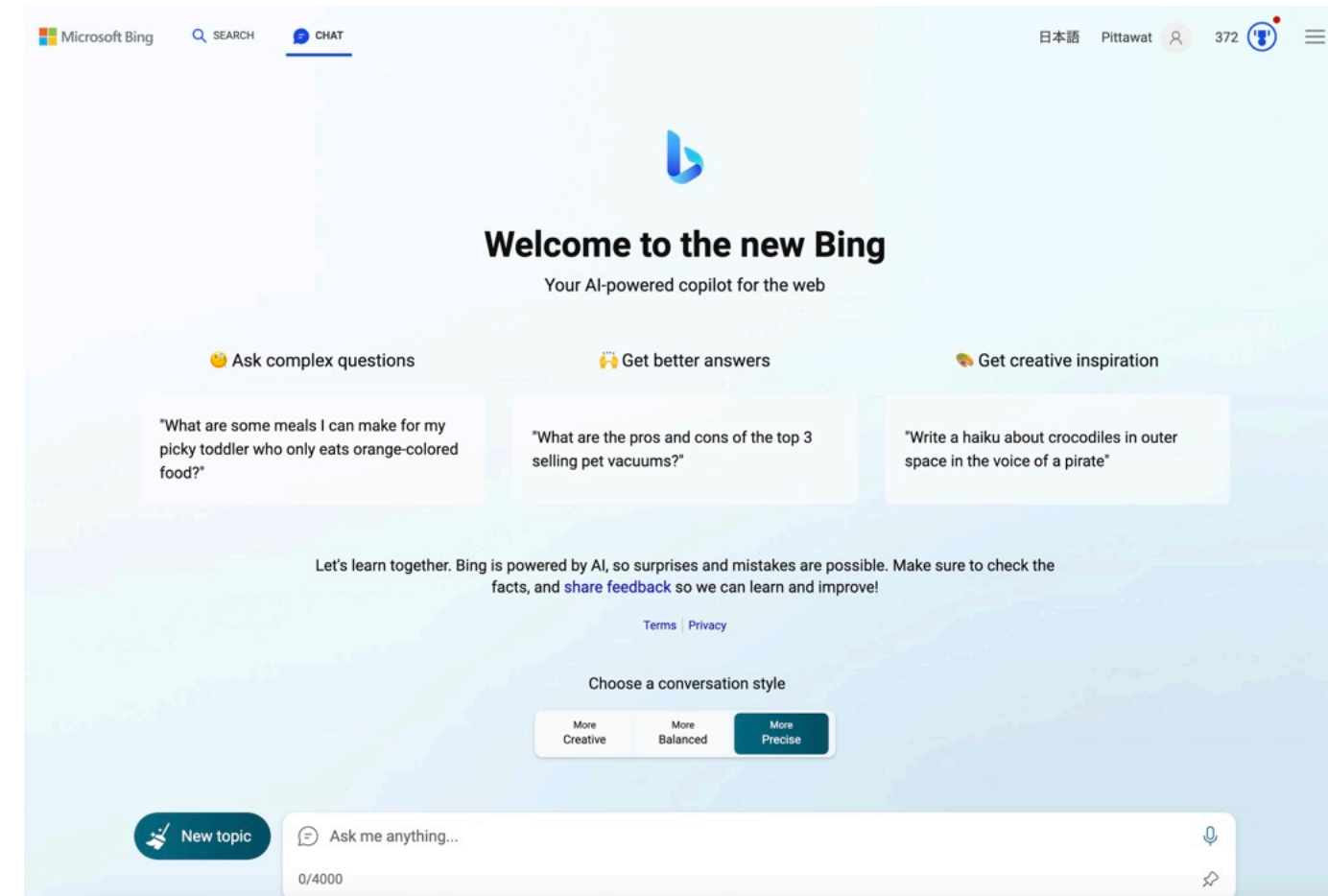
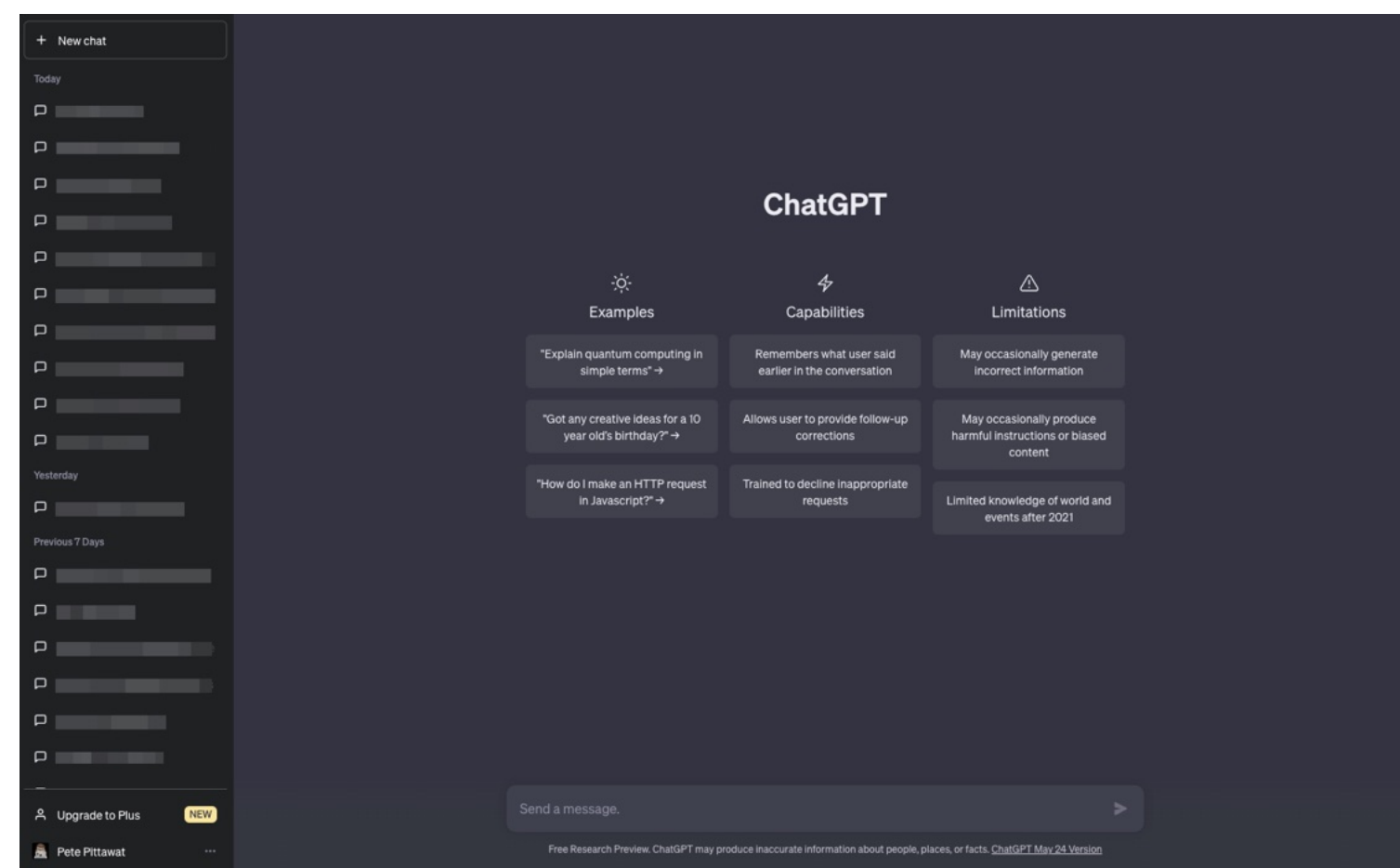
## AI 'prompt engineer' jobs can pay up to \$375,000 a year and don't always require a background in tech

Britney Nguyen May 2, 2023, 12:34 AM GMT+9



# PE for Instruction-aligned LLMs

- Focus on crafting **text** for chat-based LLMs, for example, ChatGPT, Bing Chat, Google Bard



---

"The simplest explanation is usually the best one."

**Occam's razor**

---

---

# General Prompt Structure

- Instruction/Question
- Output indicator
- Context
- (Optional) Input
- (Optional) Example

---

# Instruction/Question

- Instruction
  - Start with **verb**: think of it like a teacher designing a question for a student
  - Be **concise**: think of it like explaining to a child
  - Be **explicit**: what do you want and don't want
- Question
  - Start with **5W1H**
  - End with **question mark (?)** may help

---

# Output Indicator

- Be clear about what do you **want** in which **format**
  - Modifier: formal, modern, funny, etc.
  - Constraint: one sentence, within code block (`` ` ``), etc.
  - Format: JSON, XML, email, a list of 10 keywords, etc.
- Output **example** might be useful



---

# Context

- The most **important** part
- Explain it in a sense that LLMs know **nothing** about a topic
  - Explain all the **constraints**, what **possible** and not possible, and **relevant** information, etc.
- **Role** prompting: Make LLMs become something/someone
  - Ex. teacher, Steve Jobs, Enigma, computer science professor
  - **Expert** role prompting: Ask the LLM to assume the role of an expert in order to obtain more **factual** and **accurate** results.

---

# General Tips

- **Experiment** a lot!
- Some time it is useful to ask it in **code** (output/input)
- **Multi-turn** conversation might be useful
- Use uncommon **delimiters** to signify important information or separate sections, e.g., #####

---

# Example 1: Information Retrieval

<https://chat.openai.com/share/9e03fcdd-8952-48fc-887a-2563d6ea51f1>

---

---

# User Message 1

**Model:** ChatGPT

**Prompt**

What is a language model?

**Response**

A language model is a type of artificial intelligence (AI) model designed to understand and generate human language. It uses statistical techniques and pattern recognition to learn the structure, grammar, and semantics of natural language. Language models are trained on large amounts of text data and are capable of predicting the probability of a given sequence of words or generating coherent and contextually relevant text.

Language models are widely used in various natural language processing (NLP) tasks, such as machine translation, text summarization, sentiment analysis, question answering, and chatbot systems. They provide a way to process and understand human language in a computationally efficient manner.

With the advent of deep learning and neural networks, more advanced language models have been developed. These models, such as OpenAI's GPT (Generative Pre-trained Transformer), are capable of capturing complex dependencies and long-range context in text, resulting in more accurate and contextually aware language generation.

Language models like GPT-3.5, upon which I am based, have been trained on a vast amount of diverse text from the internet and other sources, allowing them to generate human-like responses and provide useful information on a wide range of topics.

---

# User Message 2

**Model:** ChatGPT

## **Prompt**

Summarize it

## **Response**

A language model is an AI system that understands and generates human language. It learns from large amounts of text data and can predict word sequences or generate coherent text. Language models are used in various applications like translation, summarization, and chatbots. Advanced models, such as GPT-3.5, capture complex language patterns and context. They provide accurate responses and information on diverse topics.

