



---

# Prompting 101: Fundamentals & Advanced Techniques

## Introduction to Generative AI

Associate Professor Peerapon Vateekul, Ph.D.  
[peerapon.v@chula.ac.th](mailto:peerapon.v@chula.ac.th)

Department of Computer Engineering,  
Faculty of Engineering, Chulalongkorn University  
Credit to: TA.Prommin & TA.Thanaphat

---

# Assoc. Prof. Peerapon Vateekul (Aj.Pop)

- Associate professor at Department of Computer Engineering since 2012 specialized in data analytics including data mining/machine learning, deep learning, big data architecture, data science
  - **Specialized in applying AI/ML/DL into various domains & design Big Data platform (DB/DW)**
  - Course: Data Science and Data Engineering, Data Warehousing, Natural Language Processing, etc.
- DATA MIND Lab is a part of Machine Intelligence and Knowledge Discovery Lab (MIND)
- Visiting professor at Osaka University (2014)
- Chula MOOC: Data Analytics and Big Data, Practical Data Analytics Using Rapid Miner
- Certified SAS instructor
- NVIDIA AI Technology Center (NVAITC)



# ทนายสาย AI ChatGPT

นำกฎหมายให้เป็นเรื่องง่ายด้วย



ศูนย์บริการเรียนรู้  
LEARNING INNOVATION CENTER  
Chulalongkorn University

CHULA ENGINEERING  
Foundation toward Innovation

ห้อง 209 ชั้น 2 (อาคาร 3) คณะวิศวกรรมศาสตร์ จุฬาฯ วันที่ 19 ก.ค. 2567 เวลา 12:00-13:00 น.

## หมายเหตุ (หมาย AI)



ศ. ดร.พีรวัฒน์ เวียรากoot  
คณะวิศวกรรมศาสตร์  
จุฬาลงกรณ์มหาวิทยาลัย



## คำอุปนิสัยที่ใช้ข้อมูลเท็จ: จริง = **HALLUCINATION** (ภาพหลอน)

BIAS จากข้อมูล  
ที่มีมากเกินไป  
ข้อมูลที่ไม่ถูกต้อง<sup>\*</sup>  
การเรียน  
LLM  
ไม่เข้มงวด

รูปแบบ  
ค่าความ  
ไม่เข้มงวด

## ทางแก้ภาพหลอน

- ✓ สอนตัวเองข้อมูลเฉพาะเจาะจง
- ✓ ปรับแต่ง LLM ให้เข้มงวด
- ✓ ปรับแต่งคำถ้า

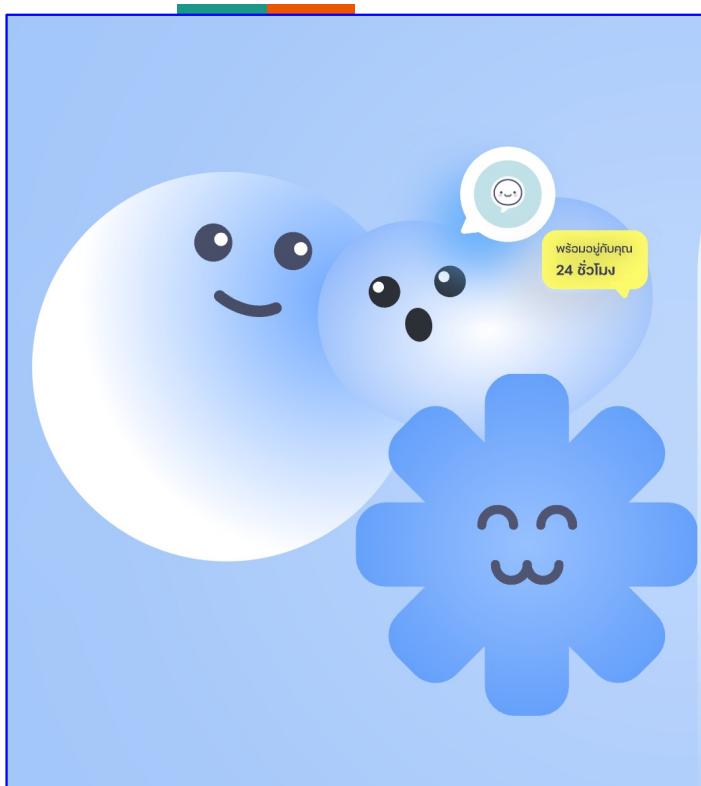
## การสร้าง หมายเหตุ

- สร้างฐานข้อมูลส่วนตัว
- เลือก LLM ที่เข้มงวด เช่น GPT-4
- สอน LLM (FINE-TUNE)
- เพิ่มความน่าเชื่อถือด้วย RAG
- ตั้งค่าความแบบนักกฎหมาย
- ปรับแต่ง LLM เล็กๆ น้อยๆ

LINE  
CHATBOT



# Dmind-LLM: Mental Health Support Chatbot



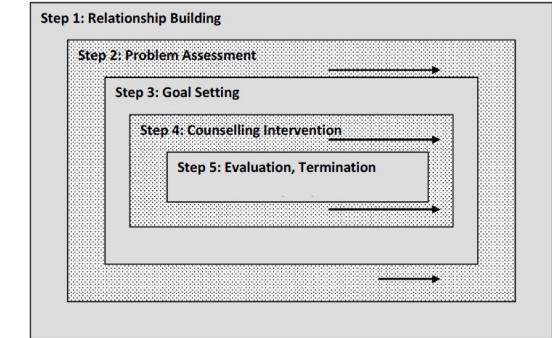
พร้อมบุกบ้าน  
24 ชั่วโมง

## คุณกำลังรู้สึกไม่ สบายใจ อุย...

เปล่า?

มาพูดคุยหรือปรึกษาเรื่อง  
สหายใจกับ DMIND Chatbot.  
สามารถพิมพ์ข้อความโดย  
คุณได้เหมือนคุยกับ  
กันเพื่อน เราพร้อมจะช่วย  
รู้สึกดีและสนับยใจขึ้น

เริ่มต้นพูดคุย



Dmind Chatbot  
Connected

สวัสดีค่ะฉันคือ DMind Chatbot สามารถช่วยคุณได้อย่างไรบ้าง?

▶ พิจารณา

สวัสดีครับ เครียดต้องมีสอนวันพุธนี้ครับ

▶ พิจารณา

สวัสดีค่ะ ติมายด์อินดี้ที่ได้พบกันอีกครั้งนึงค่ะคุณเจย์ กำลังเครียดเรื่องการสอนที่จะมีในวันพุธนี้ ติมายด์เข้าใจความรู้สึกนั้นค่ะ อย่างให้คุณเจย์ลองเล่าให้ฟังได้ไหมคะว่าเป็นการสอนเกี่ยวกับอะไร?

▶ พิจารณา

สอนเกี่ยวกับ Generative AI ครับ

▶ พิจารณา

เข้าใจแล้วค่ะ การสอนเกี่ยวกับ Generative AI เป็นหัวข้อที่น่าสนใจ

พิมพ์ข้อความของคุณที่นี่...

▶ ตอบ

▶ ลบ

# QA for Water Management at HII

## Text-Based QA & Image Search

กรุณาดำเนินการนี้เพื่อแนบมาและคาดการณ์สี่ปีต่อไป

**HIGHLIGHT**

สืบค้นที่ลับลึกว่าเว็บไซต์ที่พัฒนาและคาดการณ์สี่ปีต่อไปนี้  
มีความน่าเชื่อถือมากแค่ไหน ทำให้เราสามารถประเมินความเสี่ยงของภัยธรรมชาติได้ดีขึ้น  
และคาดเดาผลลัพธ์ที่อาจเกิดขึ้นในอนาคต

