

NLP Course

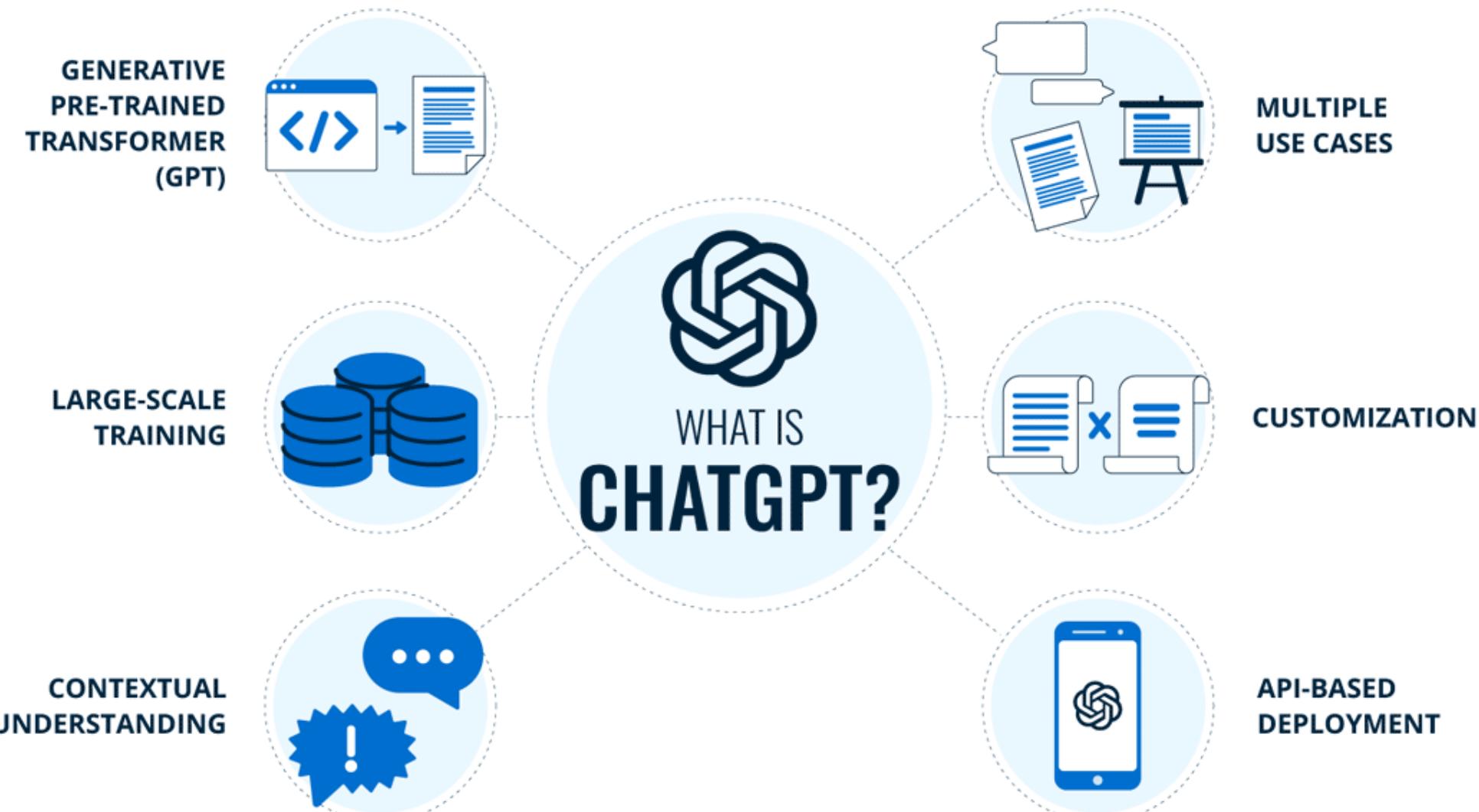
ChatGPT-based Applications

Nguyen Quoc Thai

CONTENT

1	Introduction
2	Prompt Engineering
3	ChatGPT Applications
4	Limitations
5	Generative AI Models

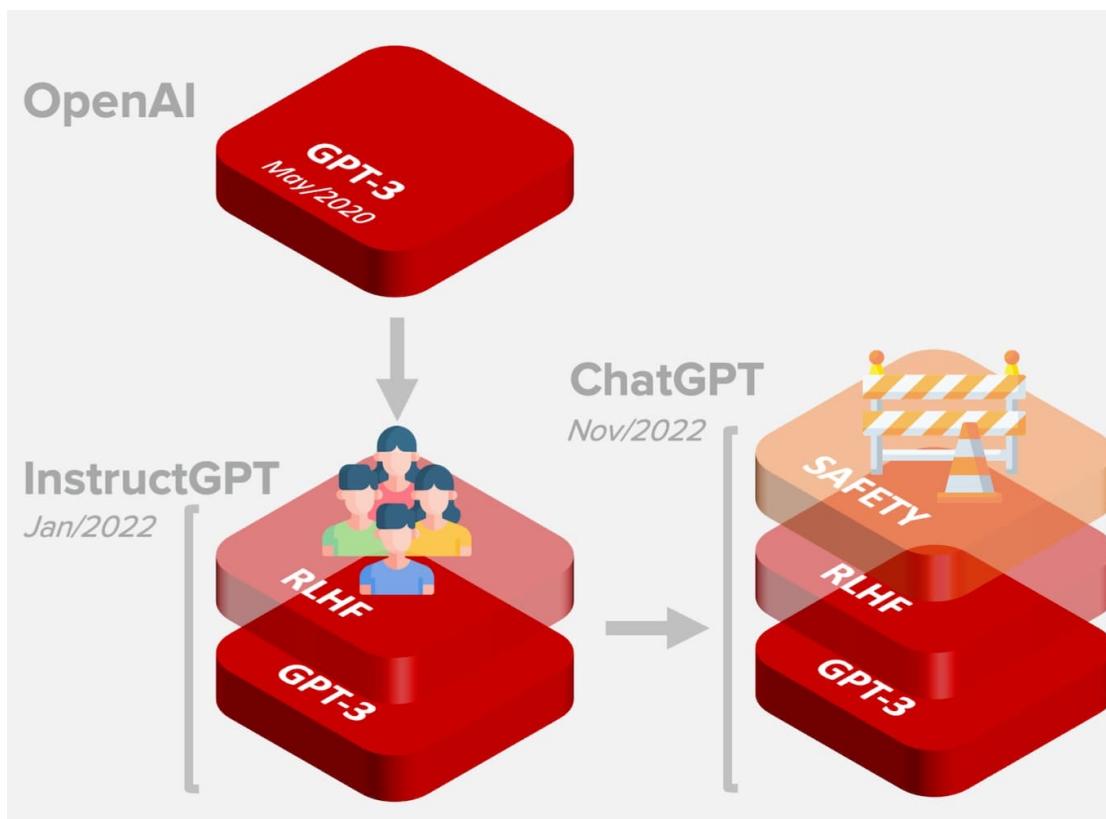
1 – Introduction



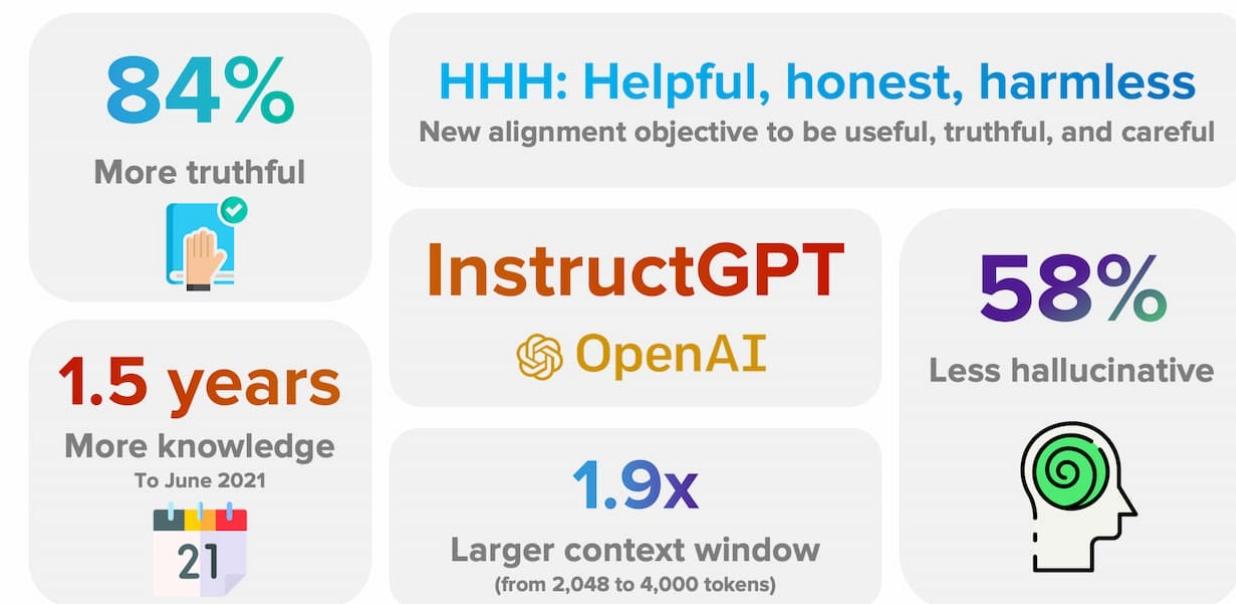
1 – Introduction

!