---

# User Message 3

**Model:** ChatGPT

## **Prompt**

Explain it like I'm five years old

## **Response**

Sure! Imagine a language model as a very smart robot that can talk and understand the things we say. It has learned a lot from reading many books and listening to people talking. This robot can help us by answering our questions and even having conversations with us. It's like having a really clever friend who knows a lot about many different things. The more the robot learns, the better it becomes at understanding and talking like a real person.

---

# Example 2: Information Extraction

<https://chat.openai.com/share/06dc11b6-ee34-4d41-9fca-3b1dacad07c1>

---

---

# User Message 1

**Model:** ChatGPT

**Prompt**

What is a company being mentioned in this content?

As Apple prepares its long-rumored jump into augmented reality on Monday, doubts have shadowed every step of the way. There are reports of frequent changes in direction and skepticism inside Apple's ranks. The device has allegedly been hard to manufacture and required numerous compromises. The process has taken years longer than Apple expected. And at a rumored \$3,000, even Apple reportedly expects slow short-term sales.

But among AR professionals, the mood is jubilant. "This is the single greatest thing that could happen to this industry," says Jay Wright, CEO of VR / AR collaboration platform Campfire 3D. "Whether you make hardware or software. We're excited about it."

**Response**

The company being mentioned in the given content is Apple.



---

# User Message 1

**Model:** ChatGPT

**Prompt**

What is a company being mentioned in this content?

**Question**

**Input**

As Apple prepares its long-rumored jump into augmented reality on Monday, doubts have shadowed every step of the way. There are reports of frequent changes in direction and skepticism inside Apple's ranks. The device has allegedly been hard to manufacture and required numerous compromises. The process has taken years longer than Apple expected. And at a rumored \$3,000, even Apple reportedly expects slow short-term sales.

But among AR professionals, the mood is jubilant. "This is the single greatest thing that could happen to this industry," says Jay Wright, CEO of VR / AR collaboration platform Campfire 3D. "Whether you make hardware or software. We're excited about it."

**Response**

The company being mentioned in the given content is Apple.

---

# User Message 2

**Model:** ChatGPT

## **Prompt**

What is the main topic in of this excerpt?

## **Response**

The main topic of this excerpt is Apple's rumored entry into the augmented reality (AR) market.

---

# Example 3: Code Generation

<https://chat.openai.com/share/06dc11b6-ee34-4d41-9fca-3b1dacad07c1>

---

---

# User Message 1

**Model:** ChatGPT

**Prompt**

Write a TypeScript code to validate format of email string. Output only code without explanation.

**Response**

Sure! Here's the TypeScript code to validate the format of an email string:

```
...  
  
function validateEmail(email: string): boolean {  
  const emailRegex = /^[A-Za-z0-9._%+-]+@[A-Za-z0-9.-]+\.[A-Za-z]{2,}$/;  
  return emailRegex.test(email);  
}  
...
```

Please note that this code uses a regular expression to validate the email format. The regular expression `/^[A-Za-z0-9._%+-]+@[A-Za-z0-9.-]+\.[A-Za-z]{2,}$/` ensures that the email string contains alphanumeric characters, dots, underscores, percent signs, plus signs, and hyphens in the local part before the "@" symbol. The domain part should contain alphanumeric characters, dots, and hyphens, followed by a dot and at least two characters for the top-level domain.

The `validateEmail` function takes an email string as an argument and returns a boolean value indicating whether the email format is valid.

---

# User Message 1

**Model:** ChatGPT

**Instruction**

**Output**

**Prompt**

Write a TypeScript code to validate format of email string. Output only code without explanation.

**Response**

Sure! Here's the TypeScript code to validate the format of an email string:

```
...  
function validateEmail(email: string): boolean {  
  const emailRegex = /^[A-Za-z0-9._%+-]+@[A-Za-z0-9.-]+\.[A-Za-z]{2,}$/;  
  return emailRegex.test(email);  
}  
...
```

Please note that this code uses a regular expression to validate the email format. The regular expression `/^[A-Za-z0-9._%+-]+@[A-Za-z0-9.-]+\.[A-Za-z]{2,}$/` ensures that the email string contains alphanumeric characters, dots, underscores, percent signs, plus signs, and hyphens in the local part before the "@" symbol. The domain part should contain alphanumeric characters, dots, and hyphens, followed by a dot and at least two characters for the top-level domain.

The `validateEmail` function takes an email string as an argument and returns a boolean value indicating whether the email format is valid.

---

# Prompt Engineering (PE)

---

---

# PE Techniques: A Summary

- **Basics**

- Zero-shot prompting
- One-shot prompting
- Few-shot prompting

- **Chain-of-Thought**

- Chain-of-Thought (CoT) prompting
- Zero-shot CoT prompting
- Self-consistency
- Tree of Thoughts (ToT) prompting

- Active-Prompt

- Multi-modal CoT

- Least-to-Most (LtM) prompting

- **Others**

- Generated Knowledge prompting

- Automatic Prompt Engineer (APE)

- Directional Stimulus prompting

- ReAct

---

# PE: Basics

---



---

# Zero-shot Prompting

- Instruct a LLM to perform a task that it has **not been seen** in training set

## Prompt

Classify the text into neutral, negative or positive.

Text: I think the vacation is okay.

Sentiment:

## Response

The sentiment of the text "I think the vacation is okay" is neutral.

---

# One-shot Prompting

- Instruct a LLM to perform a task that it has **not been seen** in training set by providing **one example**

## Prompt

Classify the text into neutral, negative or positive.

Text: I think the vacation is okay.

Sentiment: Neutral

Text: I love this hotel!

Sentiment:

## Response

Positive

---

# One-shot Prompting

- Instruct a LLM to perform a task that it has **not been seen** in training set by providing **one example**

## Prompt

Classify the text into neutral, negative or positive.

Text: I think the vacation is okay.

Sentiment: Neutral

**Example**

Text: I love this hotel!

Sentiment:

## Response

Positive

---

# Few-shot Prompting

- Instruct a LLM to perform a task that it has **not been seen** in training set by providing **a few examples**

## Prompt

Classify the text into neutral, negative or positive.

Text: I think the vacation is okay.

Sentiment: Neutral

Text: I love this hotel!

Sentiment: Positive

Text: I don't like this food.

Sentiment:

## Response

Negative

---

# Few-shot Prompting

- Instruct a LLM to perform a task that it has **not been seen** in training set by providing **a few examples**

## Prompt

Classify the text into neutral, negative or positive.

Text: I think the vacation is okay.

Sentiment: Neutral

Text: I love this hotel!

Sentiment: Positive

Text: I don't like this food.

Sentiment:

**Examples**

## Response

Negative

---

# PE: Chain-of-Thoughts

---

# Chain-of-Thought (CoT) Prompting

- Improve performance through **reasoning steps**

Standard Prompting	Chain-of-Thought Prompting
<p><b>Model Input</b></p> <p>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?</p> <p>A: The answer is 11.</p> <p>Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?</p>	<p><b>Model Input</b></p> <p>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?</p> <p>A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. <math>5 + 6 = 11</math>. The answer is 11.</p> <p>Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?</p>
<p><b>Model Output</b></p> <p>A: The answer is 27. ❌</p>	<p><b>Model Output</b></p> <p>A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had <math>23 - 20 = 3</math>. They bought 6 more apples, so they have <math>3 + 6 = 9</math>. The answer is 9. ✅</p>

# Zero-shot CoT Prompting

- Improve performance through reasoning steps **without** providing any examples

## (d) Zero-shot-CoT (Ours)

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?