ภาพที่แสดงถึงความเสี่ยงต่อภัยธรรมชาติที่คาดการณ์ไว้

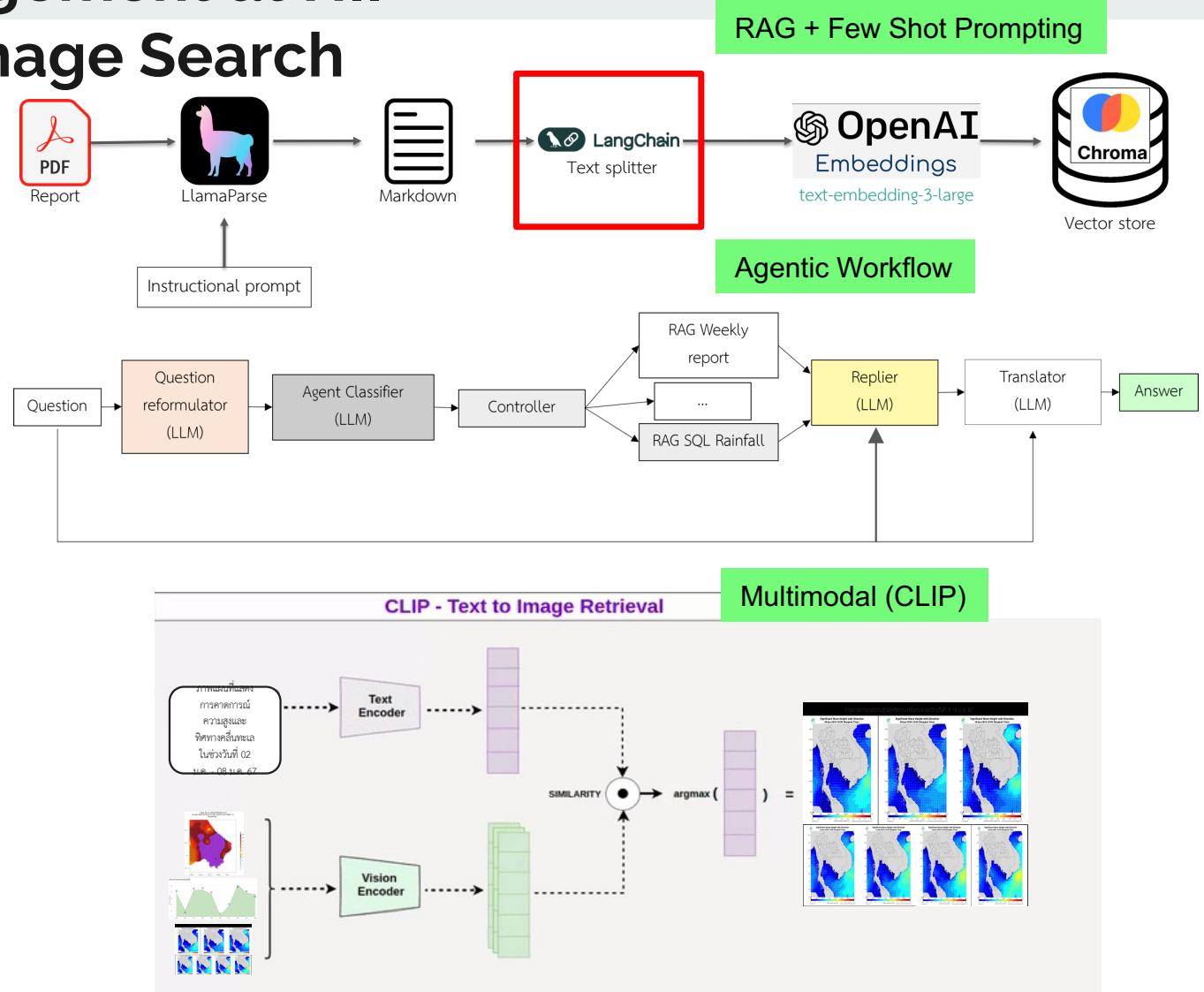
Enter text to search image related to your keyword (please identify date of the documents to get better result).

แผนที่แสดงการคาดการณ์ความเสี่ยงและพื้นที่ทางด้านประเทศไทยช่วงวันที่ 16 ม.ค. 2022

Submit

ดูเพิ่มเติม

Image Caption: แผนที่แสดงตัวหารและความเสี่ยงของภัยธรรมชาติในประเทศไทยเดือนมกราคม.  
Image Source: 1046\_20220117\_Predict\_SendRid.pdf\_page\_17\_img\_11.png



---

## วัตถุประสงค์ของหลักสูตร

- เข้าใจหลักการในการเขียน prompt ที่มีประสิทธิภาพ เพื่อใช้กับโมเดลภาษา (LLMs) เช่น ChatGPT
- ฝึกทักษะเชิงปฏิบัติในการประยุกต์ใช้ prompting กับงานด้านกฎหมาย เช่น งานเอกสาร, การสืบค้น, การสรุปข้อเท็จจริง และการวิเคราะห์คำพิพากษา



# หัวข้อเนื้อหาหลักสูตร

## 1. บทนำสู่ Generative AI (9:00 - 9:45)

Generative AI คืออะไร?

ความสามารถและข้อจำกัดของโมเดลภาษา LLMs

ความสำคัญของการเขียน prompt ที่ดีในบริบทงานกฎหมาย

## 2. พื้นฐานของ Prompting (9:45 - 10:30; Lab1 + Lab2)

ประเภทของ prompt: zero-shot, few-shot, instruction-based

การสมบทบาท (Role prompting): เช่น “คุณเป็นนิติกรประจำหน่วยงานราชการ...”

การทำหนดรูปแบบผลลัพธ์: ให้ตอบในรูปแบบ memo, ตารางเปรียบเทียบ, หรือ bullet points

--- 10.30 - 10.45: Break ---

## 3. เทคนิคการเขียน Prompt ให้ตอบโจทย์งานนิติการและคดี (10:45 - 11:15; Lab3)

การใช้ Chain-of-Thought prompting เพื่อให้โมเดลอธิบายเหตุผลทางกฎหมาย

การใช้ prompt แบบ iterative เพื่อปรับปรุงคำตอบให้ตรงตามความต้องการ

การตั้งคำถามหลายรอบ (multi-turn conversation) ในกรณีที่ต้องการเจาะลึกประเด็นกฎหมาย

## 4. เทคนิคขั้นสูงและการทำงานกับเอกสาร (11:15 - 12:00; Lab 4)

การสร้าง prompt template

การประยุกต์ใช้ Custom instructions ให้โมเดลเข้าใจบทบาท

การใช้ Retrieval-Augmented Generation (RAG) เพื่อสืบค้นข้อมูลจากคำพิพากษาหรือเอกสารทางกฎหมายที่อัปโหลดเข้าไป

--- 12.00 - 13.00: Lunch Break ---

## 5. ตัวอย่าง Use Case สำหรับกลุ่มงานนิติการและคดี และสรุปภาพรวมเนื้อหาที่เรียนทั้งหมด

(13:00 - 14:30) Lab5: Custom Prompt + Lab6: Legal Prompting

--- 14.30 - 14.45: Break ---

(14:45 - 16:00) Lab 7: Gen AI Tools: Legal Chatbot, Notebook LM, Google AI Studio

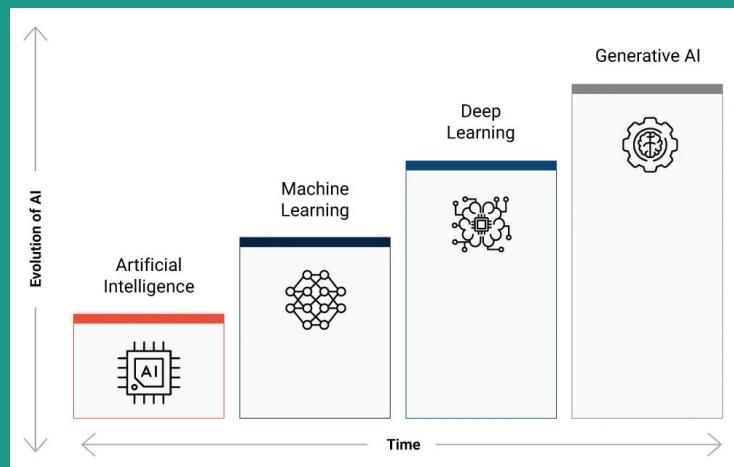


## **Outlines: Introduction to Generative AI**

- Generative AI
- AI Assistants
- Conclusion

# Generative AI: LLM

Credit to Asst. Prof.Ekapol's slide





## ChatGPT Surges in Popularity: Now Boasting Over 200 Million Weekly Users

Onome · September 2, 2024

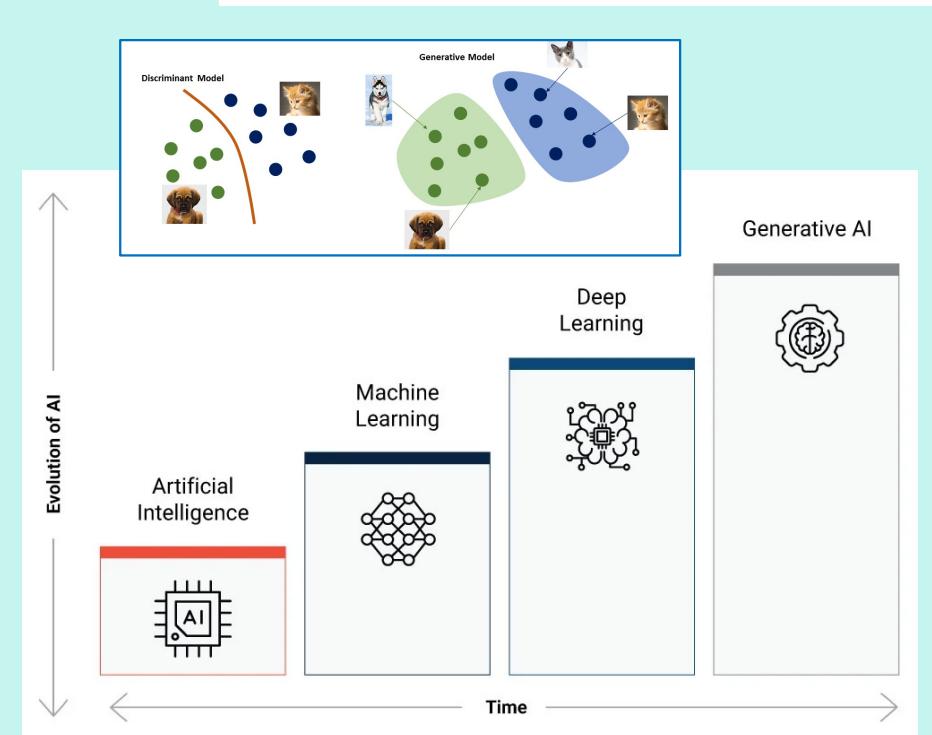
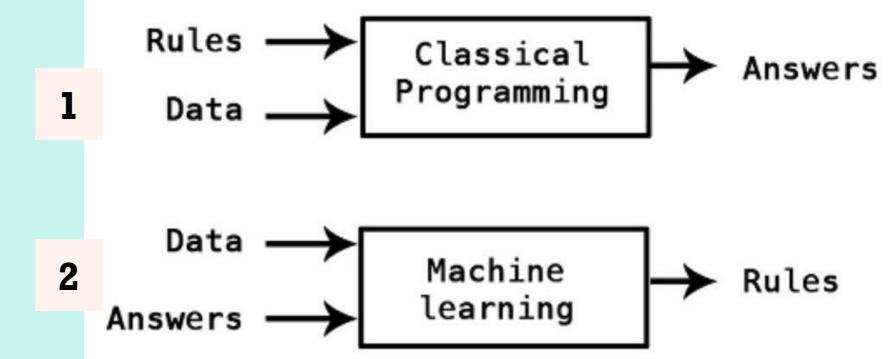
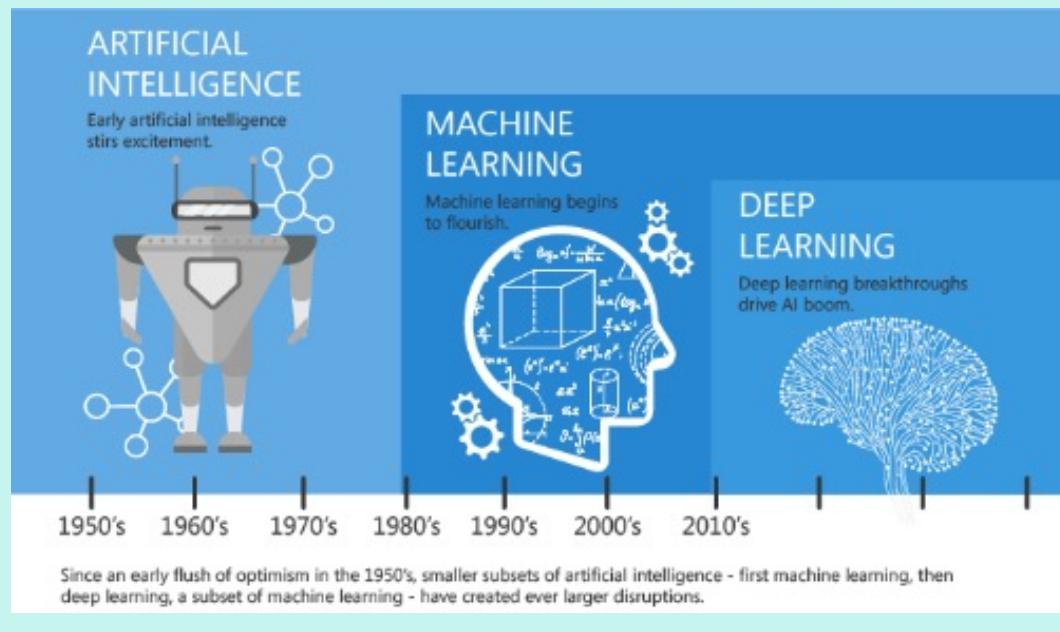
AI NEWS 3 MINS READ