GPT3, InstructGPT, ChatGPT, GPT4



Source: <https://www.theinsaneapp.com/2023/05/instructgpt-vs-chatgpt.html>

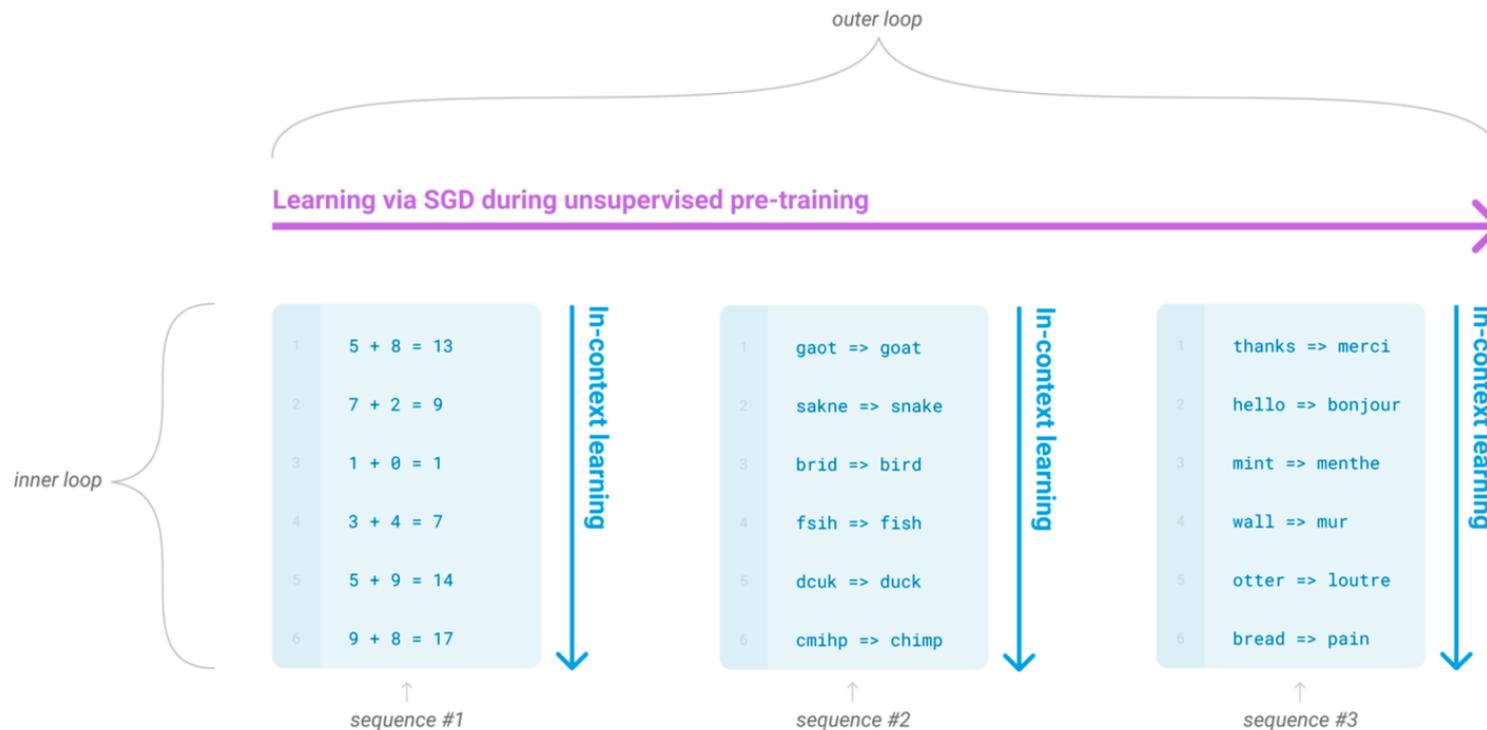


1 – Introduction



Language model meta-learning

- Model develops a broad set of skills and pattern recognition abilities at training time
- Uses those abilities at inference time to rapidly adapt to or recognize the desired task

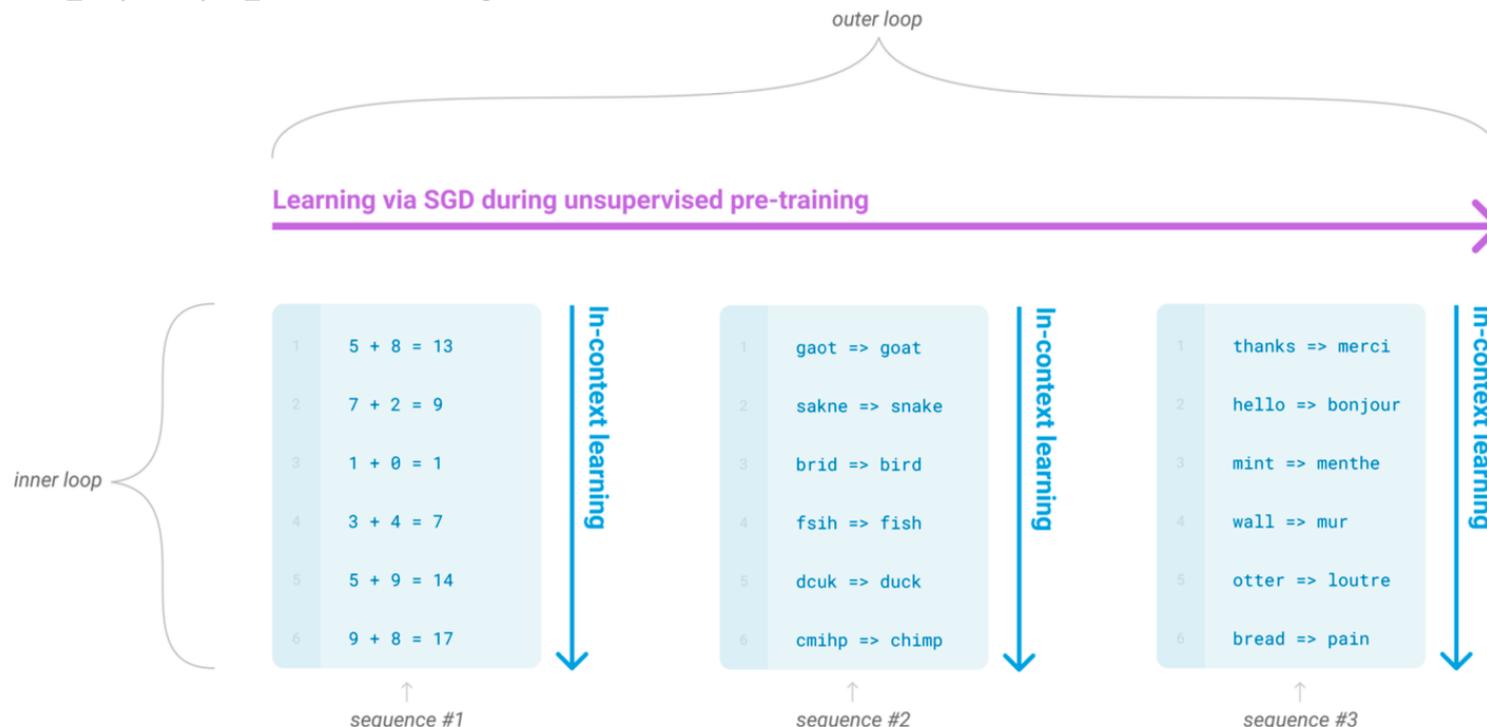


1 – Introduction



In-context Learning (ICL)

- Using the text input as a form of task specification
- Based on “**prompts**” of the task and is then expected to complete further instances of the task simply by predicting what comes next

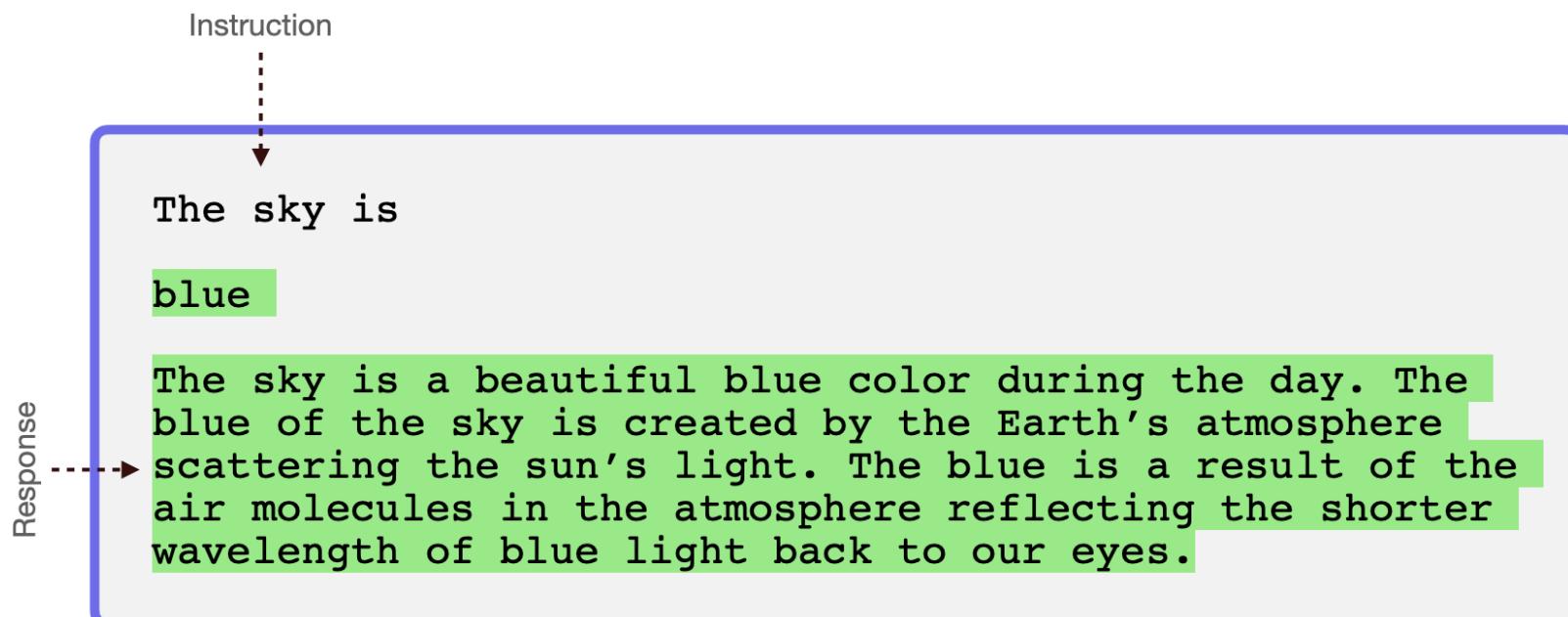


1 – Introduction

!

In-context Learning (ICL)

- “Prompts”: a natural language instruction and/or a few demonstrations



1 – Introduction



In-context Learning (ICL)

- Three settings for in-context learning

Few-shot

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

1 Translate English to French: ← task description
2 sea otter => loutre de mer ← examples
3 peppermint => menthe poivrée
4 plush girafe => girafe peluche
5 cheese => ← prompt

Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.