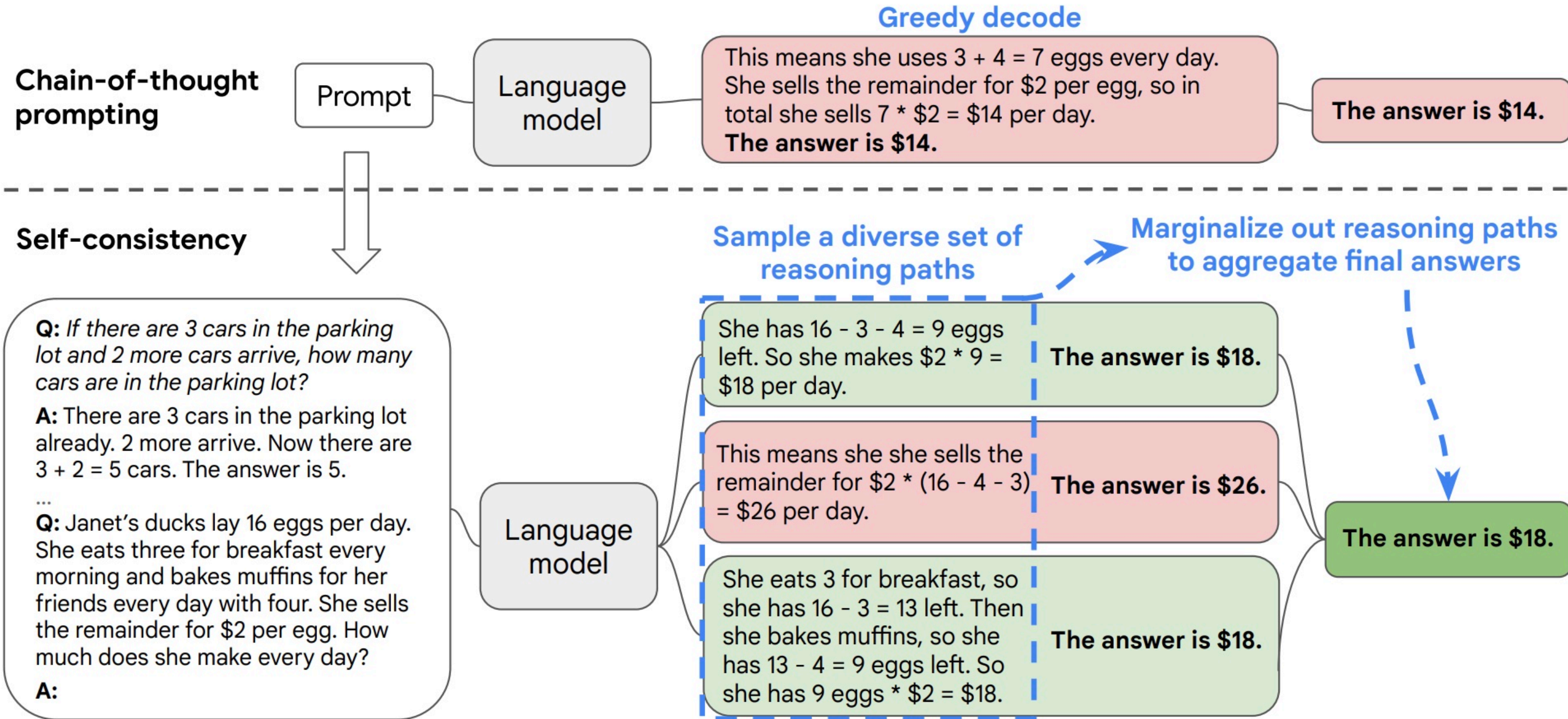
A: **Let's think step by step.**

*(Output) There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. ✓*



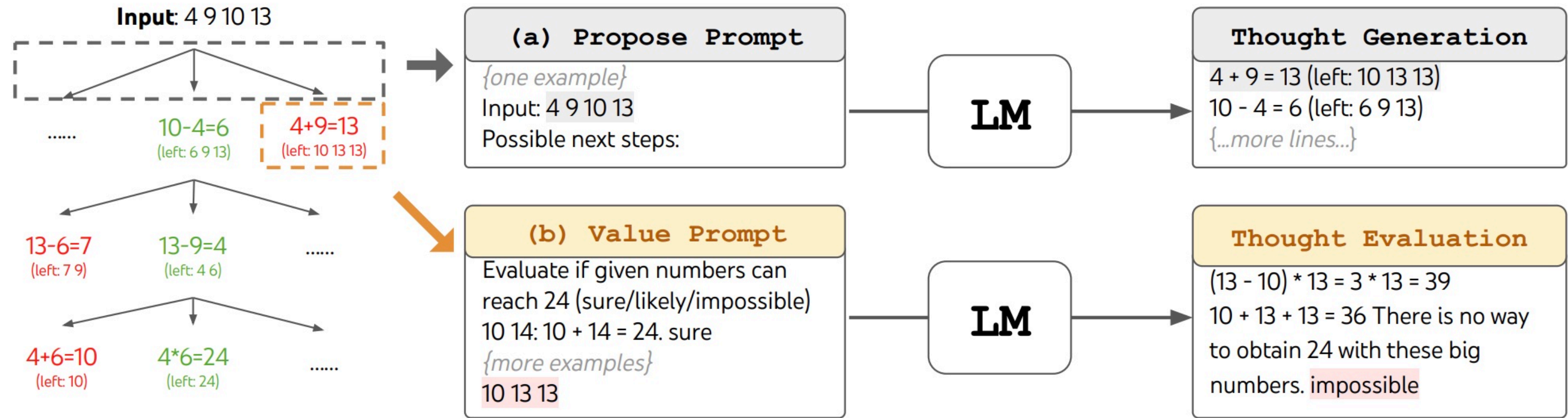
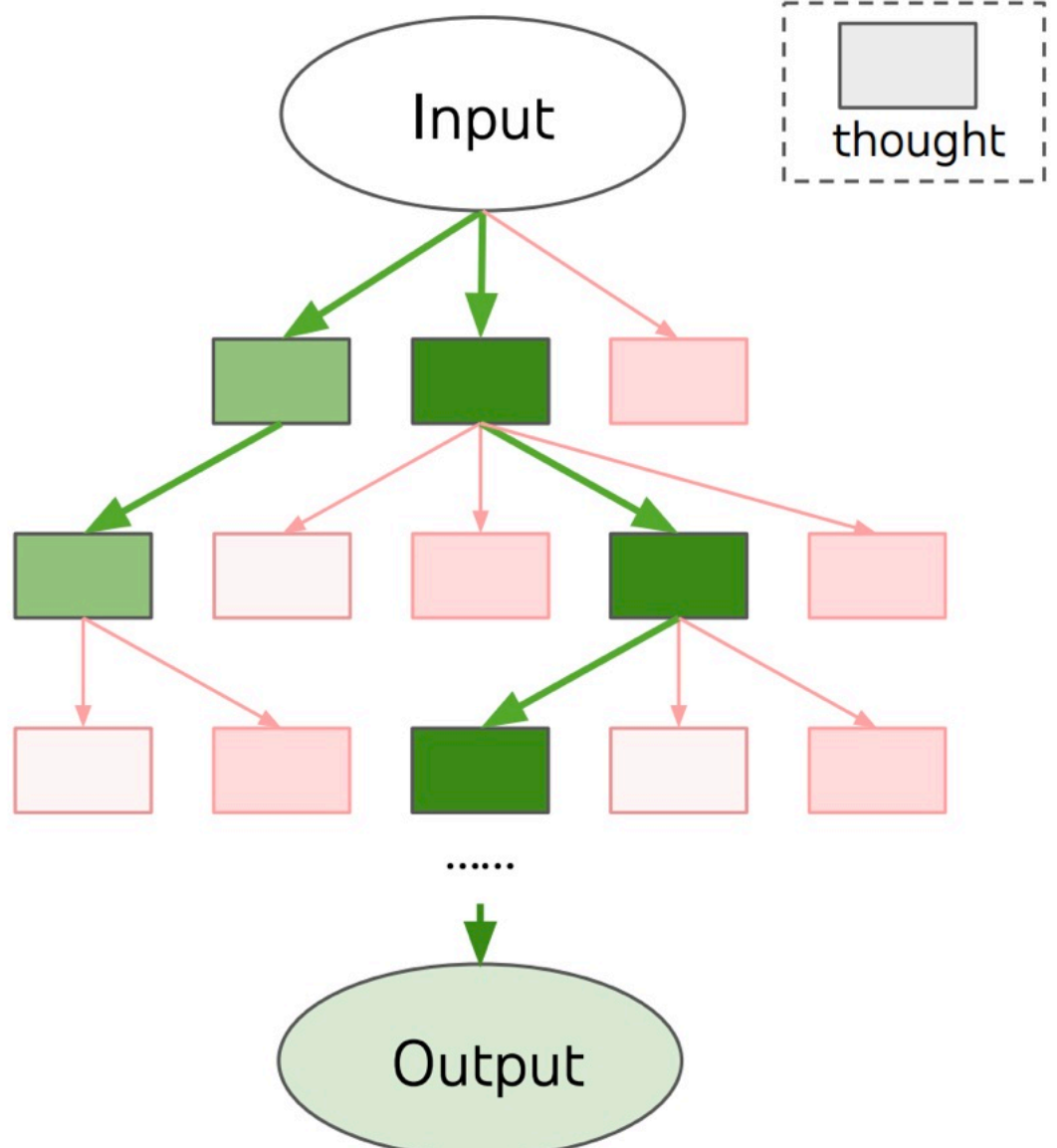
# Self-consistency

- Since LLMs are stochastic, sample responses **multiple times** and perform **majority vote** may be useful