<https://autogpt.net/chatgpt-surges-in-popularity-now-boasting-over-200-million-weekly-users/>

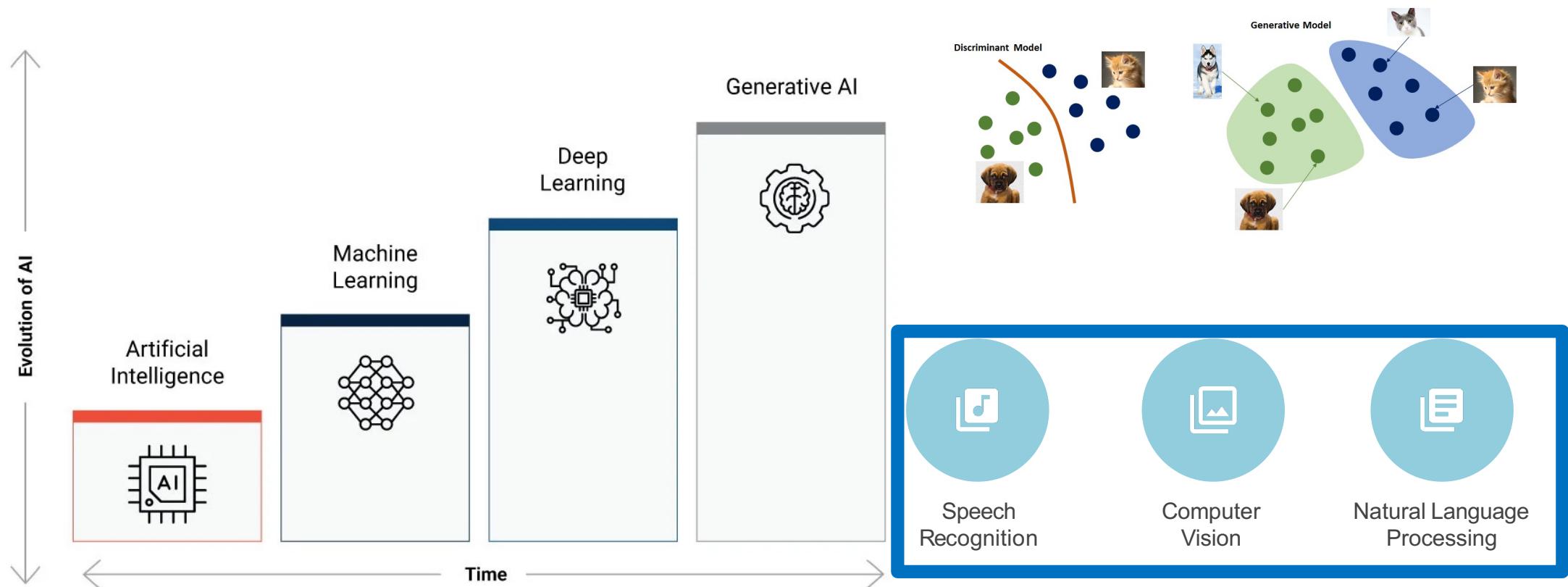
# AI = Automation

- 1) Rule-based AI
- 2) Machine Learning (ML)



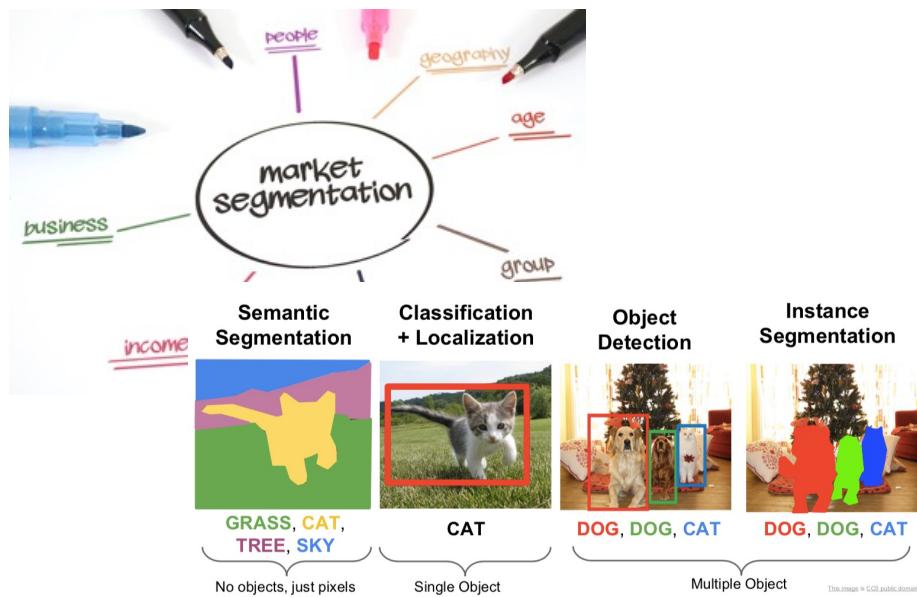
<https://mc.ai/machine-learning-basics-artificial-intelligence-machine-learning-and-deep-learning/>