1 Translate English to French: ← task description
2 cheese => ← prompt

One-shot

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

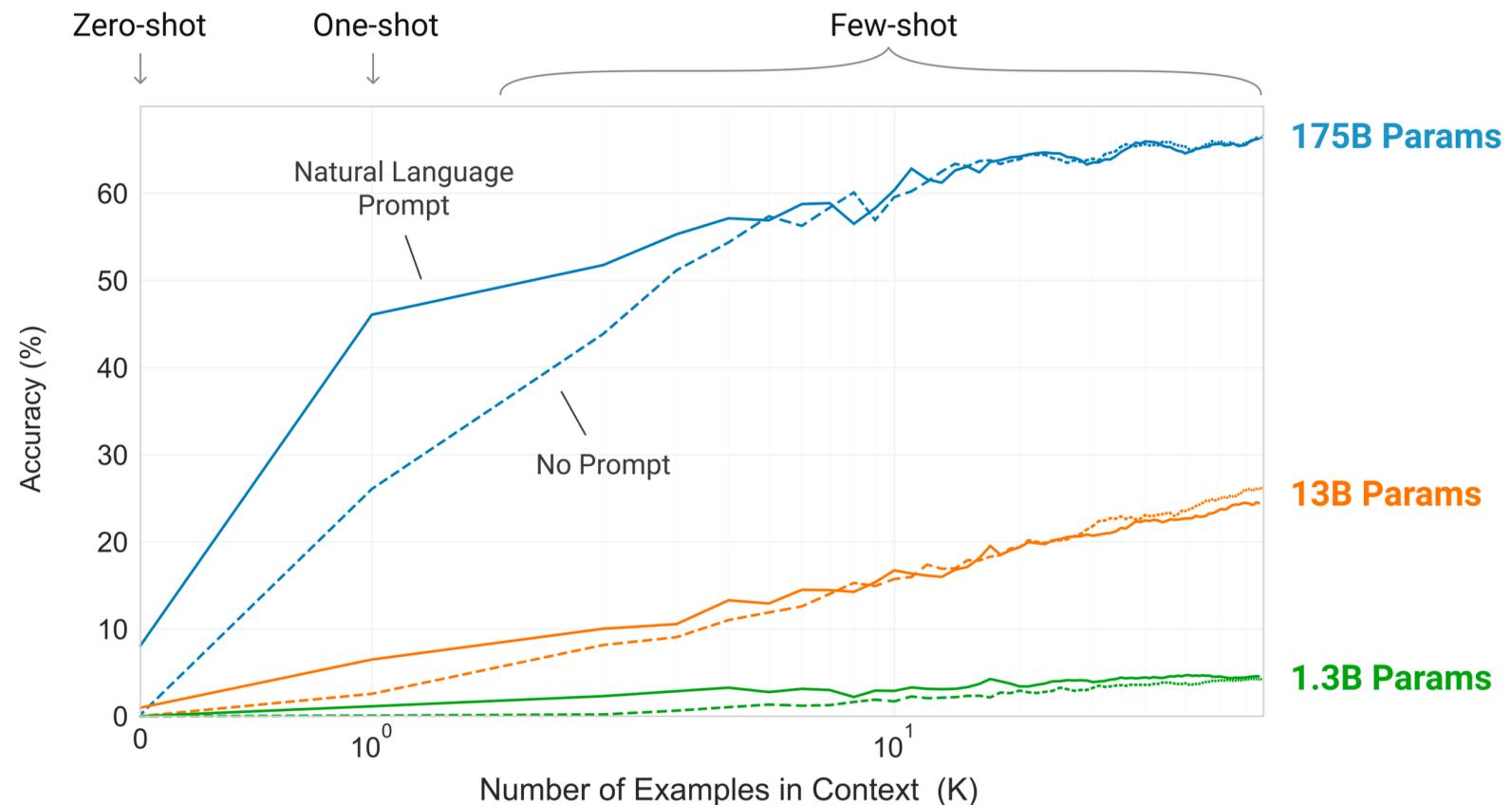
1 Translate English to French: ← task description
2 sea otter => loutre de mer ← example
3 cheese => ← prompt

1 – Introduction



In-context Learning (ICL)

- Three settings for in-context learning: **Few-shot is best**



2 – Prompt Engineering

!

What are prompts?

- 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

What is prompt engineering?

Prompt engineering is a process of creating a set of prompts, or questions, that are used to guide the user toward a desired outcome. It is an effective tool for designers to create user experiences that are easy to use and intuitive. This method is often used in interactive design and software development, as it allows users to easily understand how to interact with a system or product..

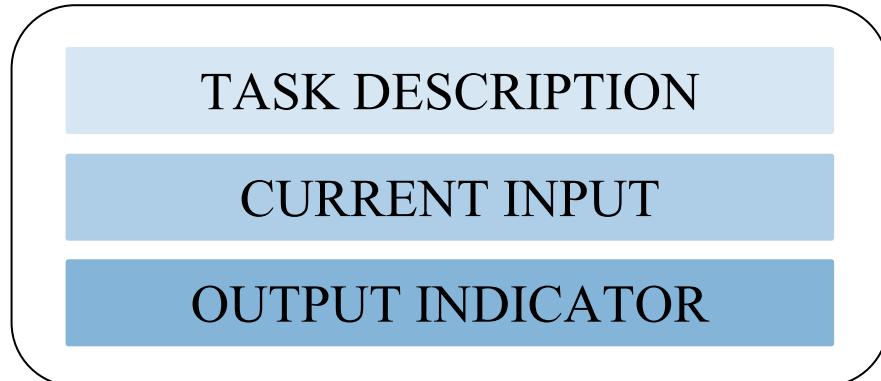
2 – Prompt Engineering



Elements of a Prompt

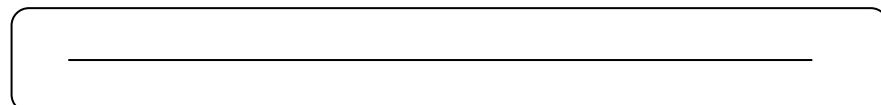
(1) Describe the task and the general setting (Zero-Shot)

Prompt



Language
Model

Response



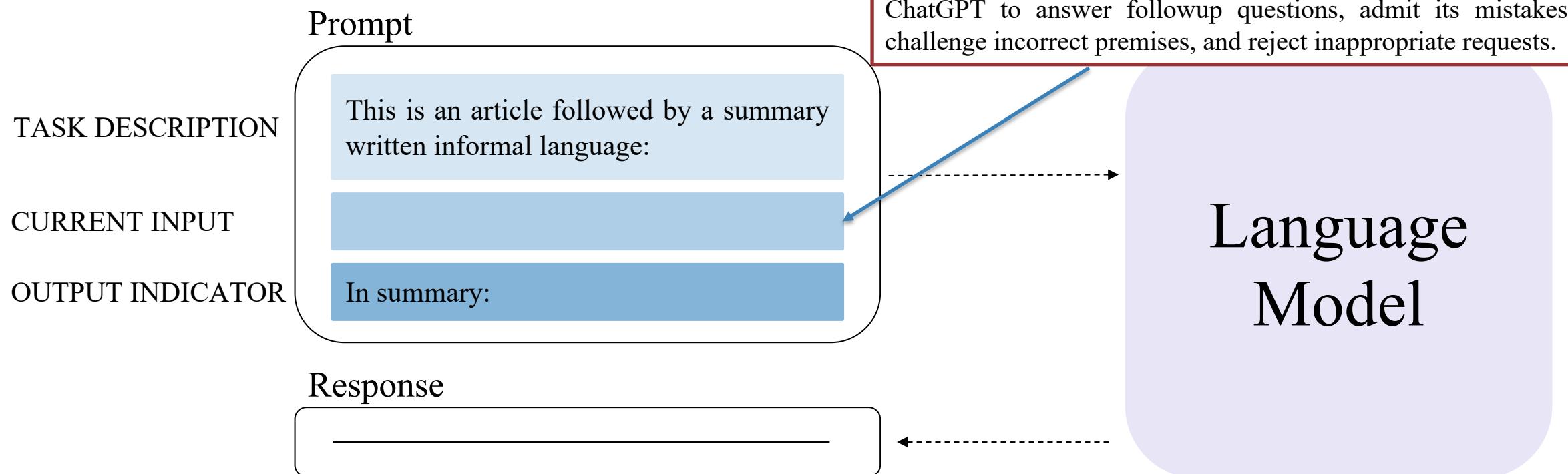
2 – Prompt Engineering



Elements of a Prompt

(1) Describe the task and the general setting (Zero-shot)

➤ Example: Summarization Task

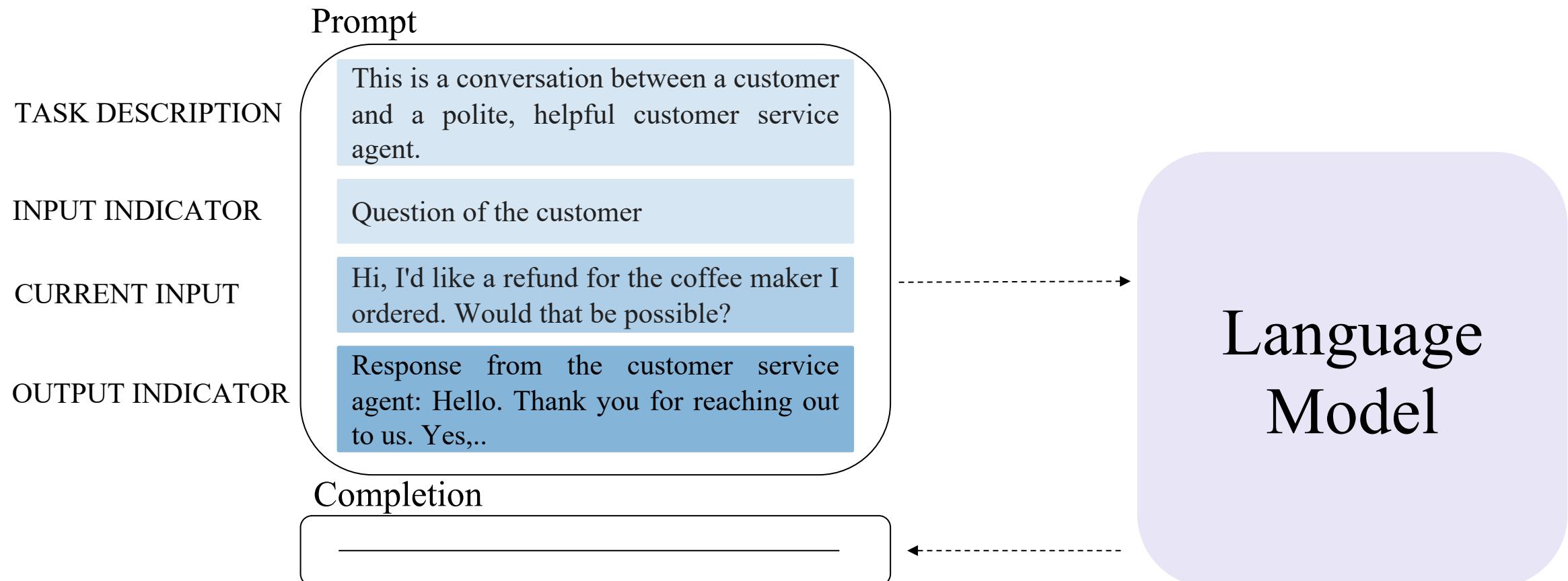


2 – Prompt Engineering



Elements of a Prompt

(1) Describe the task and the general setting (Zero-shot)