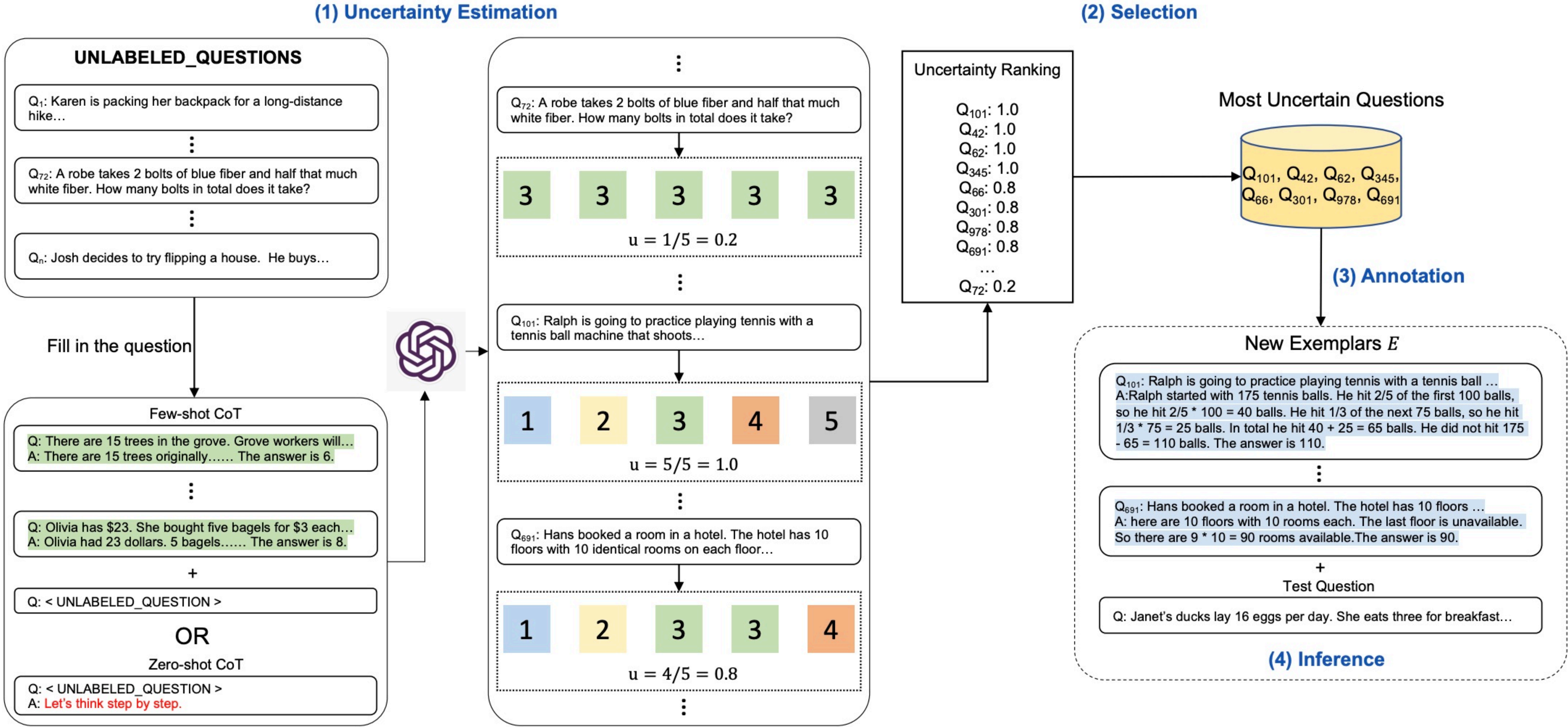


# Tree of Thoughts (ToT) Prompting

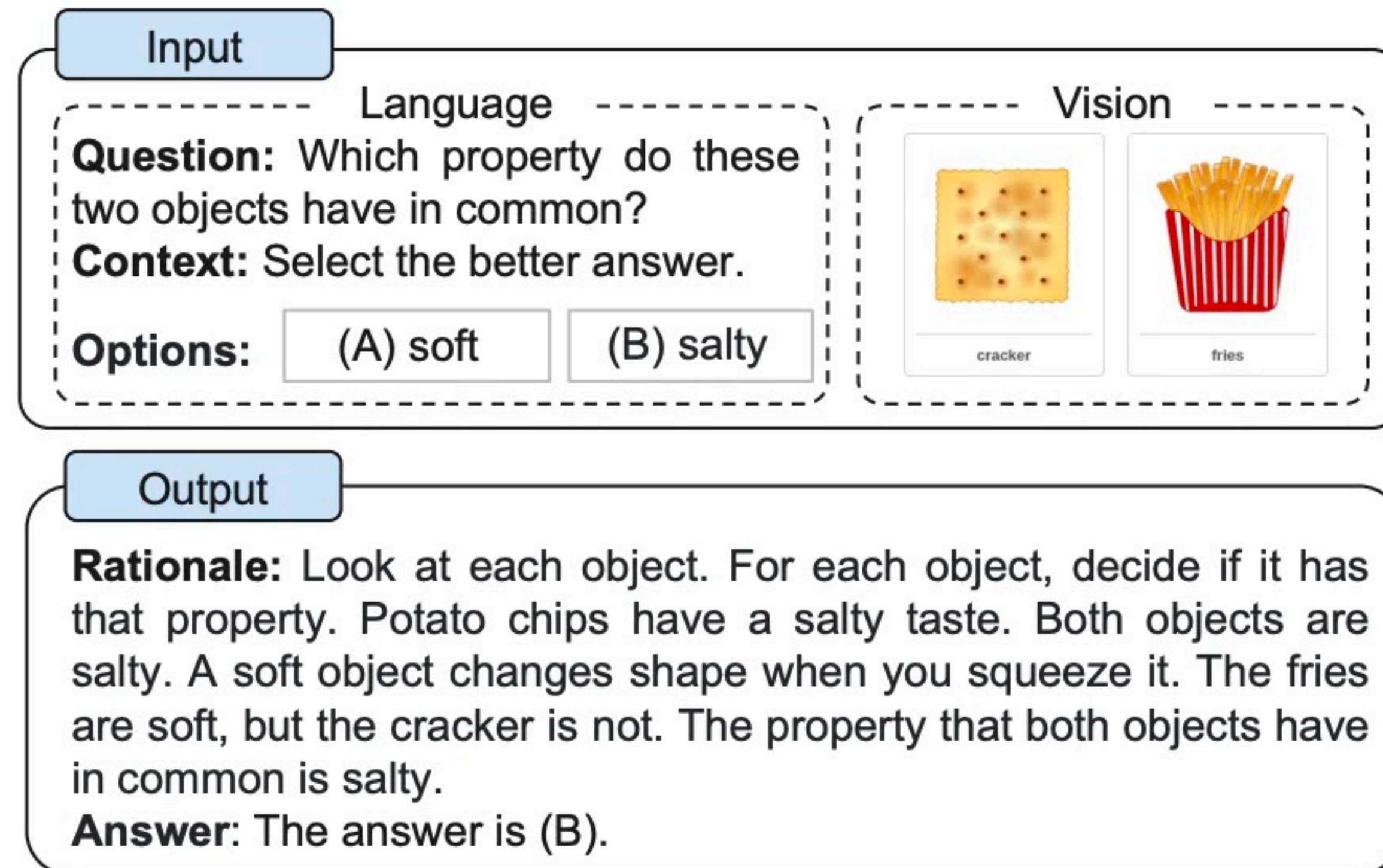
- Since CoT's reasoning steps may take a **diverse path** due to the stochastic nature of LLMs, exploring **many strategies** might be useful



# Active-Prompt

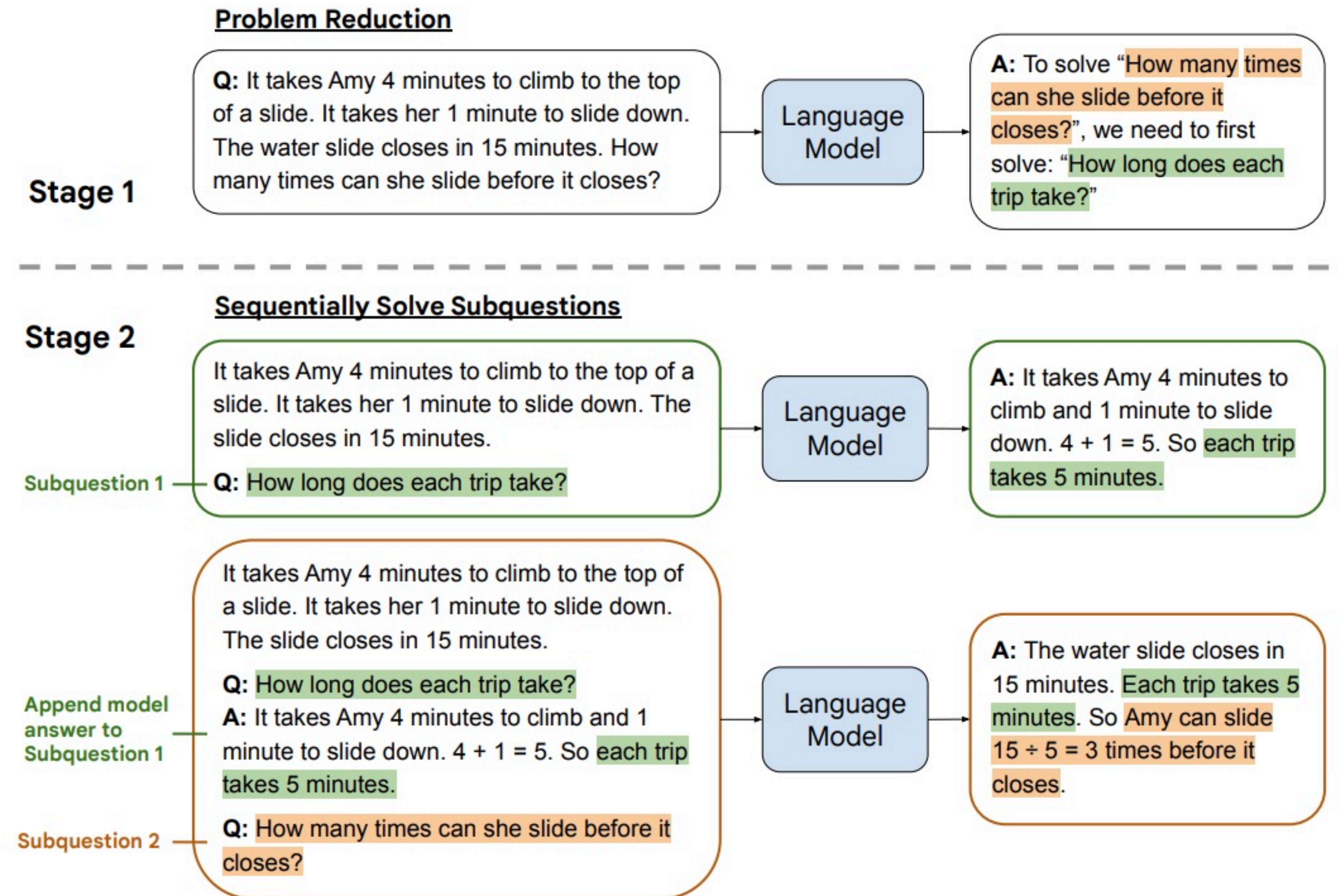


# Multi-modal CoT



# Least-to-Most (LtM) Prompting

- Break a problem into **sub-problems** and solve each sub-problems
- Similar to Divide-and-Conquer