# Recent Research: DL & Generative AI

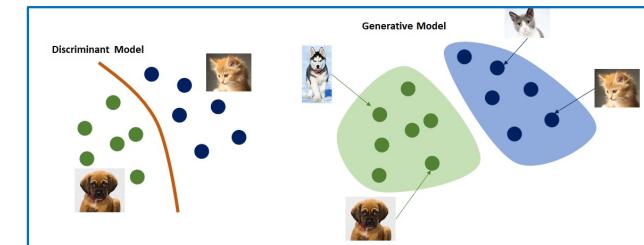


# Predictive AI vs. Generative AI

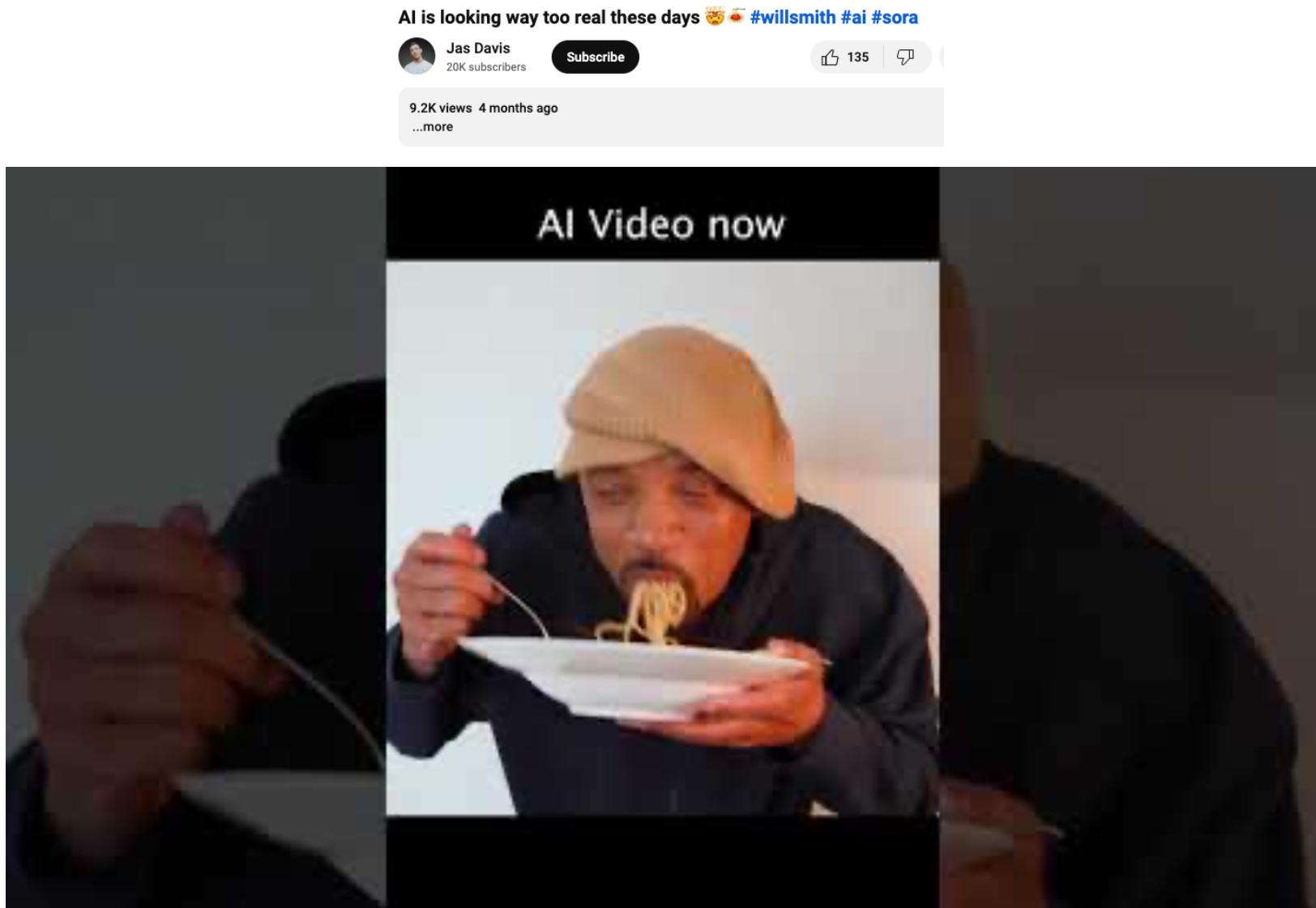
- Predictive AI
  - Predict, Forecast, Detect
  - Cluster, Classify, Segment



- Generative
  - Generate, Generalize



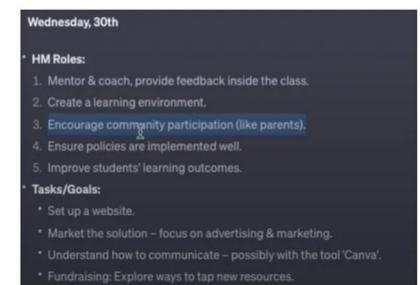
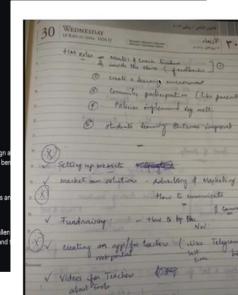
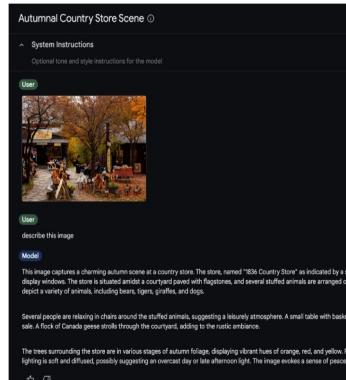
<https://www.youtube.com/watch?v=XQr4Xklqzw8>



<https://openai.com/index/sora/>

# What can be generated?

- Generate image
  - Midjourney by the Midjourney team
  - Imagen 3 by Google
  - DALLE 3 by OpenAI
- Generate text
  - ChatGPT by OpenAI
  - Claude by Anthropic
  - Gemini by Google
  - DeepSeek by the Deepseek team





## ChatGPT

- ChatGPT was launched by OpenAI on 30 Nov 2022.
- ChatGPT is a large language model (LLM) for conversational AI applications.
- Generates human-like text and performs NLP tasks.
- Scalable and flexible for various use cases.
- ChatGPT didn't enclose the details.



OpenAI

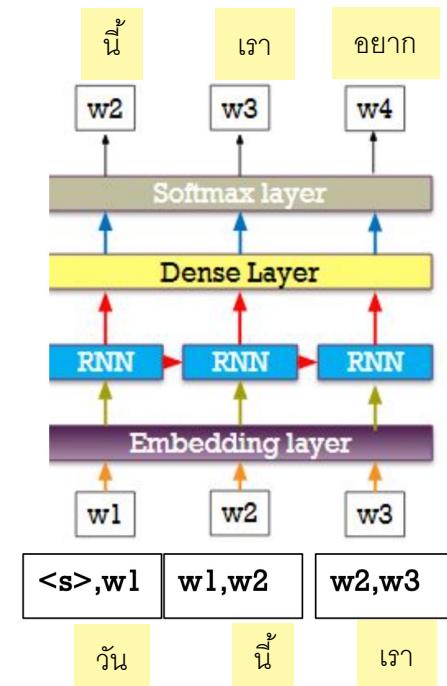


## Language Model (LM)

- It is the model that aims to predict next word based on the given previous words.
- So, the model can understand grammar & context.