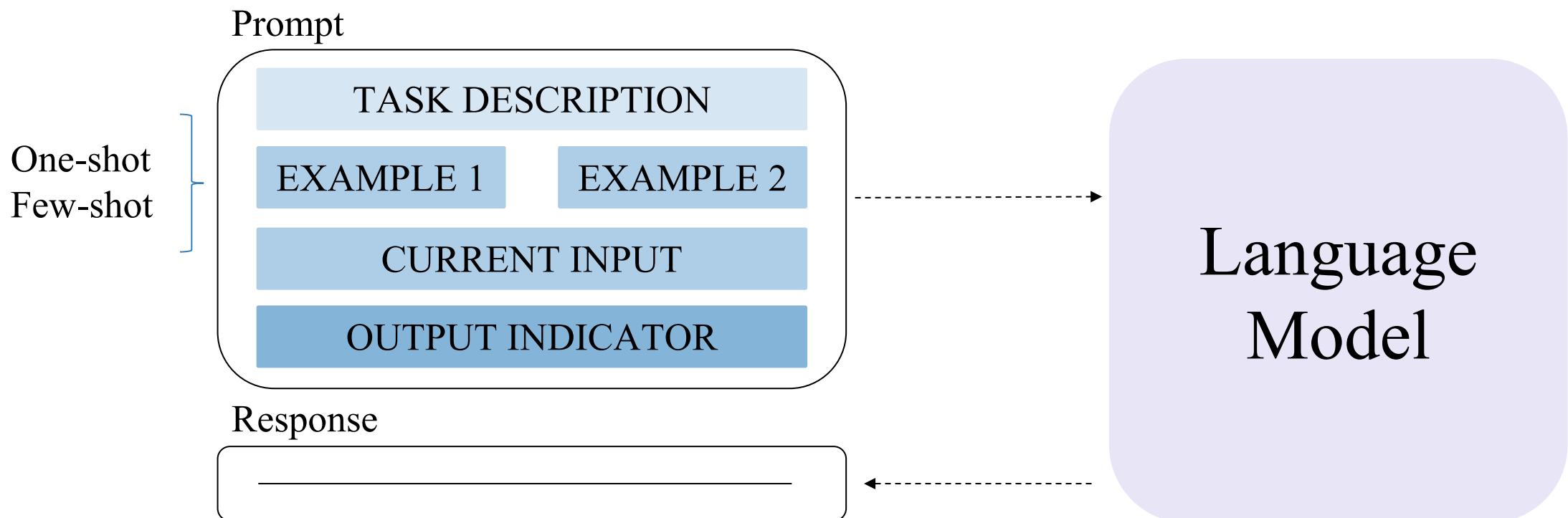


2 – Prompt Engineering



Elements of a Prompt

(2) Describe the task, the general setting and examples

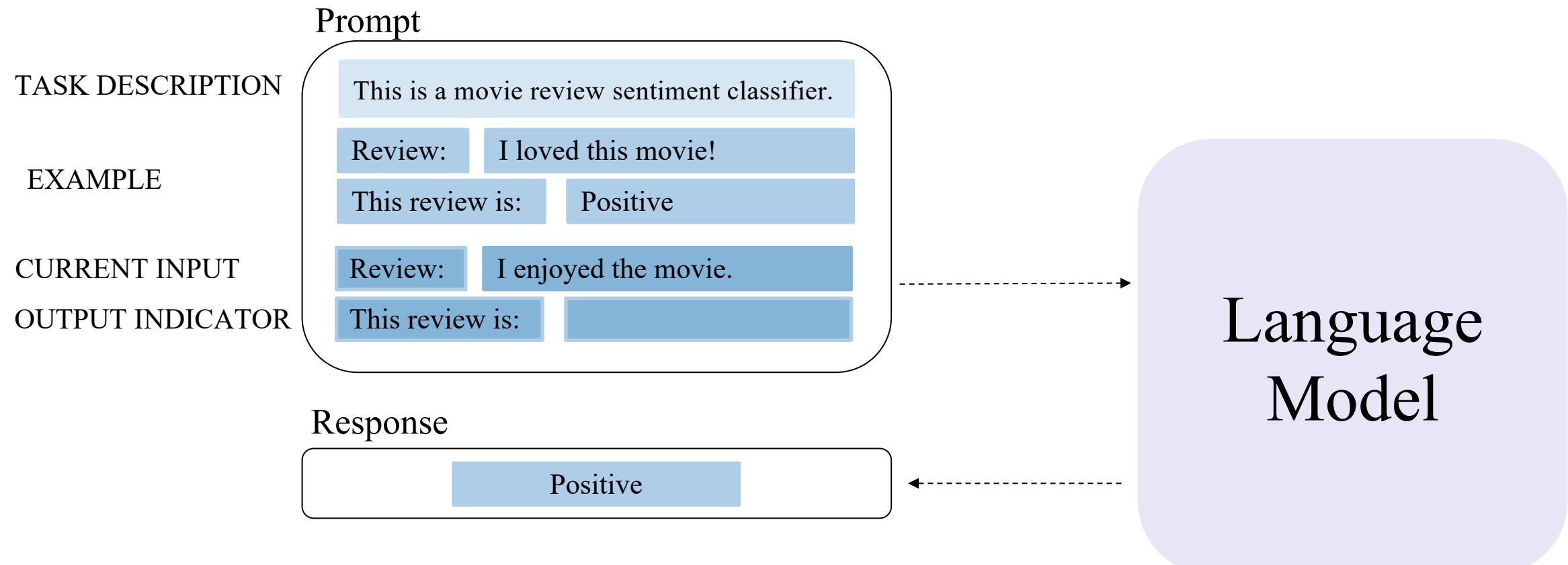


2 – Prompt Engineering



Elements of a Prompt

(2) Describe the task, the general setting and examples (One-shot)

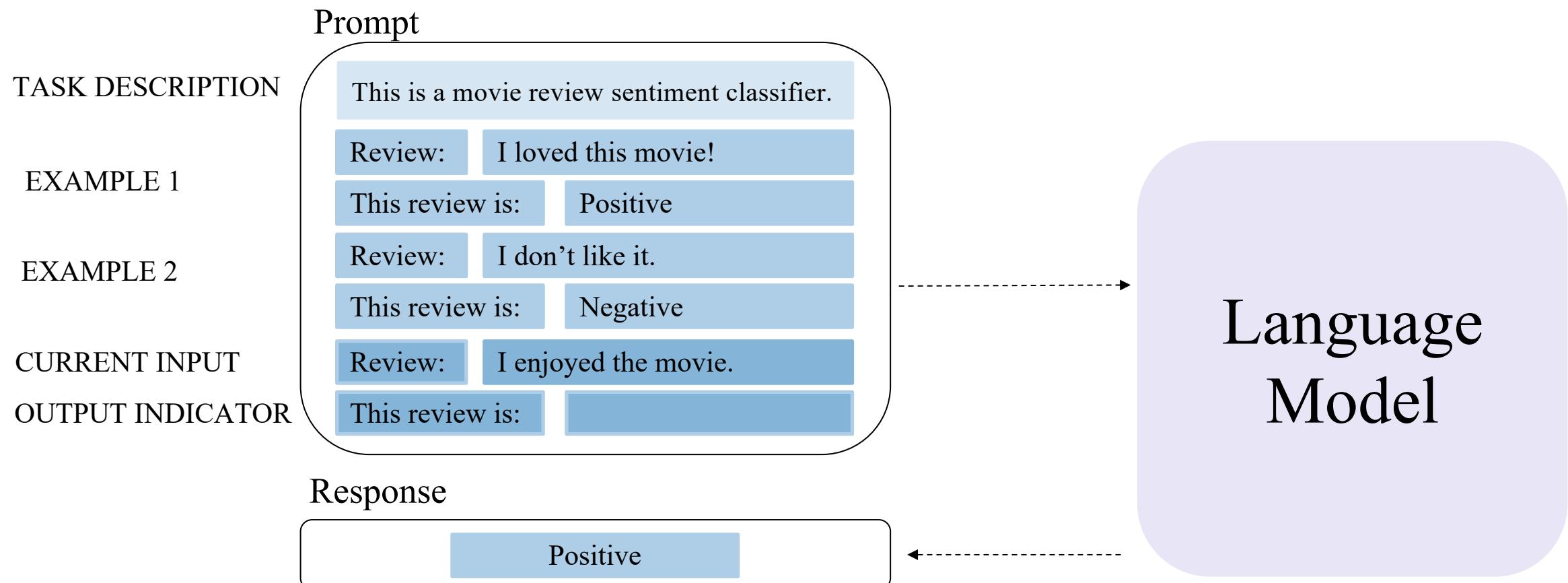


2 – Prompt Engineering



Elements of a Prompt

(2) Describe the task, the general setting and examples (Few-shot)



3 – ChatGPT Applications



Text
Classification.



Information
Extraction



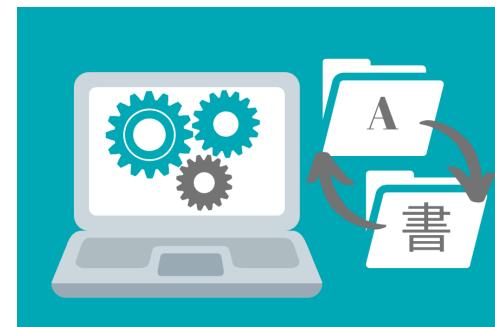
Conversational
Agent



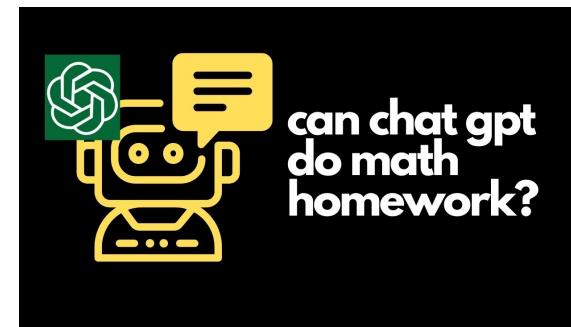
Information
Retrieval



Question Answering
Systems



Machine Translation



Reasoning

```
def calculate_average(numbers):
    """Calculate the average of a list of numbers.

    Args:
        numbers (list): List of numbers

    Returns:
        float: Average of the numbers
    """
    total = 0
    for num in numbers:
        total += num
    return total / len(numbers)
```

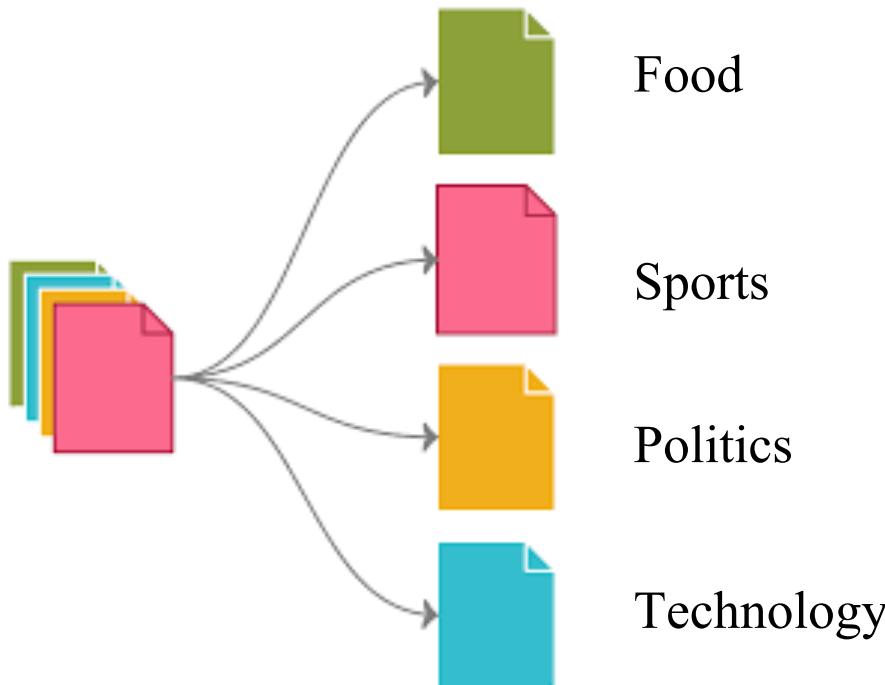
Coding Applications

3 – ChatGPT Applications



3.1. Text Classification

Text Classification is the task of assigning a label or class to a given text



Sentiment Analysis



My experience so far has been fantastic!