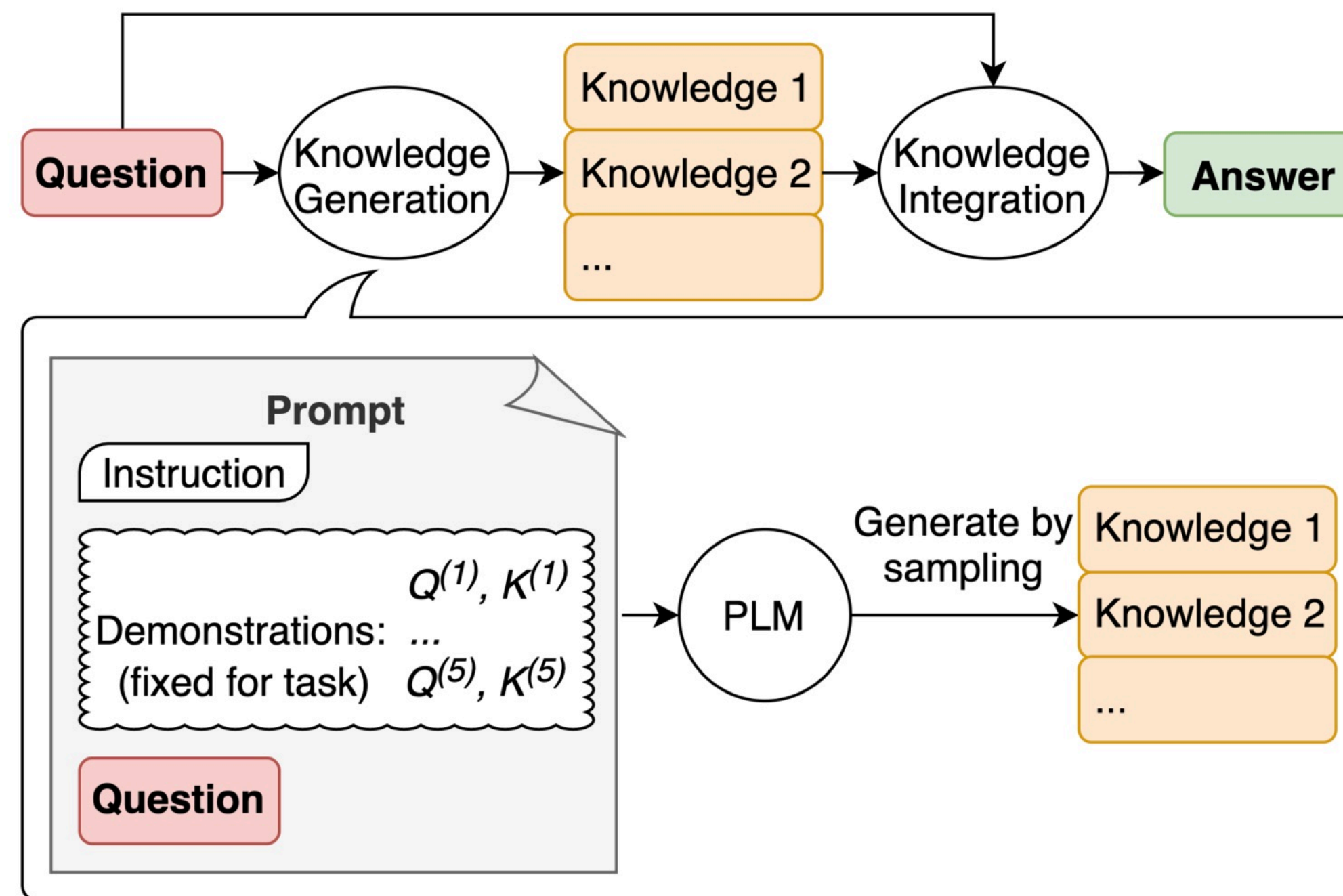
---

**PE: Others**

---

# Generated Knowledge Prompting

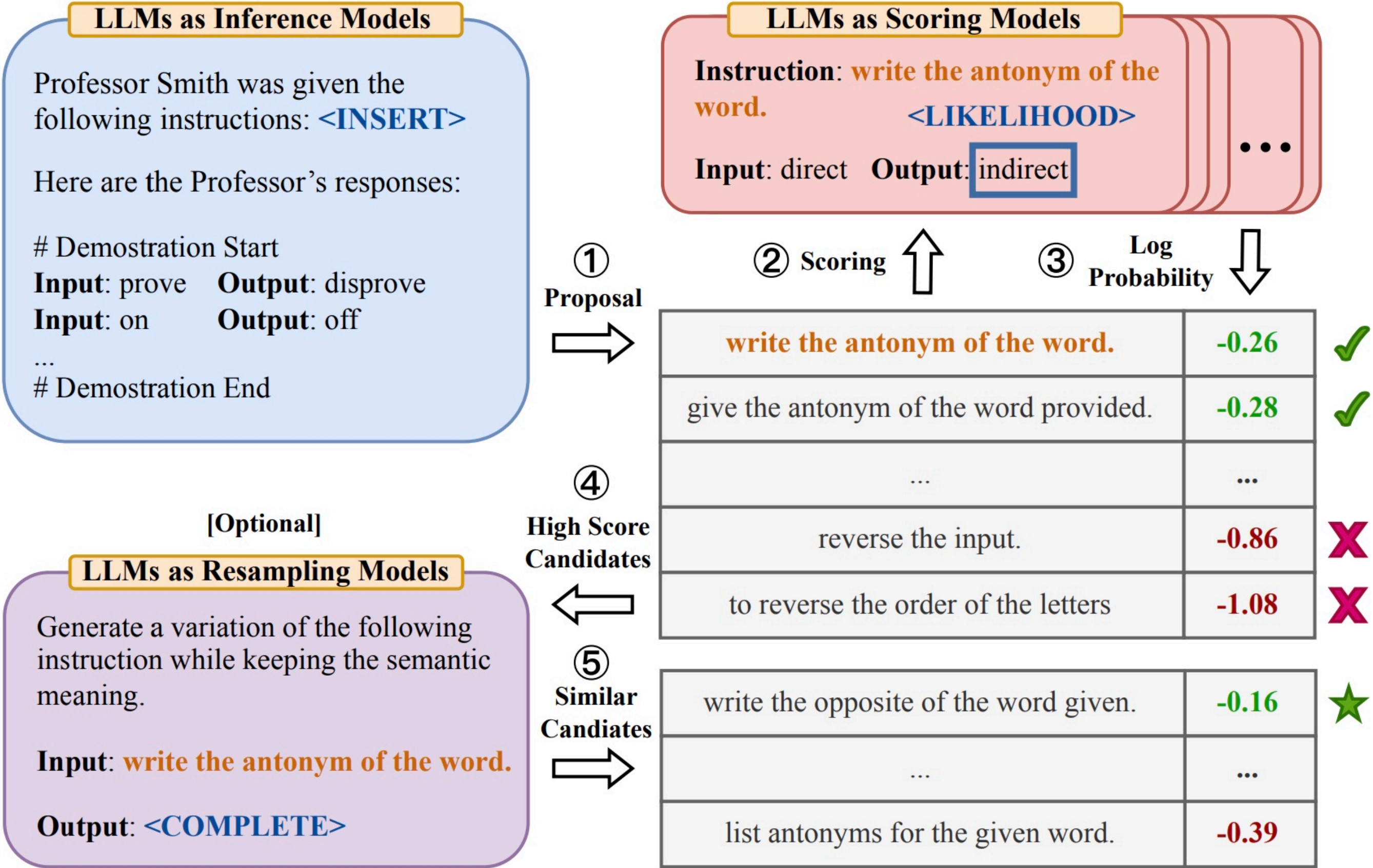
- Ask a model to **generate new knowledge** based on facts, e.g., summary, and use generated facts to improve performance



# Automatic Prompt Engineer (APE)

- **Automatically** improve prompts via LLMs

✓ Keep the high score candidates    
 ✗ Discard the low score candidates    
 ★ Final selected prompt with highest score





# Directional Stimulus Prompting

- Utilize reinforcement learning (RL) to tune one LLM to **generate hint/stimulus** for another LLM

**Input text**

Article: (CNN) For the first time in eight years, a TV legend returned to doing what he does best. Contestants told to "come on down!" on the April 1 edition of "[The Price Is Right](#)" encountered not host Drew Carey but another familiar face in charge of the proceedings. Instead, there was [Bob Barker](#), who hosted the TV game show for 35 years before stepping down in [2007](#). Looking spry at [91](#), Barker handled the first price-guessing game of the show, the classic "Lucky Seven," before turning hosting duties over to Carey, who finished up. Despite being away from the show for most of the past eight years, Barker didn't seem to miss a beat.

**Reference**

[Bob Barker](#) returned to host "[The Price Is Right](#)" on Wednesday . Barker, [91](#), had retired as host in [2007](#) .

---

**Directional Stimulus Prompting**

Q: Summarize the above article briefly in 2-3 sentences **based on the hint.**

**Hint:** [Bob Barker](#); [TV](#); [April 1](#); "[The Price Is Right](#)"; [2007](#); [91](#).

**Model output**

A: On [April 1](#), [Bob Barker](#) returned to the [TV](#) show "[The Price Is Right](#)" after eight years . Looking spry at [91](#), Barker handled the first price-guessing game . Barker stepped down as host in [2007](#), but didn't seem to miss a beat .

ROUGE-1: 48.39 👍

**Standard Prompting**

Q: Summarize the above article briefly in 2-3 sentences.

**Model output**

A: Bob Barker, the former host of \"The Price Is Right\", returned to the show for the first time in 8 years. He hosted the first game of the show, and then handed the hosting duties over to Drew Carey.

ROUGE-1: 34.48 👎

# ReAct

- Generate interleaving Reasoning traces and task-specific Actions

**(1) Hotspot QA**

**Question:** Aside from the Apple Remote, what other device can control the program Apple Remote was originally designed to interact with?

**(1a) Standard**

**Answer:** iPod ❌

**(1b) CoT (Reason Only)**

**Thought:** Let's think step by step. Apple Remote was originally designed to interact with Apple TV. Apple TV can be controlled by iPhone, iPad, and iPod Touch. So the answer is iPhone, iPad, and iPod Touch.

**Answer:** iPhone, iPad, iPod Touch ❌

**(1c) Act-Only**

**Act 1:** Search[Apple Remote]  
**Obs 1:** The Apple Remote is a remote control ...

**Act 2:** Search[Front Row]  
**Obs 2:** Could not find [Front Row]. Similar: ...

**Act 3:** Search[Front Row (software)]  
**Obs 3:** Front Row is a discontinued media center software ...

**Act 4:** Finish[yes] ❌

**(1d) ReAct (Reason + Act)**

**Thought 1:** I need to search Apple Remote and find the program it was originally designed to interact with.  
**Act 1:** Search[Apple Remote]  
**Obs 1:** The Apple Remote is a remote control introduced in October 2005 by Apple ... originally designed to control the Front Row media center program ...

**Thought 2:** Apple Remote was originally designed to control the Front Row media center program. I need to search Front Row next and find what other device can control it.  
**Act 2:** Search[Front Row]  
**Obs 2:** Could not find [Front Row]. Similar: ['Front Row Seat to Earth', 'Front Row Motorsports', 'Front Row (software)', ...]

**Thought 3:** Front Row is not found. I need to search Front Row (software) .  
**Act 3:** Search[Front Row (software)]  
**Obs 3:** Front Row is a discontinued media center software ...

**Thought 4:** Front Row (software) is controlled by an Apple Remote or the keyboard function keys. So the answer is keyboard function keys.  
**Act 4:** Finish[keyboard function keys] ✓

---

**(2) AlfWorld**

You are in the middle of a room. Looking quickly around you, you see a cabinet 6, a cabinet 1, a coffee machine 1, a countertop 3, a stove burner 1, and a toaster 1.  
**Your task is to:** Put some pepper shaker on a drawer.

**(2a) Act-Only**

**Act 1:** Go to drawer 1  
**Obs 1:** The drawer 1 is closed.

**Act 2:** Open drawer 1  
**Obs 2:** You open the drawer 1. The drawer 1 is open. In it, you see a dish sponge 2, and a spoon 1.

**Act 3:** Go to sinkbasin 1  
**Obs 3:** On the sinkbasin 1, you see a dish sponge 3, a spatula 1, and a spoon 2.

**Act 4:** Take peppershaker 1 from sinkbasin 1  
**Obs 4:** Nothing happens.

**Act 5:** Take peppershaker 1 from sinkbasin 1  
**Obs 5:** Nothing happens. ❌

**(2b) ReAct (Reason + Act)**

**Act 1:** Think[First I need to find a pepper shaker...more likely to appear in cabinets (1-6), countertops (1-3), ... ]  
**Act 2:** Go to cabinet 1  
**Obs 2:** On the cabinet 1, you see a vase 2.  
(Here the agent go to cabinet 1, then cabinet 2, 3, then countertop 1 and 2)

**Act 6:** Go to countertop 3  
**Obs 6:** On the countertop 3, you see a apple 1, a bread 1, ..... a pepper shaker 1, and a vase 3.

**Act 7:** Take pepper shaker 1 from countertop 3  
**Obs 7:** You pick up the pepper shaker 1 from the countertop 3.

**Act 8:** Think[Now I find a pepper shaker 1. Next, I need to put it in/on drawer 1. ]  
**Act 9:** Go to drawer 1  
**Obs 9:** Drawer 1 is closed.