วันนี่เราอยากจะกินข้าวมัน \_\_\_\_\_

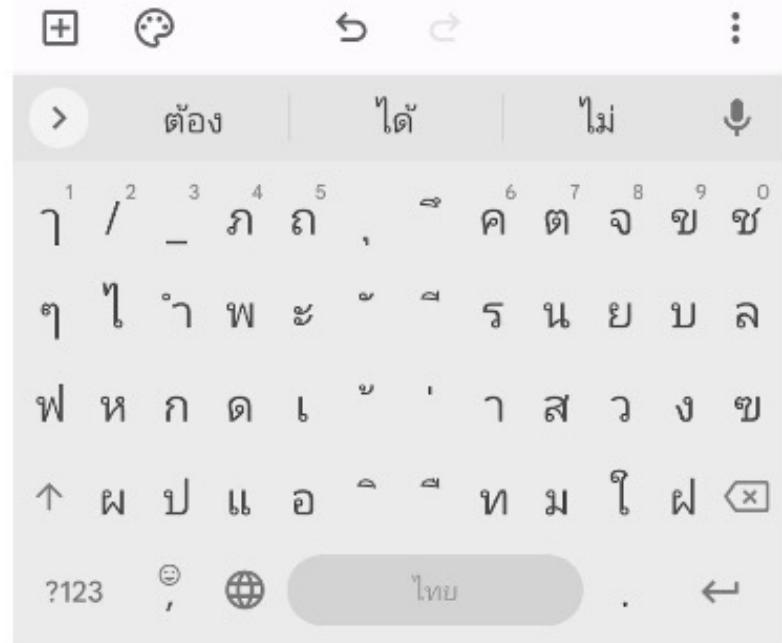
วัน นี่ เรา อยาก จะ กิน ข้าว มัน



# สร้างข้อความด้วย AI

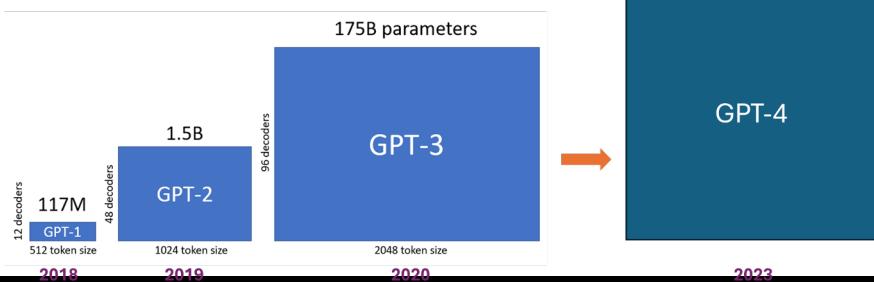


สวัสดีวันเป็นผ่านการคัดสรรร่วมกับของท่านจะ

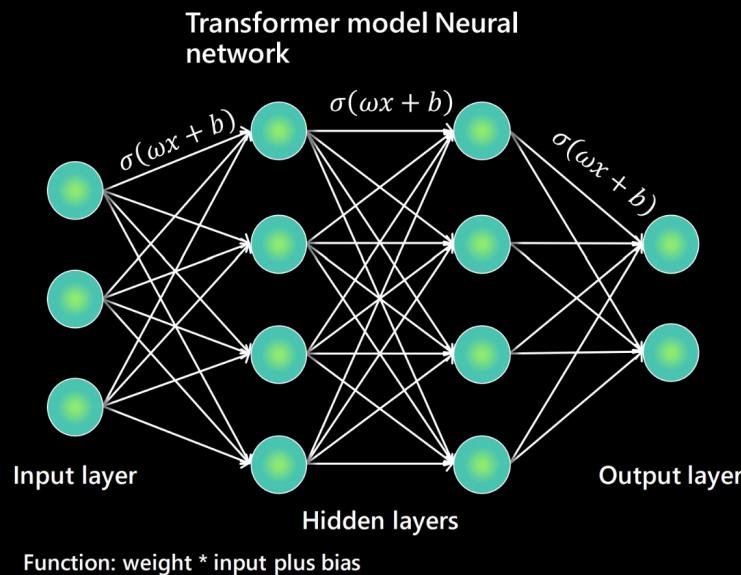


# Large + LM = LLM

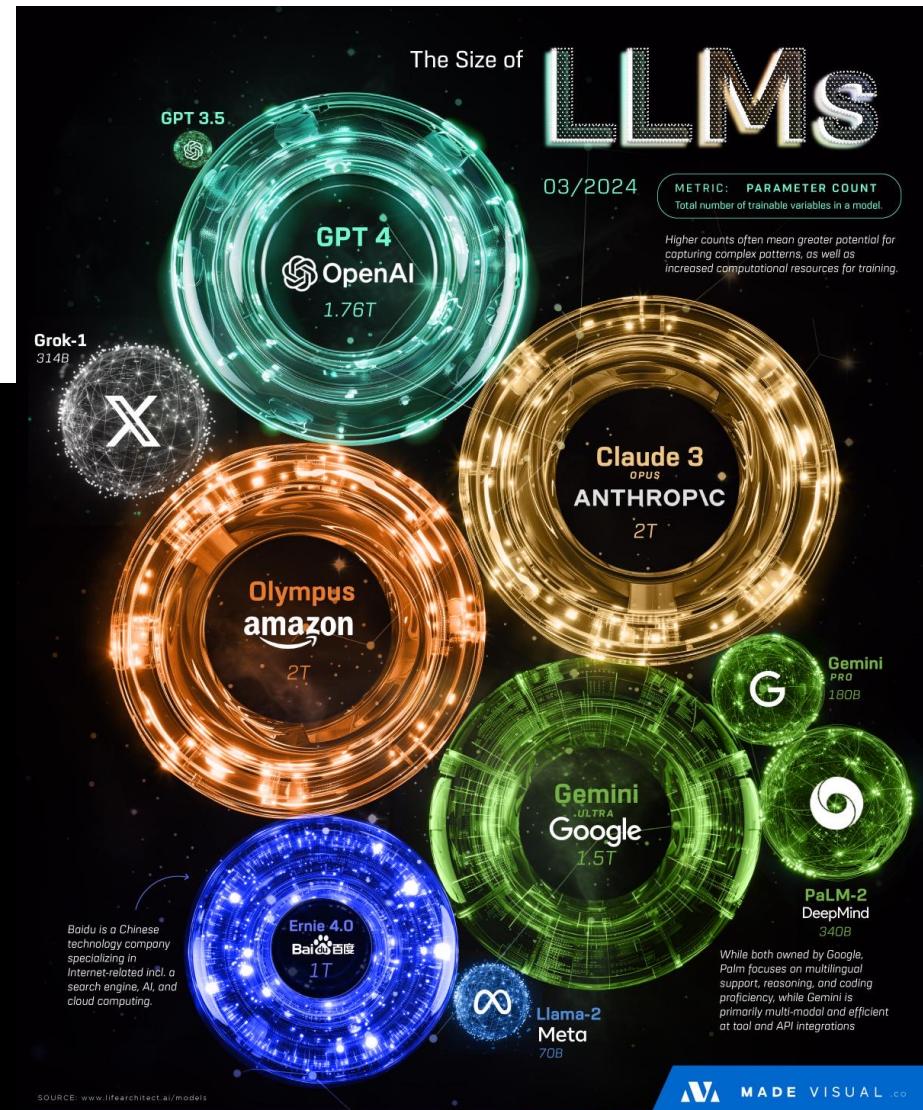
How exactly “Large” it is



How large are they?



BERT Large - 2018	<b>345M</b>
GPT2 - 2019	<b>1.5B</b>
GPT3 - 2020	<b>175B</b>
Turing Megatron NLG 2021	<b>530B</b>
GPT4 – 2023	<b>1.4T (estimated)</b>



# Nowadays there are many LLMs: API & Pretrained Models Multimodal Capabilities



Gemini

Claude

deepseek



Typhoon 2

A Family of Open Thai  
Large Language Models

Llama 4 herd: The beginning of a new era of natively multimodal AI innovation

April 5, 2025 • 12 minute read

**Openhaiqpt 1.0.0 7B Chat**  
by openhaiqpt

7b LLM, VRAM: 13.7GB, Context: 4K,  
License: apache-2.0, Quantized, Fine-Tuned

Multimodal Intelligence

Newest model suite offering unrivaled speed and efficiency

EX  
llm.extractum.io

Autumnal Country Store Scene

System Instructions

User

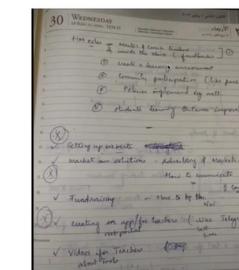
User describe this image

Model

This image captures a charming autumn scene at a country store. The store, named "1830 Country Store" as indicated by a sign above the entrance, is a dark brown wooden structure with large display windows. The store is situated amidst a courtyard paved with flagstones, and several stuffed animals are arranged on benches and picnic tables in front of the building. These plush toys depict a variety of animals, including bears, tigers, giraffes, and dogs.

Several people are relaxing in chairs around the stuffed animals, suggesting a leisurely atmosphere. A small table with baskets and a pink tablecloth sits to the right, perhaps displaying goods for sale. A flock of Canada geese strolls through the courtyard, adding to the rustic ambience.

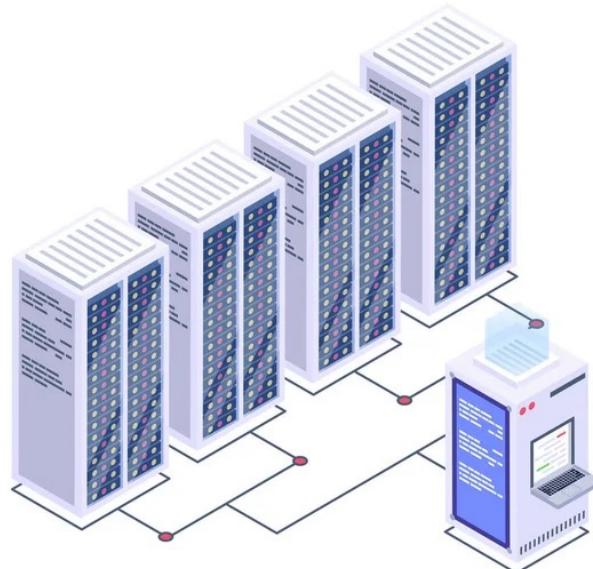
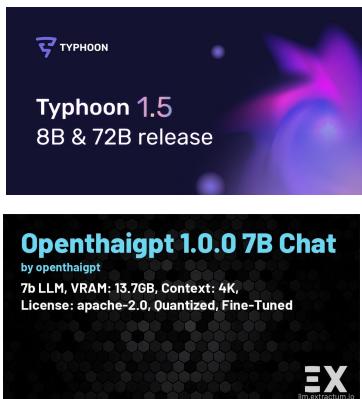
The trees surrounding the store are in various stages of autumn foliage, displaying vibrant hues of orange, red, and yellow. Fallen leaves litter the ground, further emphasizing the season. The overall lighting is soft and diffused, possibly suggesting an overcast day or late afternoon light. The image evokes a sense of peace and tranquility, characteristic of a rural setting in the fall.



- Wednesday, 30th
- HM Roles:**
- 1. Mentor & coach, provide feedback inside the class.
  - 2. Create a learning environment.
  - 3. Encourage community participation (like parents).
  - 4. Ensure policies are implemented well.
  - 5. Improve students' learning outcomes.
- Tasks/Goals:**
- Set up a website.
  - Market the solution – focus on advertising & marketing.
  - Understand how to communicate – possibly with the tool 'Canva'.
  - Fundraising: Explore ways to tap new resources.

## Loading a Pretrained Model Locally vs. Using a Cloud-based API LLM

**LLaMA**  
by  Meta



# LLM Capabilities



## ChatGPT

### 1) Assistant Chatbot

A screenshot of the ChatGPT interface. The user has typed: "Explain what ChatGPT is to me, but imagine I'm a little high." The AI response is still loading, indicated by a progress bar.

### 2) Text Summarization

A screenshot of the SCISPACE platform. A user has selected an abstract from a paper titled "CLLMate: A Multimodal Benchmark for Weather and Climate Events Forecasting". The abstract discusses the significance and innovation of the research, mentioning the challenge of translating numerical meteorological variables into textual narratives. The user has initiated a summarization task, with a progress bar indicating the process.

8v2 [cs.LG] 16 Feb 2025

### 3) Content Creation

A screenshot of the Jasper AI Script tool. A user has created a "Video Script for Jasper AI for Marketing". The script includes sections like "Opening", "Introduction to Jasper AI", and "Key Features". On the right, a "Style Guide" panel provides feedback on grammar, punctuation, and readability, with several violations highlighted.

### 4) Data Analytics

A screenshot of Gemini in Looker Conversational Analytics. It shows a "Hello, Sean" greeting and a "Chat with your data" section. Below, a "Select data to begin:" dropdown lists various data sources, including Looker Coffee datasets and Google Analytics. A "Connect to data" button is also visible.

View best practices for better answers. Gemini in Looker's Conversational Analytics is a work in progress and responses may not be complete or accurate. [Learn more](#)

---

## Prompt engineer aims to solve alignment problem.

- Language Model (LM) is originally trained to predict the next word, **NOT** answer the question.
- GPT (GPT3 is 175B parameters) is usually **frozen (not trained)**.
- Since we cannot change the model, we need to align (change) the question (also called prompt).

Input (Prompt)	Output
The patient was died.	The patient's body was found in a dark alley behind the hospital's...
"The patient was died." correct this	claim if you really believe such figures....
Poor English input: The patient was died.	Good English output: The patient died.