Your support team is useless



The product is ok I guess

3 – ChatGPT Applications



3.1. Text Classification

Prompt for Text Classification

TASK DESCRIPTION

Classify the following text.

CURRENT INPUT

Text: I think the food was okay.

OUTPUT INDICATOR

This text is:

Response

Opinion

TASK DESCRIPTION

Classify the following text into neutral, negative, positive.

CURRENT INPUT

Text: I think the food was okay.

OUTPUT INDICATOR

This text is:

Response

Neutral

3 – ChatGPT Applications



3.1. Text Classification

Prompt for Text Classification

- Practice: Modify the prompt to instruct the model to provide an explanation to the answer selected.

TASK DESCRIPTION

Classify the following text into neutral, negative, positive and explain why result is like this.

CURRENT INPUT

Text: I think the food was okay.

OUTPUT INDICATOR

This text is:

Response

Neutral

The phrase "I think the food was okay" does not express a particularly strong opinion, either positively or negatively. It is a neutral statement that does not express a strong emotion one way or the other.

3 – ChatGPT Applications



3.1. Text Classification

Prompt for Text Classification

- Practice: Modify the prompt to instruct the model to classify into neural, negative, positive (automatically)

TASK DESCRIPTION

Classify the following text:

Review: So bad.

This review is: Negative

EXAMPLE

CURRENT INPUT

Text: I think the food was okay.

OUTPUT INDICATOR

This text is:

Response

Neutral

3 – ChatGPT Applications



3.1. Text Classification

Prompt for Multi-Label Text Classification

TASK DESCRIPTION

Classify the following text in multi-label from: service, food, price, ambience, miscellaneous.

CURRENT INPUT

Text: it is very overpriced and not very tasty.

OUTPUT INDICATOR

Response

Price: Overpriced Food: Not Tasty

Label: Service

But the staff was so horrible to us.

Response

Miscellaneous

Label: Food

Each table has a pot of boiling water sunken into its surface and you get platters of thin sliced meats various vegetables and rice and glass noodles.

Food, Miscellaneous

[Source](#)

3 – ChatGPT Applications



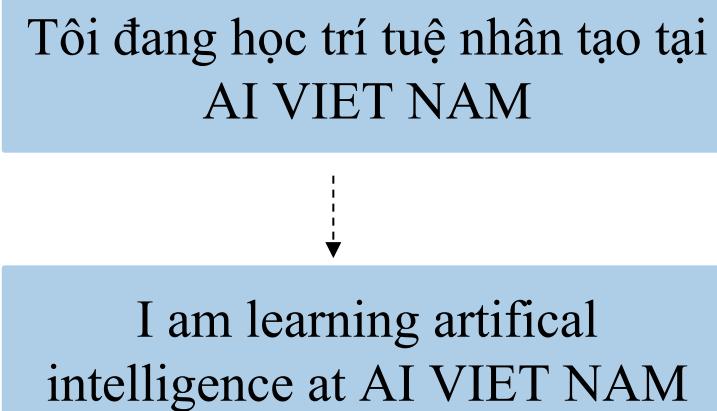
3.2. Machine Translation

- Machine Translation (MT) is a process when a computer software translates text from one language to another without human involvement.

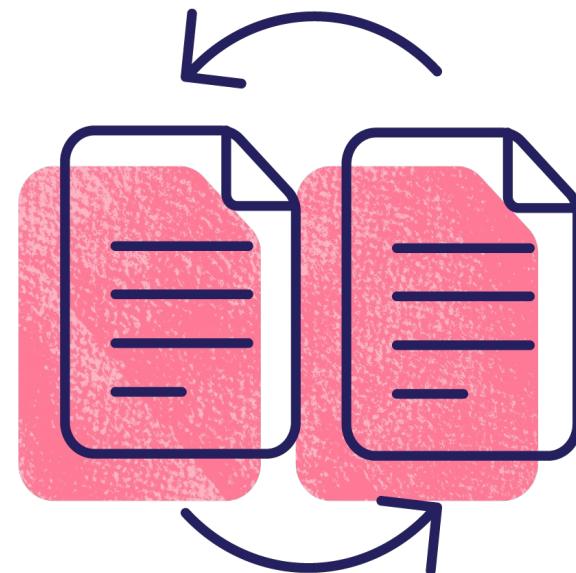
Word Translation



Sentence Translation



Document Translation



3 – ChatGPT Applications



3.2. Machine Translation

➤ Prompts for Machine Translation

Prompt: TP1

TASK DESCRIPTION

Translate these sentence from vi to en language.

CURRENT INPUT

Text

Tôi đang học trí tuệ nhân tạo tại AI VIET NAM

Response

I am learning Artificial Intelligence at AI VIET NAM.

Prompt: TP2

What do these sentences mean in en?

I am studying Artificial Intelligence at AI VIET NAM.

Prompt: TP3

Please provide the en translation for these sentences:

I am learning Artificial Intelligence at AI VIET NAM.

3 – ChatGPT Applications



3.2. Machine Translation

➤ Prompts for Machine Translation

Translation Prompt	
TP1	Translate these sentences from [SRC] to [TGT] :
TP2	What do these sentences mean in [TGT] ?
TP3	Please provide the [TGT] translation for these sentences:

System	BLEU [↑]	ChrF++ [↑]	TER [↓]
Google	31.66	57.09	56.21
DeepL	31.22	56.74	57.84
Tencent	29.69	56.24	57.16
ChatGPT w/ TP1	23.25	53.07	66.03
ChatGPT w/ TP2	24.54	53.05	63.79
ChatGPT w/ TP3	24.73	53.71	62.84

Comparison of different prompts for ChatGPT to perform Chinese-to-English (Zh⇒En) translation

System	De-En		Ro-En		Zh-En	
	⇒	⇐	⇒	⇐	⇒	⇐
Google	45.04	41.16	50.12	46.03	31.66	43.58
DeepL	49.23 _(+9.3%)	41.46 _(+0.7%)	50.61 _(+0.9%)	48.39 _(+5.1%)	31.22 _(-1.3%)	44.31 _(+1.6%)
Tencent	n/a	n/a	n/a	n/a	29.69 _(-6.2%)	46.06 _(+5.6%)
ChatGPT	43.71 _(-2.9%)	38.87 _(-5.5%)	44.95 _(-10.3%)	24.85 _(-46.0%)	24.73 _(-21.8%)	38.27 _(-12.1%)

System	De-Zh		Ro-Zh		De-Ro	
	⇒	⇐	⇒	⇐	⇒	⇐
Google	38.71	21.68	39.05	25.59	33.31	32.27
DeepL	40.46 _(+4.5%)	22.82 _(+5.2%)	38.95 _(-0.2%)	25.39 _(-0.7%)	35.19 _(+5.6%)	34.27 _(+6.1%)
Tencent	40.66 _(+5.0%)	19.44 _(-10.3%)	n/a	n/a	n/a	n/a
ChatGPT	34.46 _(-10.9%)	19.80 _(-8.6%)	30.84 _(-21.0%)	19.17 _(-25.0%)	33.38 _(+0.2%)	29.89 _(-7.3%)

Performance of ChatGPT for multilingual translation

Source: [Is ChatGPT A Good Translator?](#)

3 – ChatGPT Applications



3.2. Machine Translation

➤ Prompts for Machine Translation

TASK DESCRIPTION

Translate these sentence.

Text: Trường học.

Translated: School.

Text Tôi đang học trí tuệ nhân tạo
tại AI VIET NAM

EXAMPLE

Response

I'm learning Artificial
Intelligence at AI VIET NAM.

CURRENT INPUT

Response

EXAMPLE

Text: Tôi thích nó.

Translated: I like it.

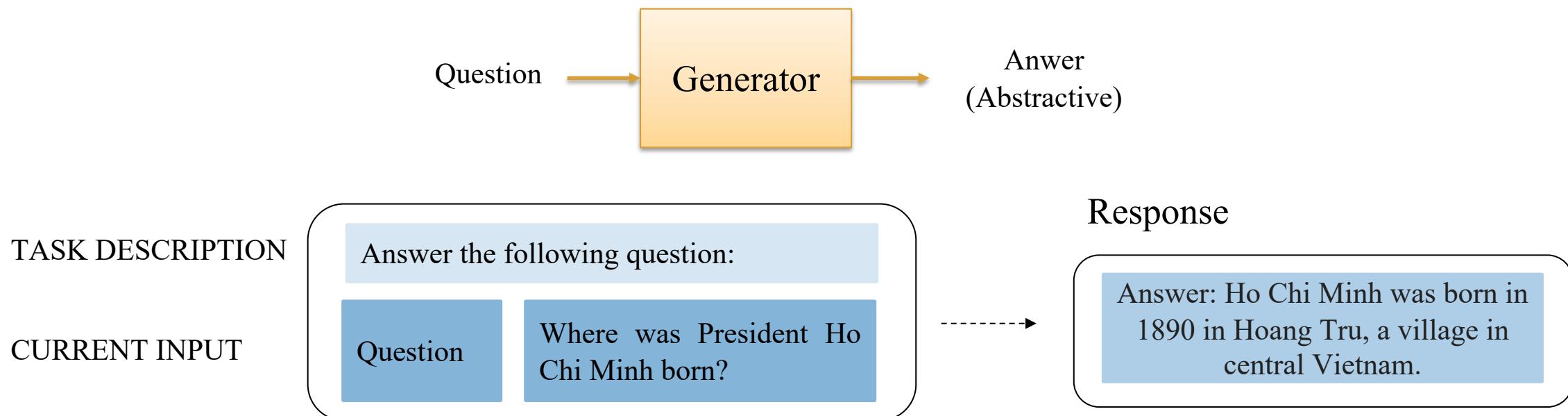
I'm learning Artificial
Intelligence at AI VIET NAM.

3 – ChatGPT Applications

!