**Act 10:** Open drawer 1  
**Obs 10:** You open Drawer 1 ...

**Act 11:** Put pepper shaker 1 in/on drawer 1  
**Obs 11:** You put pepper shaker 1 in/on the drawer 1. ✓

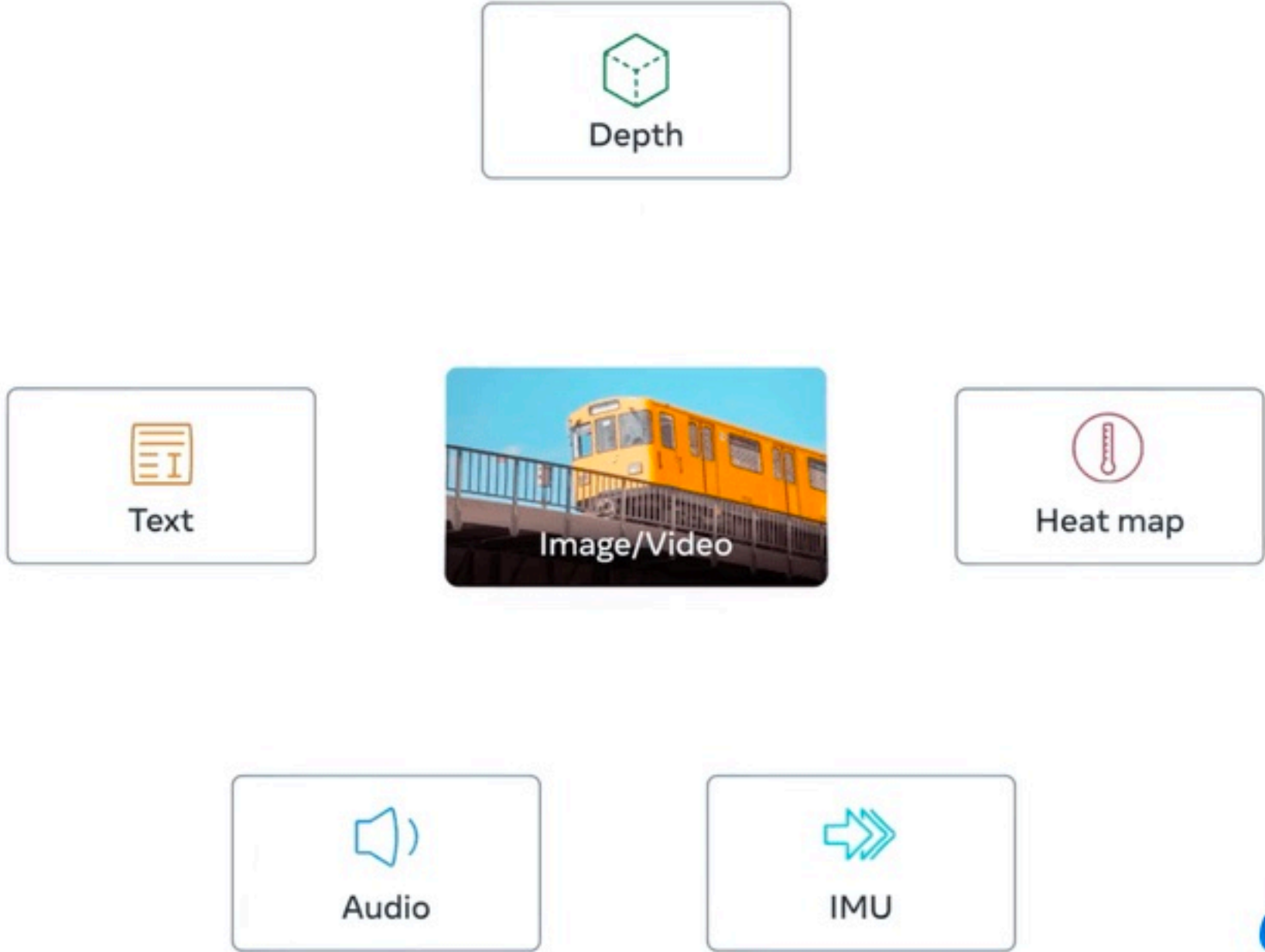
---

# Closing Remarks

---

# A Glimpse into the Future

- Multi-modal LLMs, e.g., PandaGPT



A recording of thunder and pouring rain.

Write a story to a child about what you see and hear.

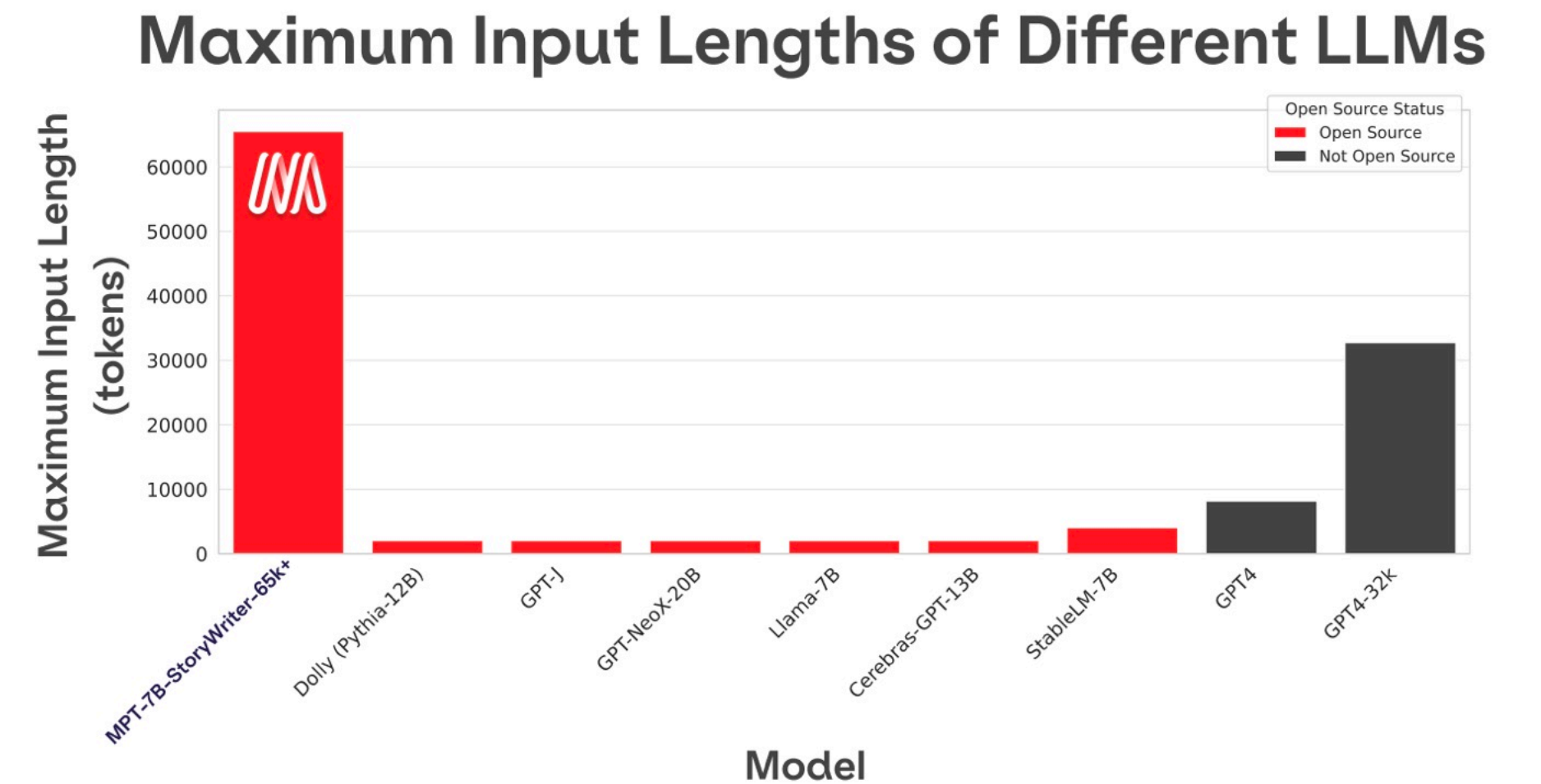
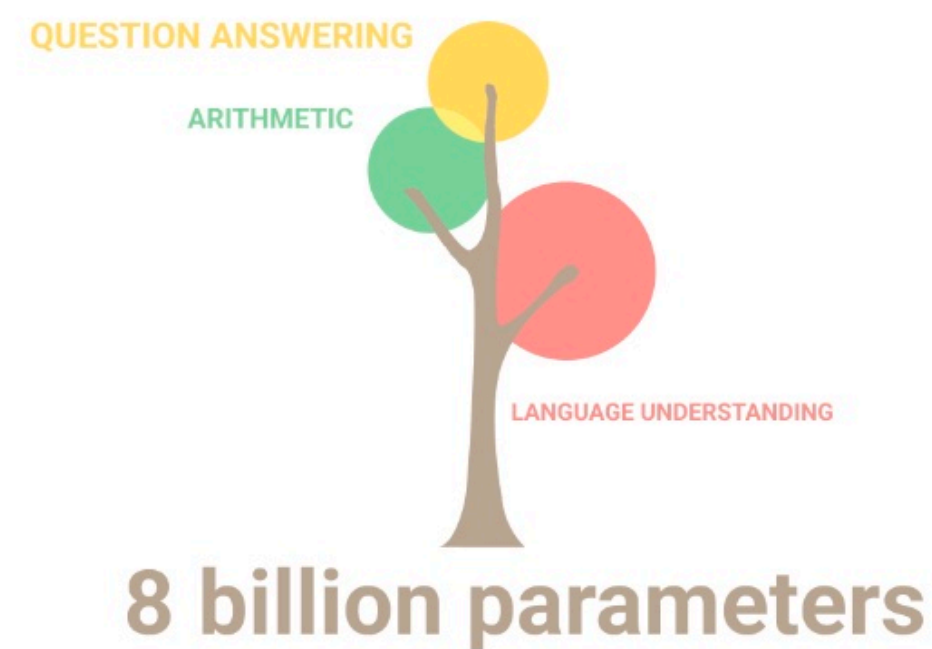


As I look out the window, I see a beautiful scene of a person walking on the road in the distance. The person is walking under a cloudy sky, which is dark and stormy. The sound of the wind is blowing, and I can hear the distant sound of thunder rumbling in the background. The person is walking with a backpack, and it seems like they are on a journey or adventure. The stormy weather might make the journey a bit more challenging, but the person seems determined to continue their journey despite the conditions.