# Jobs of the Future: AI Prompt Engineer



Cody W Burns

Emerging Technology Visionary | Distributed Systems | Privacy | Executive Leadership

10 articles

+ Follow

October 19, 2022

## JOB OF THE FUTURE: AI PROMPT ENGINEER

Cody Burns  
24

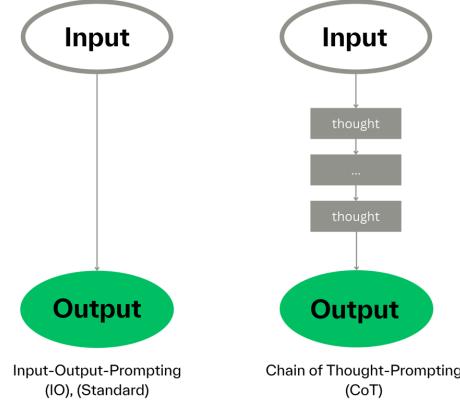
<https://www.linkedin.com/pulse/jobs-future-ai-prompt-engineer-cody-w-burns/>

# Customer service (QA & Chatbot)

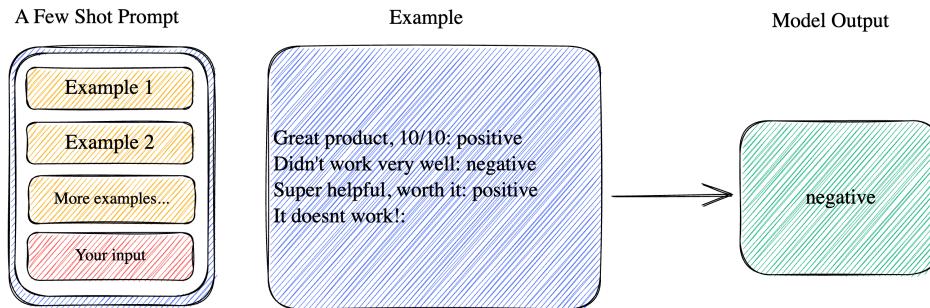
```
1 delimiter = "####"
2 system_message = f"""
3 You are a customer service assistant at a retail company. \
4 You will be provided with customer service queries. \
5 The customer service query will be delimited with {delimiter} characters.
6 Hierarchical Classify each query into a primary category and a secondary category.
7 All secondary categories must belong to primary categories.
8 You will get short explanations for each categories inside && characters. \
9 Importantly, all the category you answer must strictly satisfy the short explanations if given.\n10 Provide your output in json format with the keys: primary and secondary. \
11
12 Primary categories:
13 - Shopping
14 - Returns and Payments
15 - Promotions and Discounts
16 - Products
17 - General Inquiry
18
19 if "Shopping" is a primary category, secondary categories must follow:
20 - Order Status and Tracking
21 - Compare Products
22 - Delivery Service
23 - Missing or Damaged Items in orders
24 - Buy Products
25
26 if "Returns and Payments" is a primary category, secondary categories must follow:
27 - Returns and Refunds
28 - Billing
29
30 if "Promotions and Discounts" is a primary category, secondary categories must follow:
31 - Ongoing Sales or Promotions
32 - Applicable Discounts
33 - Classification
34
35 if "Products" is a primary category, secondary categories must follow:
36 - Product Information
37 - Product Availability
38 - Size and Color Options
39 - Warranty and Returns
40
41 if "General Inquiry" is a primary category, secondary categories must follow:
42 - Standard Questions
43 - Product Specific Questions
44 - Complaint Resolution
45 - Special Requests
46
47 """
```

```
query: 1
user query : I placed an order a few days ago, and I'm wondering when it will be delivered. Can you provide me with the shipping status?
output : {"primary": "Shopping", "secondary": "Order Status and Tracking"}
token : {'prompt_tokens': 362, 'completion_tokens': 19, 'total_tokens': 381}
-----
query: 2
user query : I received a damaged item in my order. What should I do to get a replacement or a refund?
output : {"primary": "Returns and Payments", "secondary": "Returns and Refunds"}
token : {'prompt_tokens': 355, 'completion_tokens': 21, 'total_tokens': 376}
-----
query: 3
user query : I'm interested in the new line of smartphones you have. Can you tell me more about their features and pricing?
output : {"primary": "Products", "secondary": "Product Information"}
token : {'prompt_tokens': 357, 'completion_tokens': 17, 'total_tokens': 374}
-----
query: 4
user query : Do you have any ongoing promotions or discount codes I can use for my next purchase?
output : {"primary": "Promotions and Discounts", "secondary": "Ongoing Sales or Promotions"}
token : {'prompt_tokens': 351, 'completion_tokens': 24, 'total_tokens': 375}
-----
query: 5
user query : I forgot my password and can't access my account. Can you help me reset it?
output : {"primary": "General Inquiry", "secondary": "Speak to a human"}
```

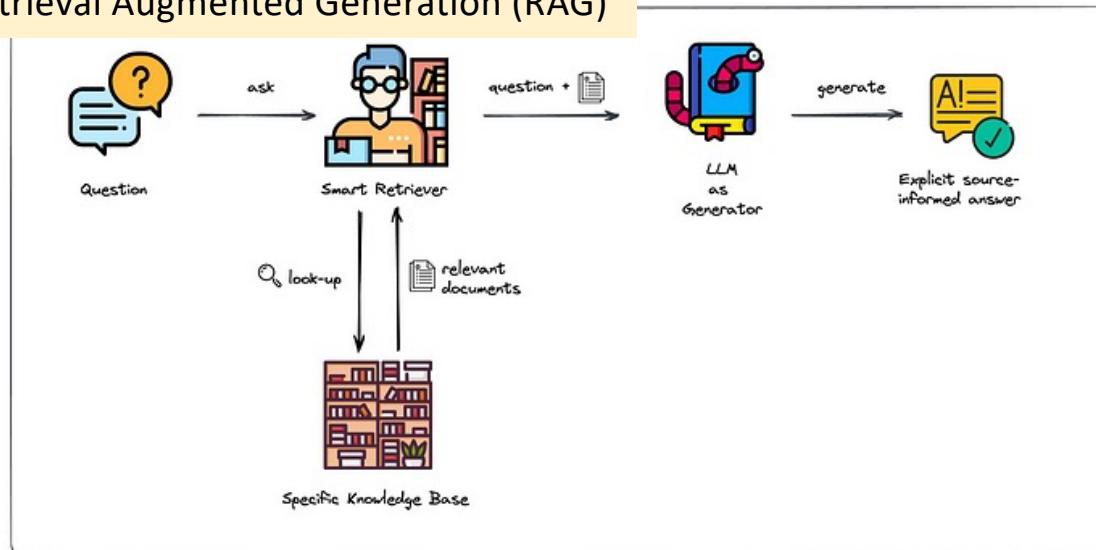
## 2) Chaing-of-Thoughts (CoT)



## 1) Few-Shot Prompting

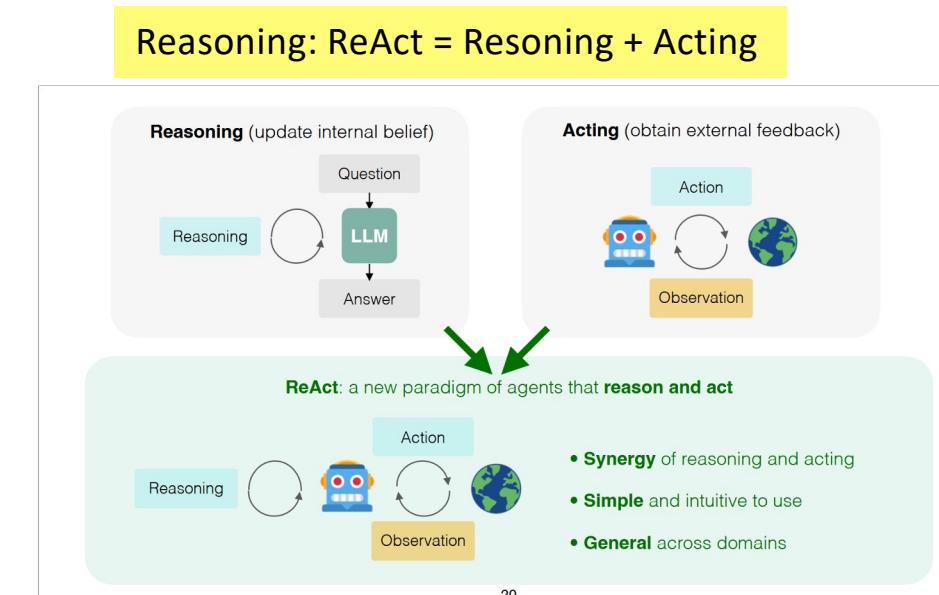
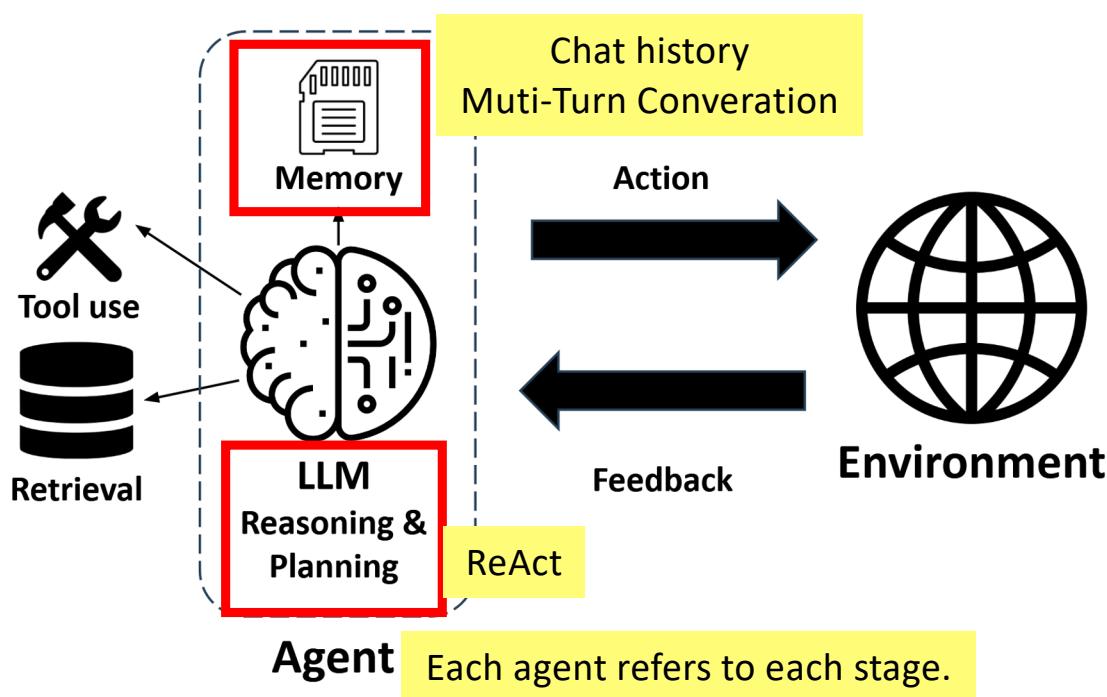