3.3. Question Answering

- Question Answering system: Automatically answer questions posed by humans in a natural language

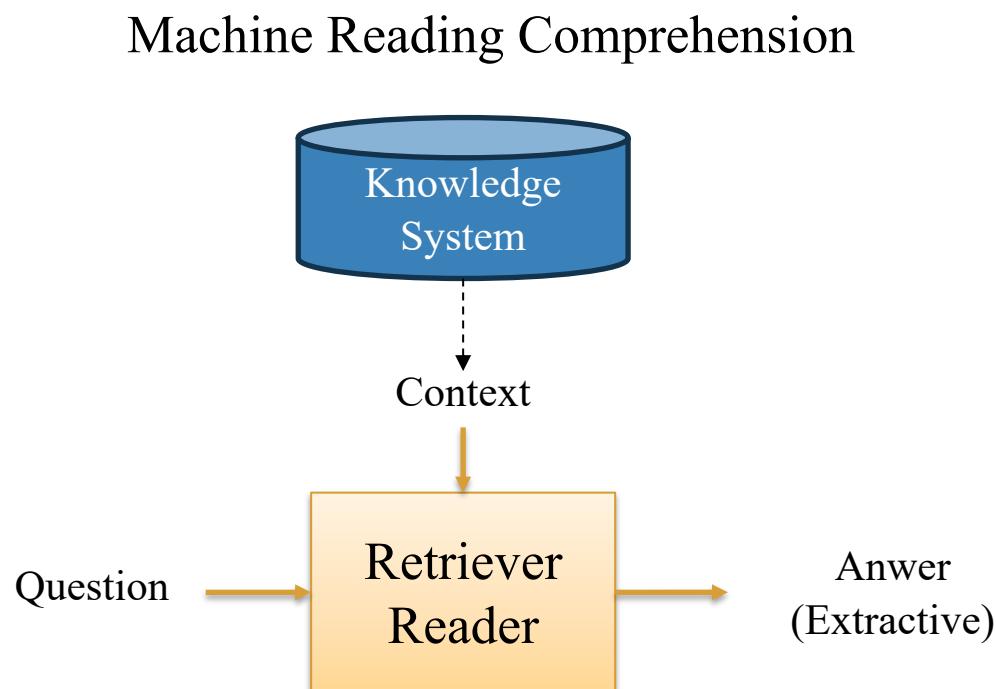


3 – ChatGPT Applications



3.3. Question Answering

- Question Answering system: Automatically answer questions posed by humans in a natural language



Context

Hồ Chí Minh (Nguyễn Sinh Cung; 19 May 1890 – 2 September 1969), commonly known as Uncle Ho (Bác Hồ), President Ho (Hồ Chủ tịch) and by other aliases and sobriquets, was a Vietnamese revolutionary and statesman. He served as Prime Minister of Vietnam from 1945 to 1955, and as President of Vietnam from 1945 until his death in 1969. Ideologically a Marxist–Leninist, he was the Chairman and First Secretary of the Workers' Party of Vietnam. Hồ Chí Minh was born in Nghê An province in the French protectorate of Annam.

Question

Where was President Ho Chi Minh born?

Answer

Nghe An province

3 – ChatGPT Applications



3.3. Question Answering

- Question Answering system: Automatically answer questions posed by humans in a natural language

TASK DESCRIPTION

Answer the following question:

Question

Where was President Ho Chi Minh born?

CURRENT INPUT

Response

Answer: Ho Chi Minh was born in 1890 in Hoang Tru, a village in central Vietnam.

TASK DESCRIPTION

Answer the following question:

Question

Where was President Ho Chi Minh born?

CURRENT INPUT

Context

Hồ Chí Minh (Nguyễn Sinh Cung; 19 May 1890 - 2 September 1969)...

Response

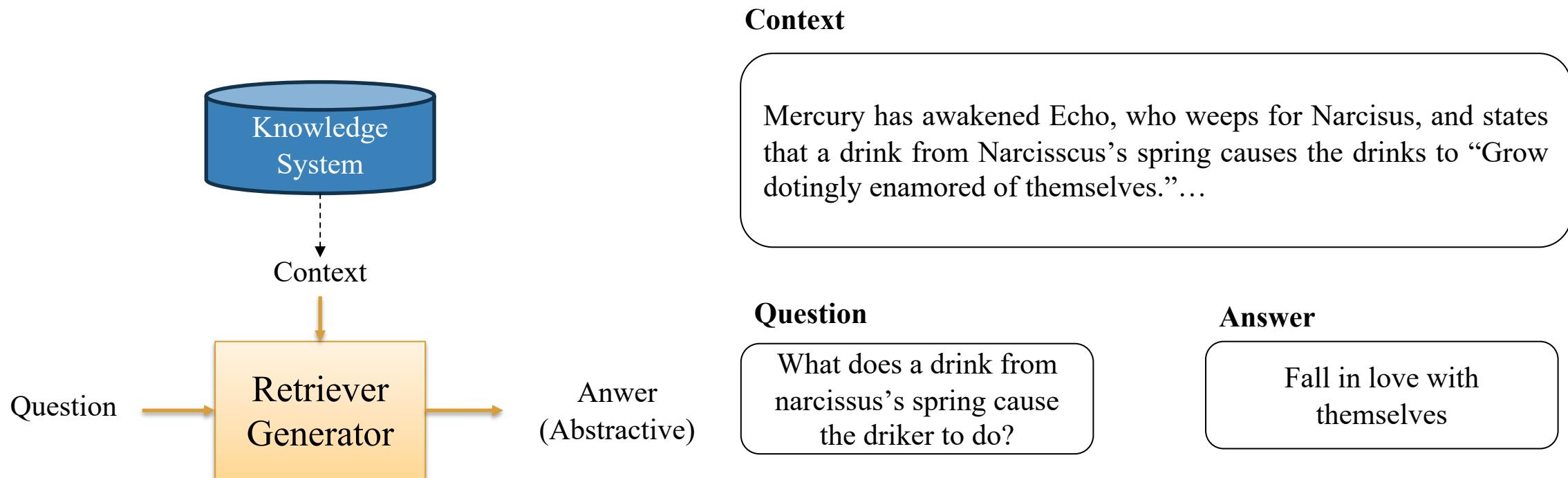
Answer: Nghệ An province, French protectorate of Annam

3 – ChatGPT Applications



3.3. Question Answering

- Question Answering system: Automatically answer questions posed by humans in a natural language



3 – ChatGPT Applications



3.3. Question Answering

- Question Answering system: Automatically answer questions posed by humans in a natural language

TASK DESCRIPTION

Answer the following question:

Question

What does a drink from narcissus's spring cause the drinker to do?

CURRENT INPUT

Response

Drinking from Narcissus's spring is said to cause the drinker to become entranced with their own reflection and unable to look away.

TASK DESCRIPTION

Answer the following question:

Question

What does a drink from narcissus's spring cause the drinker to do?

CURRENT INPUT

Context

Mercury has awakened Echo, who weeps for Narcisus,..

Response

A drink from Narcissus's spring causes the drinker to become infatuated with themselves, growing deeply in love with their own reflection.

3 – ChatGPT Applications

!

3.3. Question Answering

- Question Answering system: Automatically answer questions posed by humans in a natural language

Multiple Choice Question Answering / Yes-No

TASK DESCRIPTION

Answer the following question:

CURRENT INPUT

Question

Question: What does photosynthesis produce that helps plants grow?

Candidate: (A) Water (B) Oxygen
(C) Protein (D) Sugar.

Response

B (Oxygen).

3 – ChatGPT Applications



3.3. Question Answering

- Question Answering system: Automatically answer questions posed by humans in a natural language

TASK DESCRIPTION

Answer the following question:

CURRENT INPUT

Question

How many gold medals does Vietnam have in SEA Games 2023?

Context

Vietnam won 136 gold medals to secure the top spot in the medal tally as the 2023 Southeast Asian (SEA) Games in Cambodia came to the final official day of competition on Tuesday.

Response

Answer: At this time, it is not possible to answer this question as the SEA Games 2023 have not taken place yet.

Response

Vietnam has 136 gold medals at the 2023 Southeast Asian (SEA) Games.

3 – ChatGPT Applications



3.4. Conversation (Role Playing)

- Prompt engineering is instructing the LLM system on how to behave, its intent, and its identity.
- Create a conversational system that's able to generate more technical and scientific responses to questions.

TASK DESCRIPTION

This is a conversation between a customer and a polite, helpful customer service agent.

INPUT INDICATOR

Question of the customer

CURRENT INPUT

Hi, I'd like a refund for the coffee maker I ordered. Would that be possible?

OUTPUT INDICATOR

Response from the customer service agent: Hello. Thank you for reaching out to us. Yes,..



Response

I'm sorry to hear that you're not satisfied with your purchase. We'll be glad to process a refund for you. Please provide us with the order number so we can expedite the process. Thank you for your time.

3 – ChatGPT Applications

!

3.4. Conversation (Role Playing)

- Prompt engineering is instructing the LLM system on how to behave, its intent, and its identity.
- Create a conversational system that's able to generate more technical and scientific responses to questions.

TASK DESCRIPTION

The following is a conversation with an AI research assistant. The assistant tone is technical and scientific.

EXAMPLE

Human: Hello, who are you?

CURRENT INPUT

AI: Greeting! I am an AI research assistant. How can I help you today?

Question of the human: Can you tell me about the creation of blackholes?



Response

Absolutely! Black holes are created when a very large star collapses under its own gravity. The gravity of the star is so strong that it causes the star to collapse inward,...

3 – ChatGPT Applications



3.5. Reasoning

- One of the most difficult tasks for an LLM today is one that requires some form of reasoning
- Some improvements in tasks involving mathematical capabilities

What is $9,000 * 9,000$?

Response

81,000,000

The odd numbers in this group add up to an even number: 15, 32, 5, 13, 82, 7, 1.

Response

The odd numbers in this group add up to 106, which is an even number.

The odd numbers in this group add up to an even number: 15, 32, 5, 13, 82, 7, 1.

Response

The odd numbers add up to 118, which is an even number.

3 – ChatGPT Applications



3.5. Reasoning

- One of the most difficult tasks for an LLM today is one that requires some form of reasoning
- Some improvements in tasks involving mathematical capabilities

The odd numbers in this group add up to an even number: 15, 32, 5, 13, 82, 7, 1.

Solve by breaking the problem into steps. First, identify the odd numbers, add them, and indicate whether the result is odd or even.

Response

Odd numbers: 15, 5, 13, 7, 1
Sum of odd numbers: 41
Result: 41 is an odd number.

3 – ChatGPT Applications



3.6. Coding Applications

➤ Writing code

Write a function that calculates the average of the numbers in Python List.

Response

```
def average(nums):  
    total = 0  
    for num in nums:  
        total += num  
    return total / len(nums)
```

Write a function that calculates the average of the numbers in Python List.

Response

```
def averageList(list):  
    total = 0  
    for num in list:  
        total += num  
    return total / len(list)
```

3 – ChatGPT Applications



3.6. Coding Applications

➤ Comment code

Write a function that calculates the average of the numbers in Python List.
Add to comment for this code

Response

```
def average(nums):
    """Calculate the average of a list of numbers.
    Args: numbers (list): List of numbers
    Returns: float: Average of the numbers
    """
    total = 0
    for num in nums:
        total += num
    return total / len(nums)
```

3 – ChatGPT Applications



3.6. Coding Applications

➤ Debug code

Debug this code:

```
def average(nums):
    total = 0
    for num in nums:
        total += num
    return total * len(nums)
```

Response →

```
def average(nums):
    total = 0
    for num in nums:
        total += num
    return total / len(nums)
```

3 – ChatGPT Applications



3.7. Other Applications

- Education: Course content creation, Grammar and Writing check,...
 - Marketing & SEO: Content creation, Writing product descriptions, Chatbot for customer support,...
 - Government and Politics
 - Healthcare and Medicine
- ...

Source: [ChatGPT: Applications, Opportunities, and Threats](#)

4 – ChatGPT Limitations



4 – ChatGPT Limitations

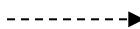


4.1. Lack of knowledge

- Despite this improvement, the model still lacks knowledge of events that occurred after September 2021 and sometimes makes simple reasoning errors.

Answer the following question:

Question How many gold medals does Vietnam have in SEA Games 2023?



Response

Answer: At this time, it is not possible to answer this question as the SEA Games 2023 have not taken place yet.