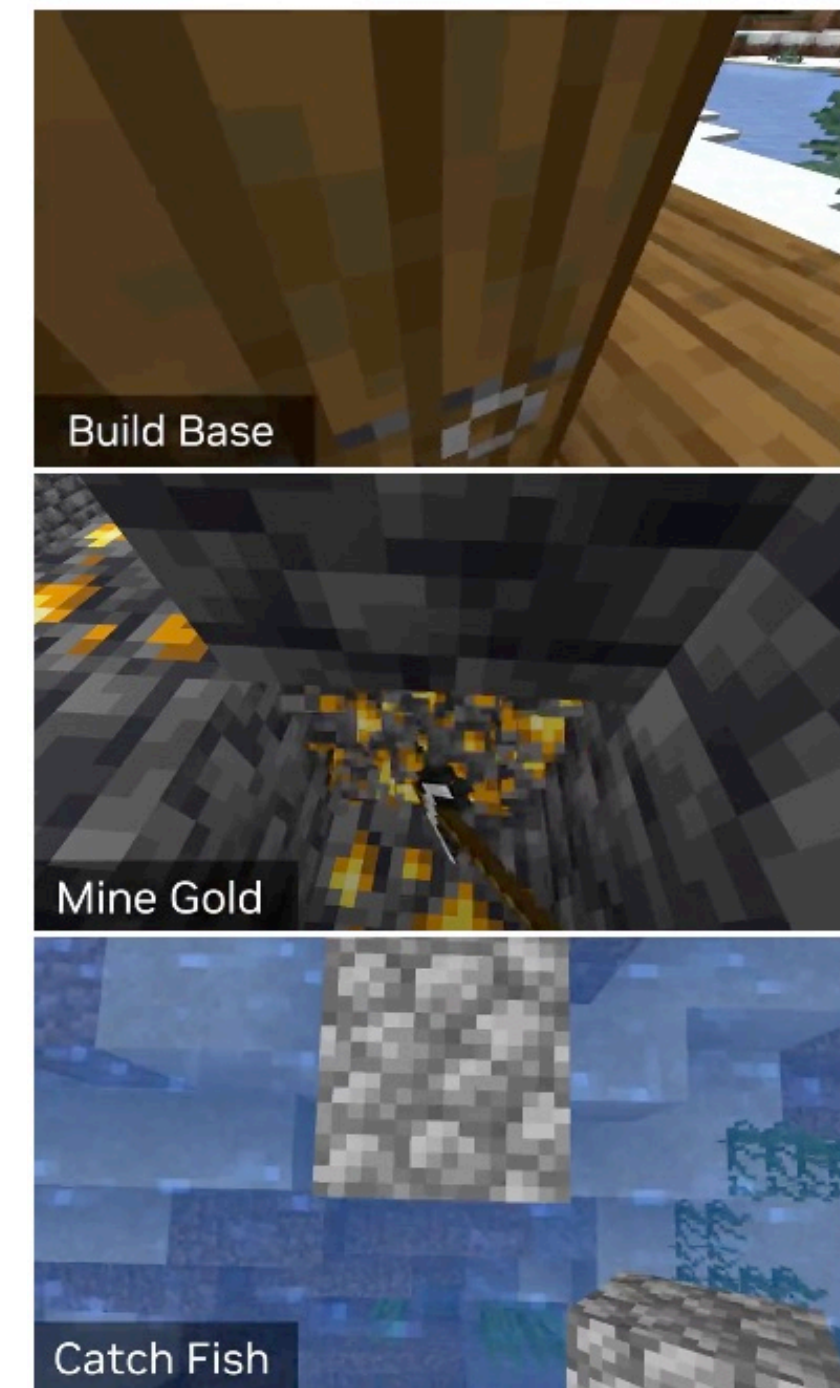
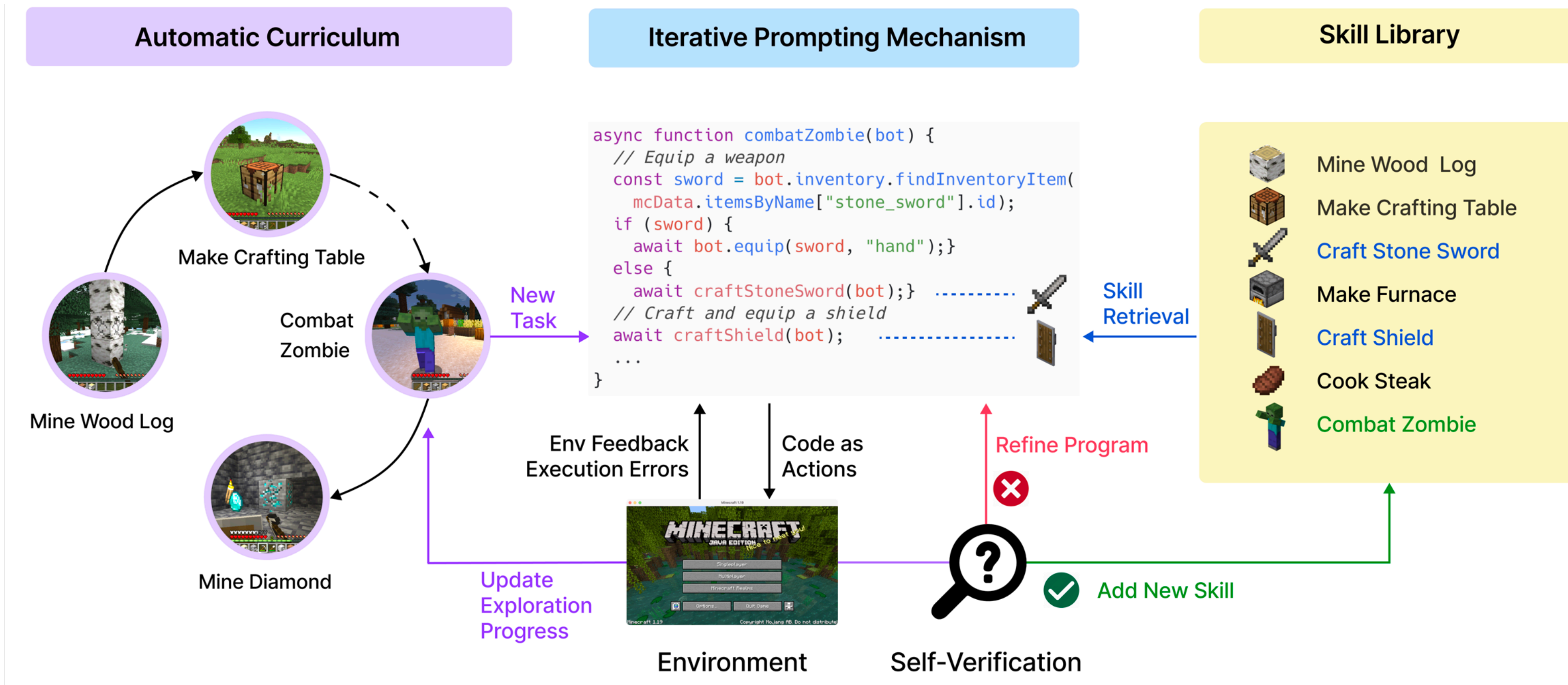
# A Glimpse into the Future

- Larger context and larger LLMs, e.g., PaLM **540B**, MPT-7B-StoryWriter-**65k+**



# A Glimpse into the Future


- LLMs as a **tool maker/user**, e.g., Voyager, LATM



# A Glimpse into the Future

- Embodied LLMs, LLMs that can affect real world, e.g., PaLM-E

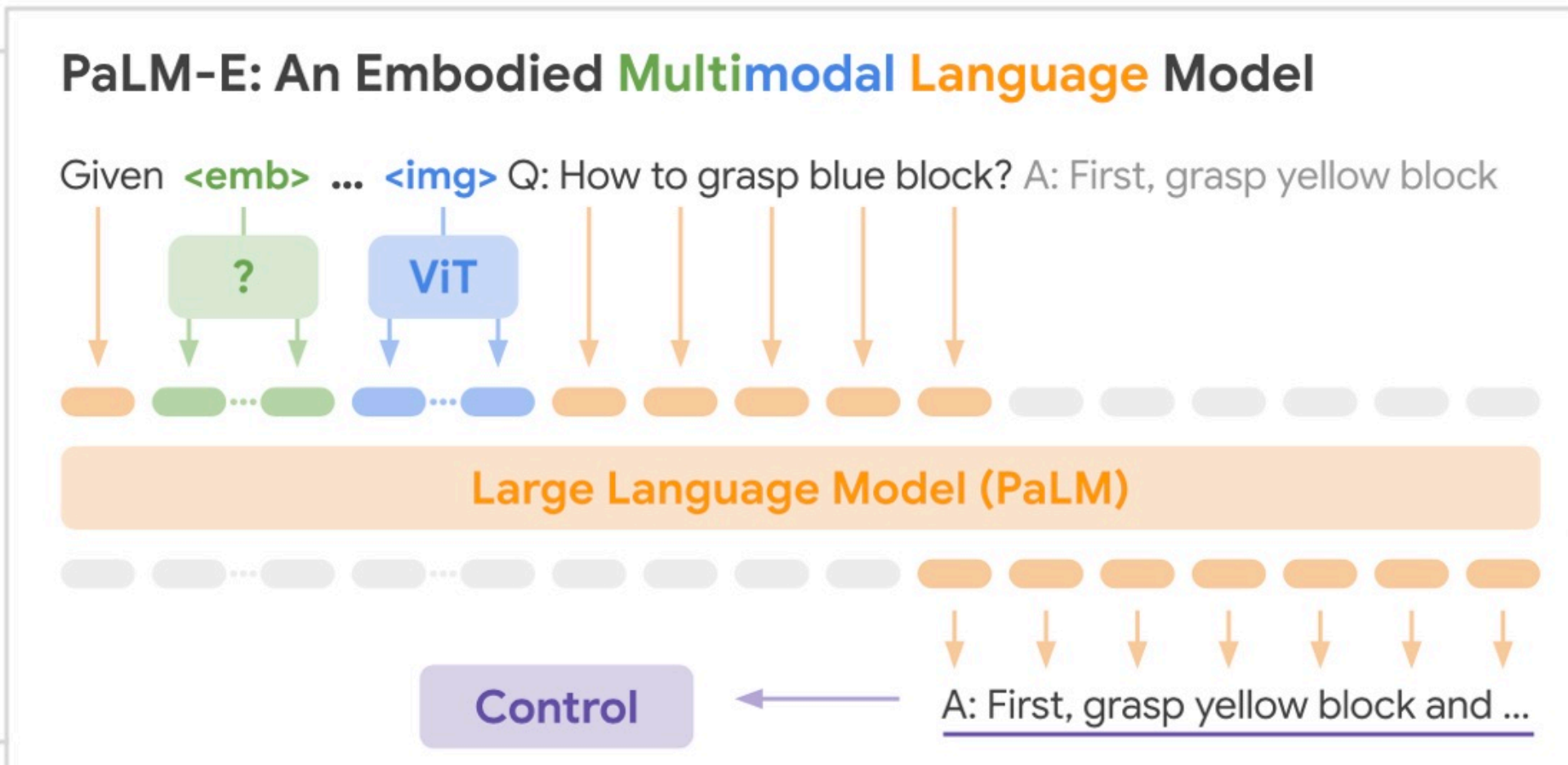
**Mobile Manipulation**



Human: Bring me the rice chips from the drawer. Robot: 1. Go to the drawers, 2. Open top drawer. I see **<img>**. 3. Pick the green rice chip bag from the drawer and place it on the counter.