## 3) Retrieval Augmented Generation (RAG)



# Agentic LLMs

An agentic LLM is a large language model (LLM) that operates as an autonomous agent, capable of performing tasks, making decisions, and interacting with external tools or APIs to achieve the specified goals.



# LLM Use Cases: Domain Specific Chatbot

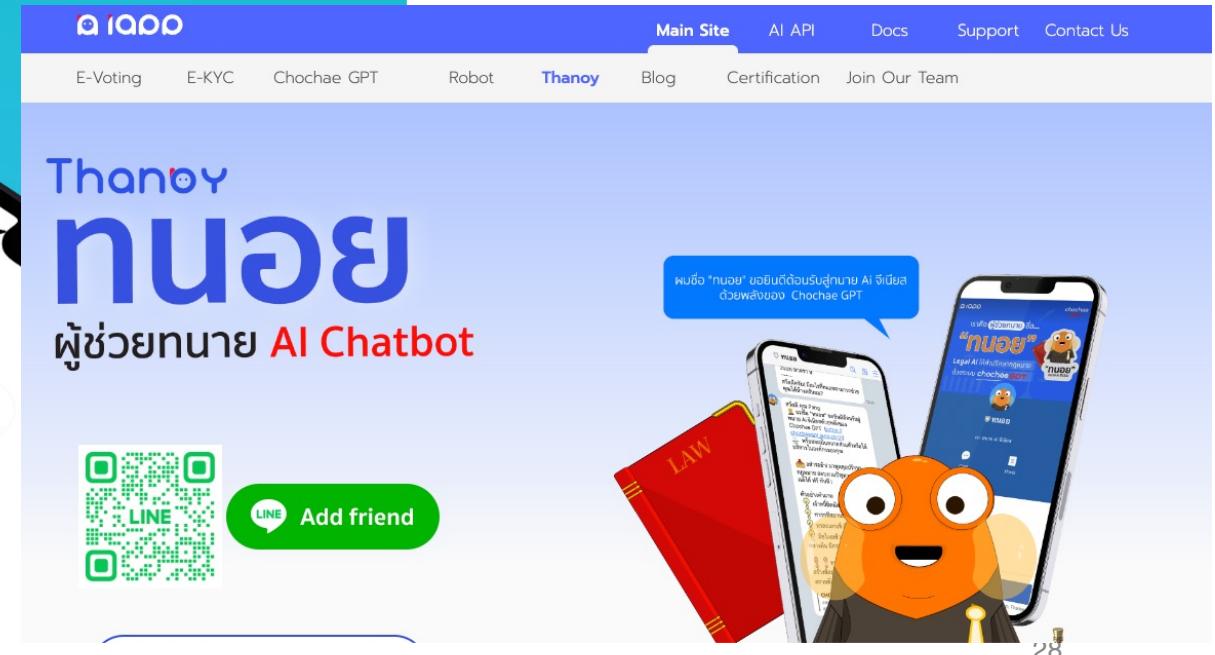


Overview Our Goals Partners Explore Try it

## MEDICAL FOCUS GENERATIVE AI

PreceptorAI is your pocket 2nd opinion for medical professionals finely tuned to Thai clinical guidelines.

Try it now



The screenshot shows the Thanoy AI Chatbot website. At the top, there's a navigation bar with links to Main Site, AI API, Docs, Support, and Contact Us. Below the navigation, the word "Thanoy" is prominently displayed in large blue letters, with "nuoy" written vertically below it. The text "ผู้ช่วยทนาย AI Chatbot" is also present. To the right, there's a smartphone displaying the Thanoy app interface, which includes a red book labeled "LAW" and a cartoon orange character with large eyes. A speech bubble above the phone says "พบว่า 'กันอย' ของคุณต้องการสุ่มนาย AI จึงได้ขอรับคำแนะนำ". On the left, there's a QR code for LINE and a green "Add friend" button.

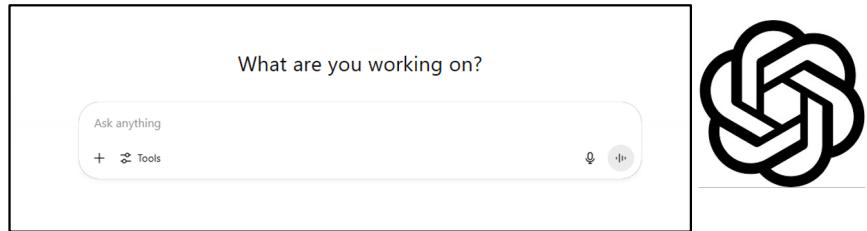
---

# AI Assistants



# 1) ChatGPT

## Overview



- The most popular and well-known AI assistant from [OpenAI](#).

## Features

- **Agent Mode:** Can perform multi-step tasks for you across different apps, like researching a topic online and creating a presentation from the results.
- **Custom GPTs:** You can build and share your own specialized versions of ChatGPT for specific tasks or topics, creating a personalized AI assistant.
- **Deep Research:** A feature for paid users that can conduct multi-stage research for complex questions, connecting to different sources to find answers.
- **Image Generation (DALL-E 3):** Creates high-quality images directly from your text descriptions.
- **Advanced Voice & Video:** Offers interactive voice conversations, and paid users get access to video generation with Sora.



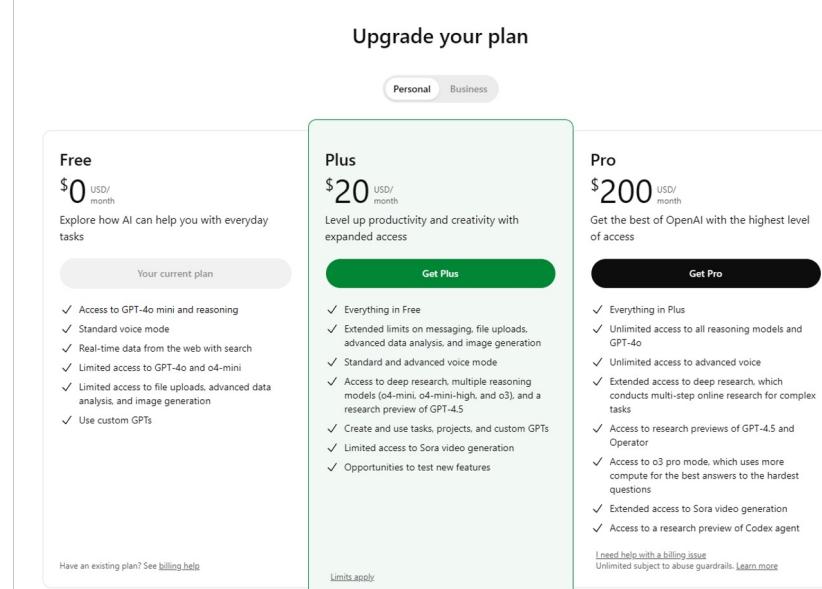
DataMining Group  
Machine Intelligence and Knowledge Discovery Lab  
Chulalongkorn University

<https://chatgpt.com/>

# 1) ChatGPT (cont.)

## Subscription plans

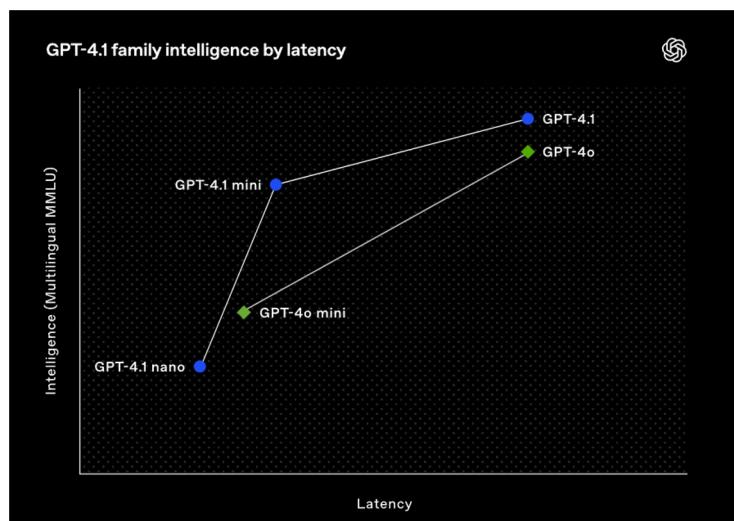
- **Free:** Great for basic chats. You get access to the standard models with **some limits on usage**.
- **Plus (\$20/month):** The best option for most people. Gives you access to the much more powerful **GPT-4o** model, which is faster and smarter. It also unlocks image creation and other advanced tools.
- **Pro (\$200/month):** For professionals who need the absolute best. It offers unlimited access to all models, including the newest and most powerful **o3-pro** for tackling the hardest questions.