Answer the following question:

Question When did the world population reach 8 billion people?



Response

The world population reached 8 billion people in March 2021.

Correct Answer: 15 November 2022

4 – ChatGPT Limitations



4.2. Incorrect Answers

- Not completely reliable

Answer the following question:

Question

How many country names start with the letter V?

Response

There are three countries that start with the letter V: Venezuela, Vietnam, and Vanuatu.

Correct Answer: Four countries: Vanuatu, Venezuela, Vatican, VietNam

4 – ChatGPT Limitations



4.3. Reasoning Errors

- Not completely reliable and makes reasoning errors

	SOTA	GPT-4	PaLM	PaLM 2
WinoGrande	87.5 ^a	87.5 ^a ₍₅₎	85.1 ^b ₍₅₎	90.9 ₍₅₎
ARC-C	96.3 ^a	96.3 ^a ₍₂₅₎	88.7 ^c ₍₄₎	95.1 ₍₄₎
DROP	88.4 ^d	80.9 ^a ₍₃₎	70.8 ^b ₍₁₎	85.0 ₍₃₎
StrategyQA	81.6 ^c	-	81.6 ^c ₍₆₎	90.4 ₍₆₎
CSQA	91.2 ^e	-	80.7 ^c ₍₇₎	90.4 ₍₇₎
XCOPA	89.9 ^g	-	89.9 ^g ₍₄₎	94.4 ₍₄₎
BB Hard	65.2 ^f	-	65.2 ^f ₍₃₎	78.1 ₍₃₎

Evaluation on reasoning tasks.

Source: [PaLM 2 Technical Report](#)

Response

Answer the following question:

Question: Count the number of letters 's' that appear last in each word of the following sentence:

Sentence: She sells seashells by the seashore, the shells she sells are surely seashells.

Answer: There are 4 letters 's' that appear last in each word of the sentence: seashore, sells, shells, seashells.

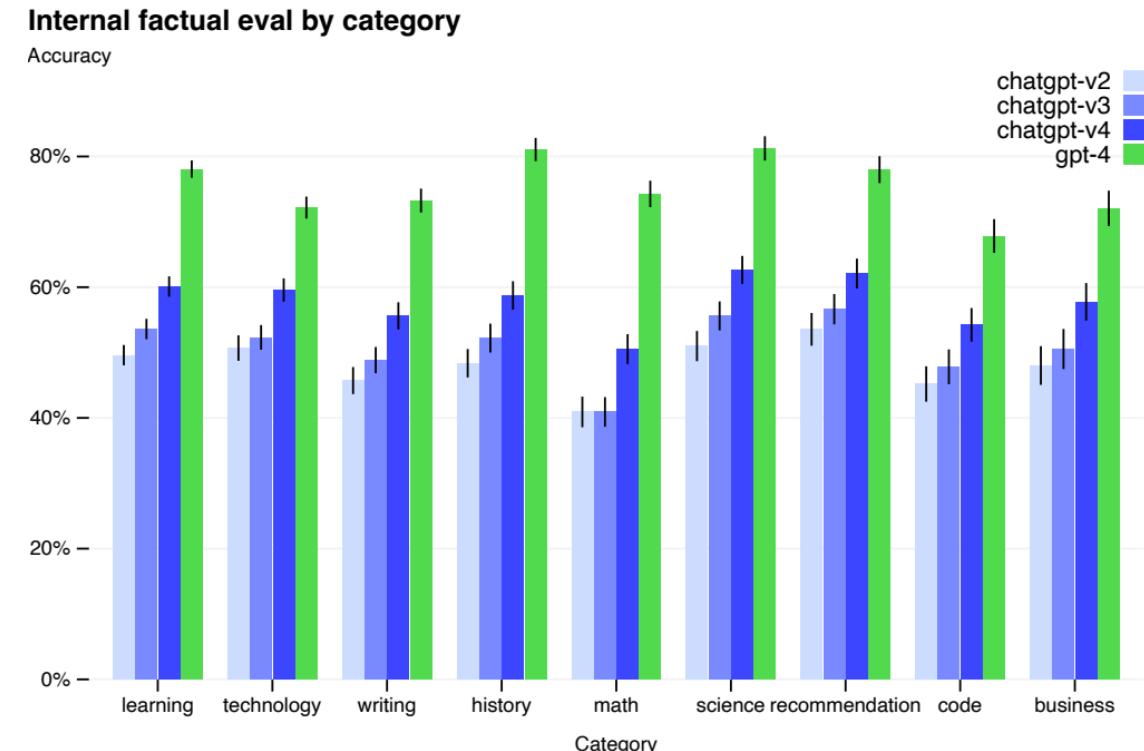
Correct Answer: 5

4 – ChatGPT Limitations



4.4. “Hallucinates” facts

- AI generates each next word based on a sequence of previous words (including the words it has itself previously generated in the current response), causing a cascade of possible hallucination as the response grows longer.
- The default length is fixed at 2048 tokens, while the maximum can be set at 4096 tokens.

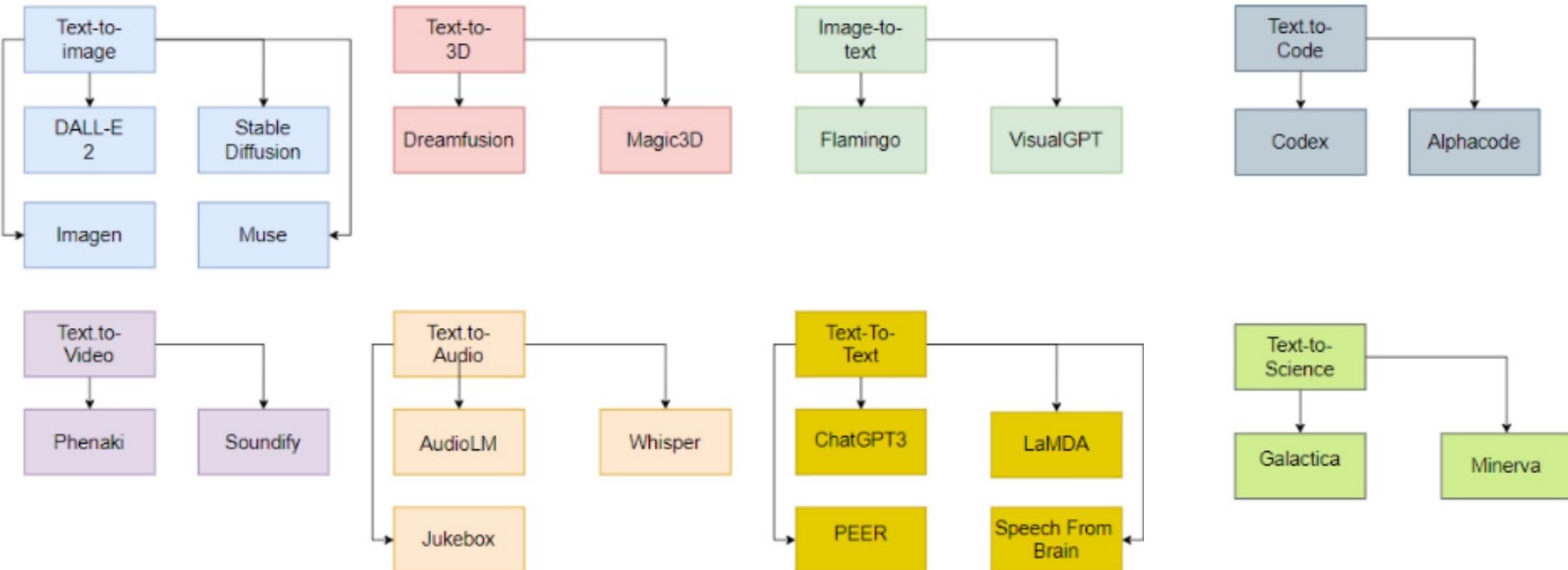


Source: [GPT-4 Technical Report](#)

4 – Generative AI Models



ChatGPT is not all you need



4 – Generative AI Models



4.1. Text-to-Image Models

➤ DALL-E 2 (OpenAI)

CLIP (Contrastive Language-Image Pre-Training)

➤ IMAGEN (Google)

CLIP (Contrastive Language-Image Pre-Training)



Prompt: “A shiba inu wearing a beret and black turtleneck”



Prompt: “A cute corgi lives in a house made out of sushi”

4 – Generative AI Models



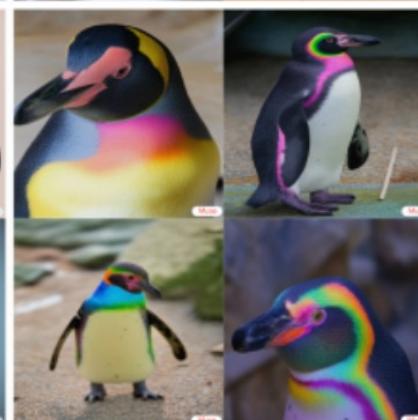
4.1. Text-to-Image Models

- Comparison of generated images by the DALL-E 2, IMAGEN and Muse models

A high contrast portrait of a very happy fuzzy panda dressed as a chef in a high end kitchen making dough. There is a painting of flowers on the wall behind him.



Rainbow coloured penguin.