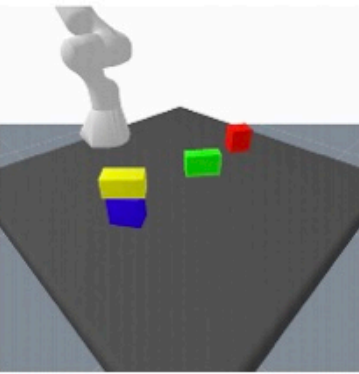
**PaLM-E: An Embodied Multimodal Language Model**

Given **<emb>** ... **<img>** Q: How to grasp blue block? A: First, grasp yellow block




**Control** ← A: First, grasp yellow block and ...

**Task and Motion Planning**

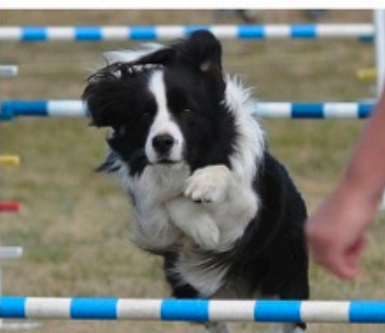


Given **<emb>** Q: How to grasp blue block? A: First grasp yellow block and place it on the table, then grasp the blue block.

**Visual Q&A, Captioning ...**



Given **<img>**. Q: What's in the image? Answer in emojis.  
A: 🍏 🍌 🍇 🍋 🍓 🍓 🍓

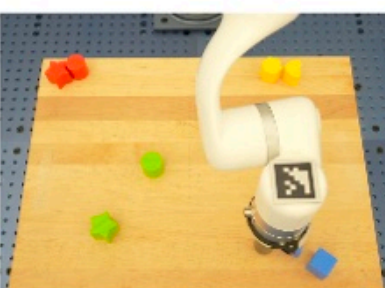


Describe the following **<img>**:  
A dog jumping over a hurdle at a dog show.

**Language Only Tasks**

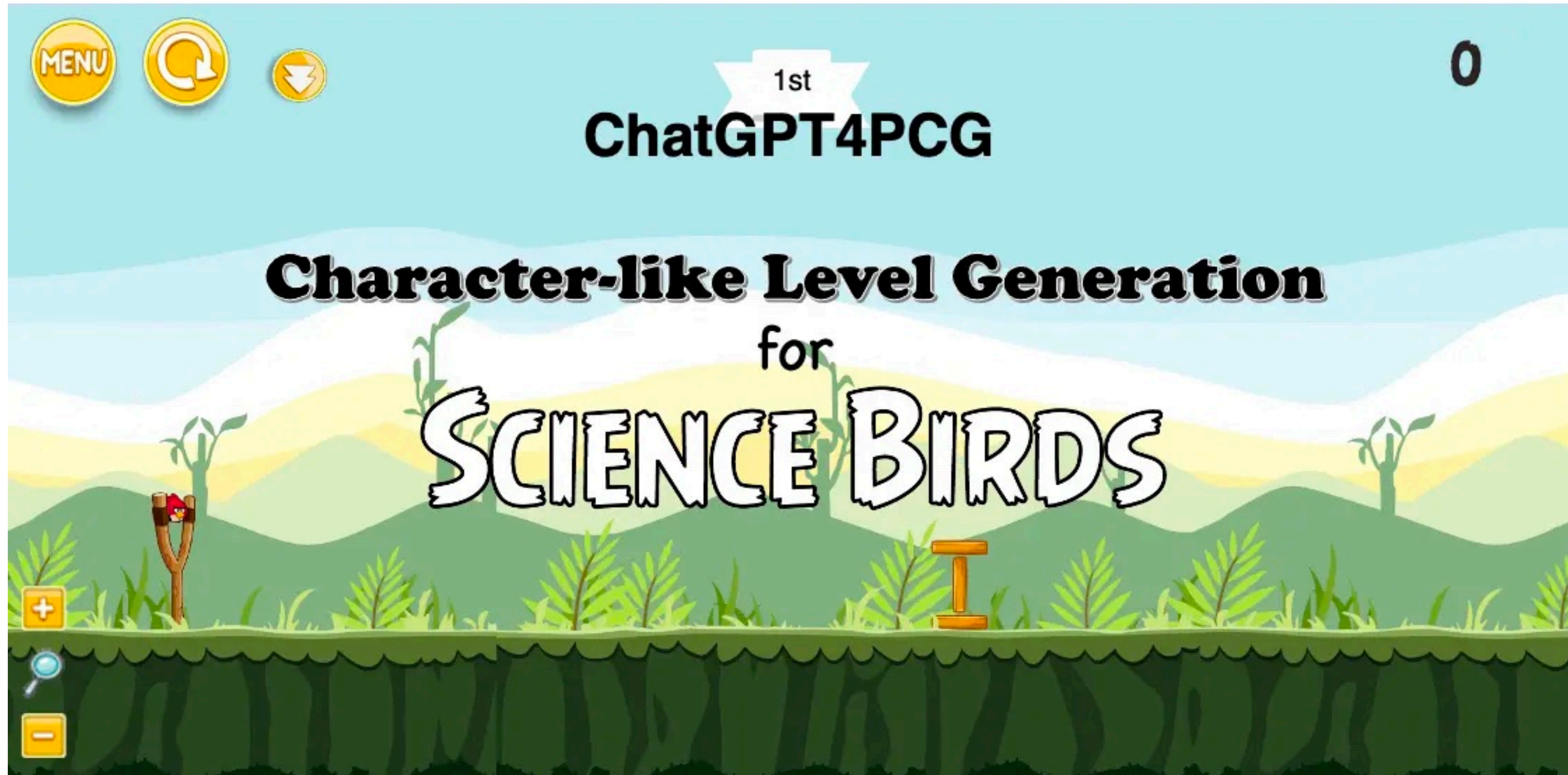
Q: Miami Beach borders which ocean? A: Atlantic. Q: What is 372 x 18? A: 6696. Q: Write a Haiku about embodied LLMs. A: Embodied language. Models learn to understand. The world around them.

**Tabletop Manipulation**



Given **<img>** Task: Sort colors into corners.  
Step 1. Push the green star to the bottom left.  
Step 2. Push the green circle to the green star.

# Need a Place to Practice PE Skills?



[chatgpt4pcg.github.io](https://chatgpt4pcg.github.io)



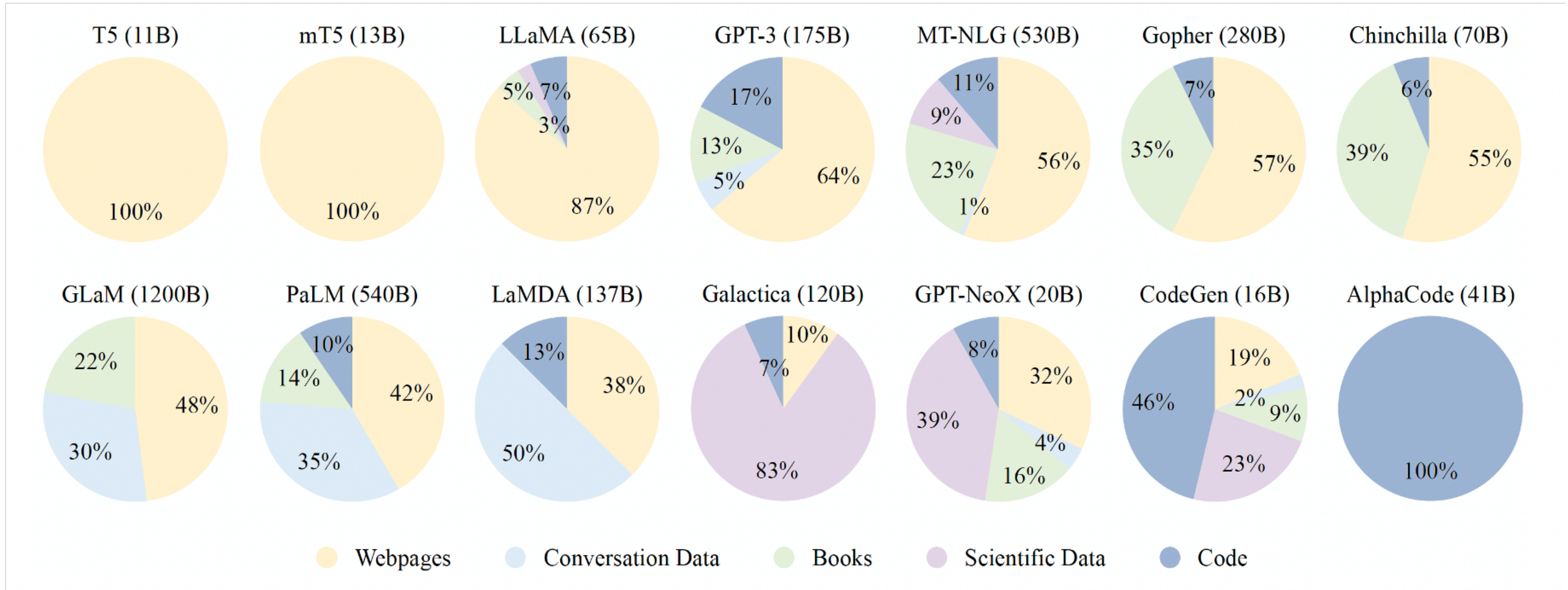


---

# Bonus: How to Choose LLMs?

- Based on these factors
  - **Context length:** Its memory
  - **Training dataset:** Its prior knowledge
  - **Model size:** Its capabilities and hardware requirements
  - **License:** Research-only or business-ok

# Bonus: How to Choose LLMs?



---

# Bonus: More Topics to Explore

- Prompt patterns (SE for prompt)
- Adversarial Prompting
  - Prompt injection
  - Prompt leaking
  - Jailbreaking
- Soft prompts
  - Interpretable soft prompts
- AI-generated content detection
- Prompt engineering for diffusion-based/GAN-based models
  - Negative prompts
  - Non-text input as a prompt
- LLM: Grounding, Alignment, Distillation