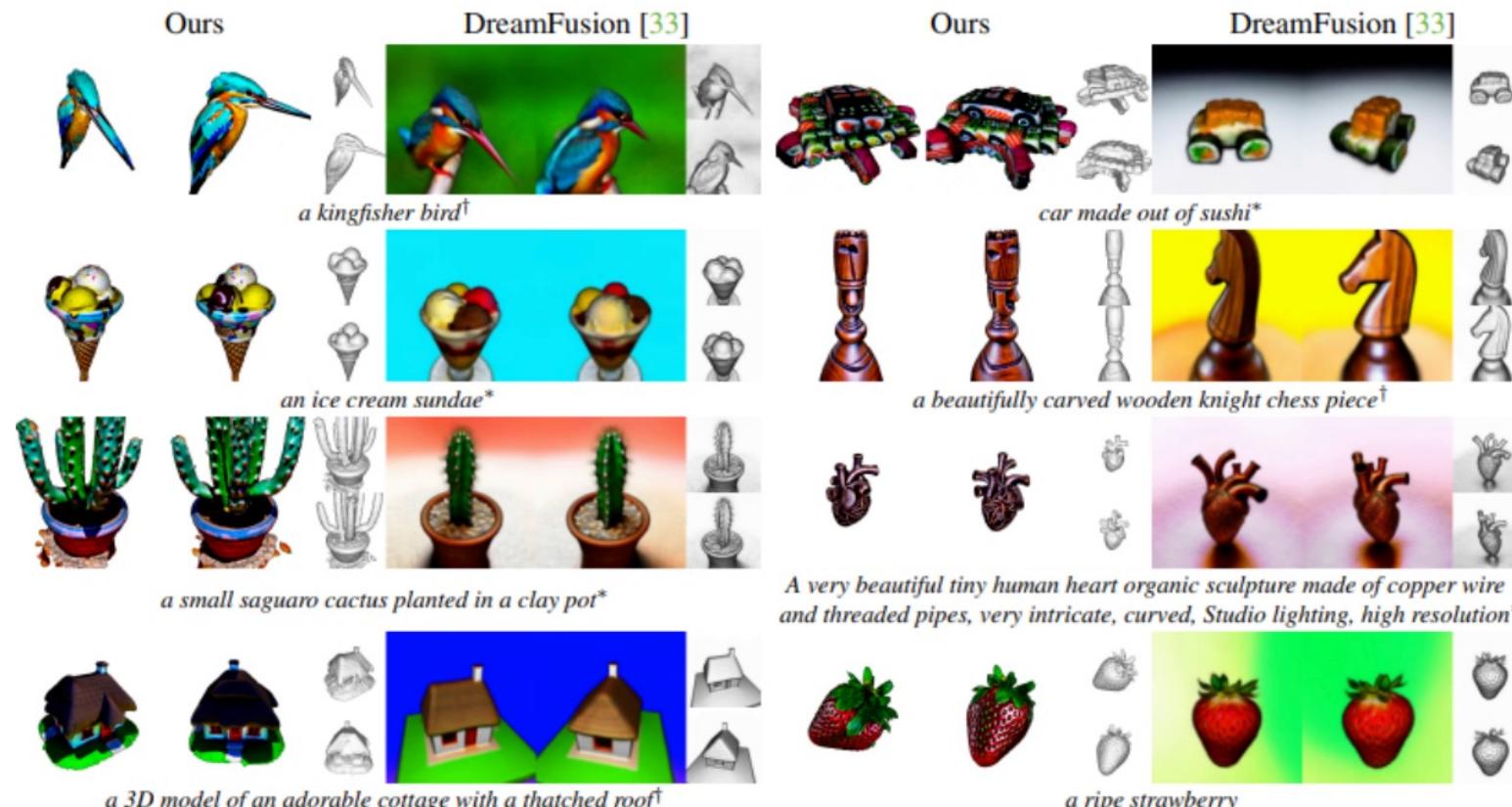
4 – Generative AI Models



4.2. Text-to-3D Models

➤ Magic3D (NVIDIA)

CLIP (Contrastive Language-Image Pre-Training)



4 – Generative AI Models



4.3. Image-to-Text Models

➤ VisualGPT (OpenAI)



GT: the lady is sitting on the wood bench

Ours	a	woman	sitting	on	a	bench	in	a	park
attention	0.7	0.78	0.82	0.76	0.8	0.96	0.8	0.69	0.85



GT: a laptop with a keyboard and mouse are on this desk

Ours	a	laptop	sitting	on	a	desk	with	a	mouse
attention	0.7	0.78	0.81	0.7	0.7	0.92	0.85	0.64	0.76



GT: a cat is sitting in front of a television

Ours	a	cat	is	sitting	in	front	of	a	television
attention	0.8	0.86	0.8	0.83	0.7	0.72	0.6	0.71	0.93



GT: a number of people sitting on a snowy surface with skis

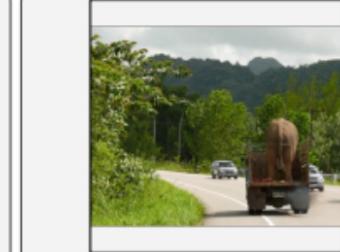
Ours	a	couple	of	people	sitting	on	a	snowy	surface
attention	0.8	0.87	0.71	0.85	0.91	0.76	0.71	0.94	0.95

➤ Flamingo (DeepMind)

Input Prompt



Question: What do you think the capacities of these are?
Answer:



Question: What is odd about this image? Explain why it is unusual.
Answer:



Question: What country is this? Why do you think so?
Answer:

Completion

The floppy disk is 1.44MB and the CD is 700MB.

The image is odd because the elephant is in the back of the truck. It is unusual because elephants are not usually transported in the back of a truck.

It is Canada. I think so because the flag is the Canadian flag.

4 – Generative AI Models



4.4. Text-to-Video Models

➤ Phenaki (Google)

1st prompt: "A photorealistic teddy bear is swimming in the ocean at San Francisco"



2nd prompt: "The teddy bear goes under water"



3rd prompt: "The teddy bear keeps swimming under the water with colorful fishes"



Sequence of images created by the Phenaki model given different prompts

Given Image



Given Image



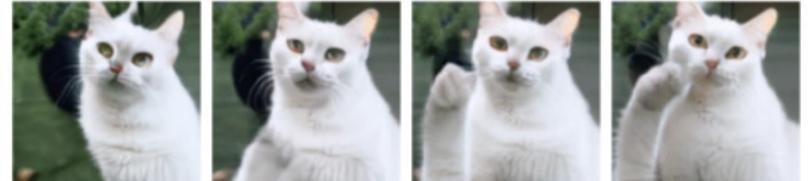
Given Image



Prompt: "Camera zooms quickly into the eye of the cat"



Prompt: "A white cat touches the camera with the paw"



Prompt: "A white cat yawns loudly"



Sequences of images created by the Phenaki model given an image and the prompt.

4 – Generative AI Models



4.5. Text-to-Text Models

Model	Release Time	Size (B)	Base Model	Adaptation IT	Adaptation RLHF	Pre-train Data Scale	Latest Data Timestamp	Hardware (GPUs / TPUs)	Training Time	Evaluation ICL	Evaluation CoT
T5 [73]	Oct-2019	11	-	-	-	1T tokens	Apr-2019	1024 TPU v3	-	✓	-
mT5 [74]	Oct-2020	13	-	-	-	1T tokens	-	-	-	✓	-
PanGu- α [75]	Apr-2021	13*	-	-	-	1.1TB	-	2048 Ascend 910	-	✓	-
CPM-2 [76]	Jun-2021	198	-	-	-	2.6TB	-	-	-	-	-
T0 [28]	Oct-2021	11	T5	✓	-	-	-	512 TPU v3	27 h	✓	-
CodeGen [77]	Mar-2022	16	-	-	-	577B tokens	-	-	-	✓	-
GPT-NeoX-20B [78]	Apr-2022	20	-	-	-	825GB	-	96 40G A100	-	✓	-
Tk-Instruct [79]	Apr-2022	11	T5	✓	-	-	-	256 TPU v3	4 h	✓	-
UL2 [80]	May-2022	20	-	-	-	1T tokens	Apr-2019	512 TPU v4	-	✓	✓
OPT [81]	May-2022	175	-	-	-	180B tokens	-	992 80G A100	-	✓	-
NLLB [82]	Jul-2022	54.5	-	-	-	-	-	-	-	✓	-
Publicly Available	GLM [83]	Oct-2022	130	-	-	400B tokens	-	768 40G A100	60 d	✓	-
	Flan-T5 [64]	Oct-2022	11	T5	✓	-	-	-	-	✓	✓
	BLOOM [69]	Nov-2022	176	-	-	366B tokens	-	384 80G A100	105 d	✓	-
	mT0 [84]	Nov-2022	13	mT5	✓	-	-	-	-	✓	-
	Galactica [35]	Nov-2022	120	-	-	106B tokens	-	-	-	✓	✓
	BLOOMZ [84]	Nov-2022	176	BLOOM	✓	-	-	-	-	✓	-
	OPT-IML [85]	Dec-2022	175	OPT	✓	-	-	128 40G A100	-	✓	✓
	LLaMA [57]	Feb-2023	65	-	-	1.4T tokens	-	2048 80G A100	21 d	✓	-
	CodeGeeX [86]	Sep-2022	13	-	-	850B tokens	-	1536 Ascend 910	60 d	✓	-
	Pythia [87]	Apr-2023	12	-	-	300B tokens	-	256 40G A100	-	✓	-

Source: [A survey of Large Language Models](#)

4 – Generative AI Models



4.5. Text-to-Text Models

Closed Source	GPT-3 [55]	May-2020	175	-	-	-	300B tokens	-	-	-	✓	-
	GShard [88]	Jun-2020	600	-	-	-	1T tokens	-	2048 TPU v3	4 d	-	-
	Codex [89]	Jul-2021	12	GPT-3	-	-	100B tokens	May-2020	-	-	✓	-
	ERNIE 3.0 [90]	Jul-2021	10	-	-	-	375B tokens	-	384 V100	-	✓	-
	Jurassic-1 [91]	Aug-2021	178	-	-	-	300B tokens	-	800 GPU	-	✓	-
	HyperCLOVA [92]	Sep-2021	82	-	-	-	300B tokens	-	1024 A100	13.4 d	✓	-
	FLAN [62]	Sep-2021	137	LaMDA-PT	✓	-	-	-	128 TPU v3	60 h	✓	-
	Yuan 1.0 [93]	Oct-2021	245	-	-	-	180B tokens	-	2128 GPU	-	✓	-
	Anthropic [94]	Dec-2021	52	-	-	-	400B tokens	-	-	-	✓	-
	WebGPT [72]	Dec-2021	175	GPT-3	-	✓	-	-	-	-	✓	-
	Gopher [59]	Dec-2021	280	-	-	-	300B tokens	-	4096 TPU v3	920 h	✓	-
	ERNIE 3.0 Titan [95]	Dec-2021	260	-	-	-	-	-	-	-	✓	-
	GLaM [96]	Dec-2021	1200	-	-	-	280B tokens	-	1024 TPU v4	574 h	✓	-
	LaMDA [63]	Jan-2022	137	-	-	-	768B tokens	-	1024 TPU v3	57.7 d	-	-
	MT-NLG [97]	Jan-2022	530	-	-	-	270B tokens	-	4480 80G A100	-	✓	-
	AlphaCode [98]	Feb-2022	41	-	-	-	967B tokens	Jul-2021	-	-	-	-
	InstructGPT [61]	Mar-2022	175	GPT-3	✓	✓	-	-	-	-	✓	-
	Chinchilla [34]	Mar-2022	70	-	-	-	1.4T tokens	-	-	-	✓	-
	PaLM [56]	Apr-2022	540	-	-	-	780B tokens	-	6144 TPU v4	-	✓	✓
	AlexaTM [99]	Aug-2022	20	-	-	-	1.3T tokens	-	128 A100	120 d	✓	✓
	Sparrow [100]	Sep-2022	70	-	-	✓	-	-	64 TPU v3	-	✓	-
	WeLM [101]	Sep-2022	10	-	-	-	300B tokens	-	128 A100 40G	24 d	✓	-
	U-PaLM [102]	Oct-2022	540	PaLM	-	-	-	-	512 TPU v4	5 d	✓	✓
	Flan-PaLM [64]	Oct-2022	540	PaLM	✓	-	-	-	512 TPU v4	37 h	✓	✓
	Flan-U-PaLM [64]	Oct-2022	540	U-PaLM	✓	-	-	-	-	-	✓	✓
	GPT-4 [46]	Mar-2023	-	-	✓	✓	-	-	-	-	✓	✓
	PanGu- Σ [103]	Mar-2023	1085	PanGu- α	-	-	329B tokens	-	512 Ascend 910	100 d	✓	-

Source: [A survey of Large Language Models](#)

4 – Generative AI Models



4.6. Other Models

Text-to-Audio Models

- AudioLM (Google)
- Jukebox (OpenAI)
- Whisper (OpenAI) – Audio-to-Speech

Text-to-Code Models

- Codex (OpenAI)
- Alphacode

Thanks!

Any questions?