# 1) ChatGPT (cont.)

## Some models (as of 28 Jul 2025)

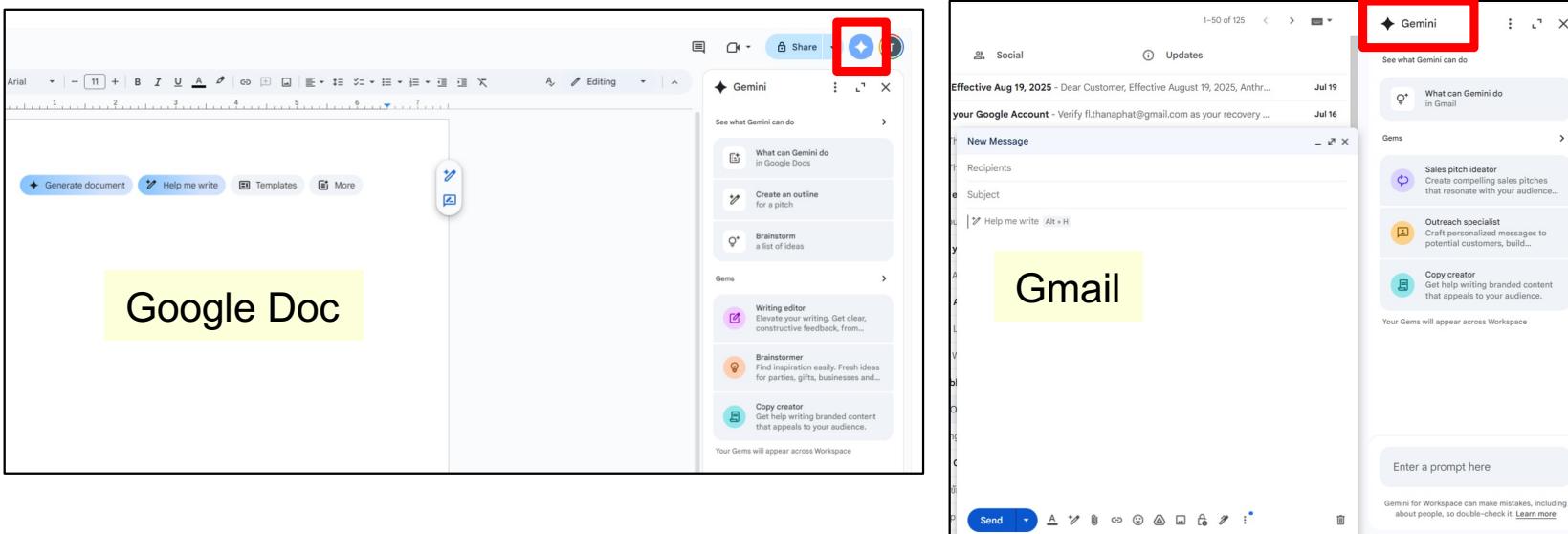
- **GPT-4o** - The main, [all-around model](#) that is perfect for almost any task ([new one is GPT-4.5 trial version](#)).
- **GPT-4.1** - Newer version of chat model. Great for quick [coding](#) and analysis
- **o3 and o4** - Built for the most complex [reasoning](#) and problem-solving



## 2) Gemini

### Overview

- This is Google's AI, and its strength is that it's built **integrated** with the Google tools many of us use every day, like Gmail, Google Docs, and Android phones.



## 2) Gemini (cont.)

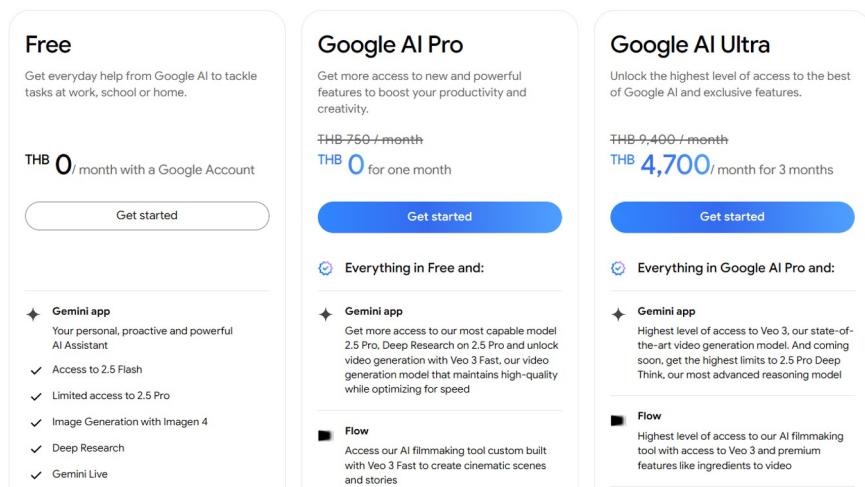
### Features

- **Deep Ecosystem Integration:** Its biggest strength is being built directly into Google apps like Gmail, Docs, Sheets, and Android phones, making your existing tools smarter.
- **Image generation (Imagen):** Creates high-quality images directly from your text descriptions.
- **Advanced Video Generation (Veo):** Paid users get access to Google's powerful text-to-video models to create high-quality video clips.
- **Grounded in Google Search:** Answers are powered by real-time information from the web, making them very current and accurate.
- **Deep research:** A feature that helps with complex questions by analyzing large amounts of information, including uploaded files, and spotting trends.

## 2) Gemini (cont.)

### Subscription plans

- **Free:** Basic access for everyday questions with a standard Google Account.
- **Google AI Pro (THB 750/month):** Upgrades you to the more capable **Gemini 2.5 Pro** model. It also integrates Gemini into Workspace apps (Gmail, Docs) and gives you 2TB of Google storage.
- **Google AI Ultra (THB 9400/month):** The super-premium tier. This gives you access to Google's most advanced reasoning model, **2.5 Pro Deep Think**, for the most complex problems, plus a massive 30TB of storage.



## 2) Gemini (cont.)

### Models

- **Gemini 2.5 Pro** - This is the powerful, state-of-the-art model available to users on the Google AI Pro plan. It excels at complex tasks, advanced coding, and understanding multiple types of input like text, images, and documents.
- **Gemini 2.5 Flash** - A model designed to offer the best balance of price and performance. It's optimized for high-volume tasks where speed and cost-efficiency are important.
- **Gemini 2.5 Pro Deep Think** - The most advanced reasoning model Google offers, designed for the most difficult and complex problems. It is exclusively available to users on the highest-priced Google AI Ultra plan.
- **Imagen 4** - A specialized model focused entirely on creating high-quality, detailed images from text descriptions.
- **Veo 2 & 3** - A family of specialized models for generating high-definition video clips from text prompts.

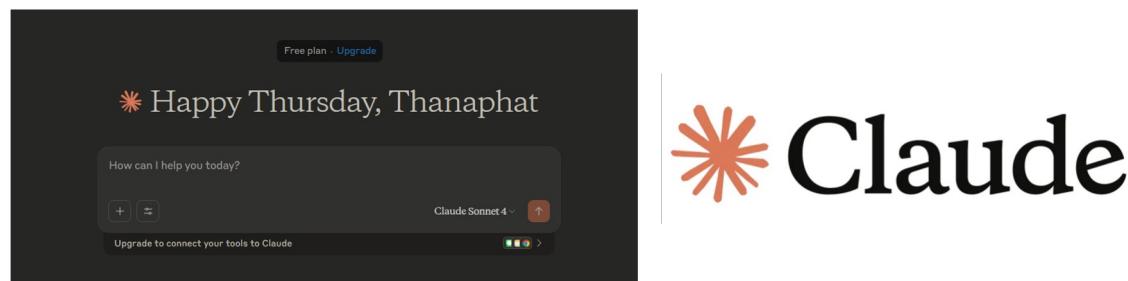
### 3) Claude

#### Overview

- An AI assistant from the company Anthropic, built with a strong focus on being safe, honest, and reliable. It's carefully designed to avoid making mistakes and giving harmful answers, making it a trusted tool for professional tasks.

#### Features

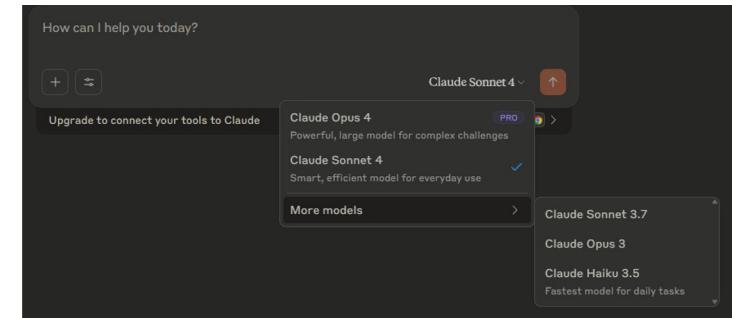
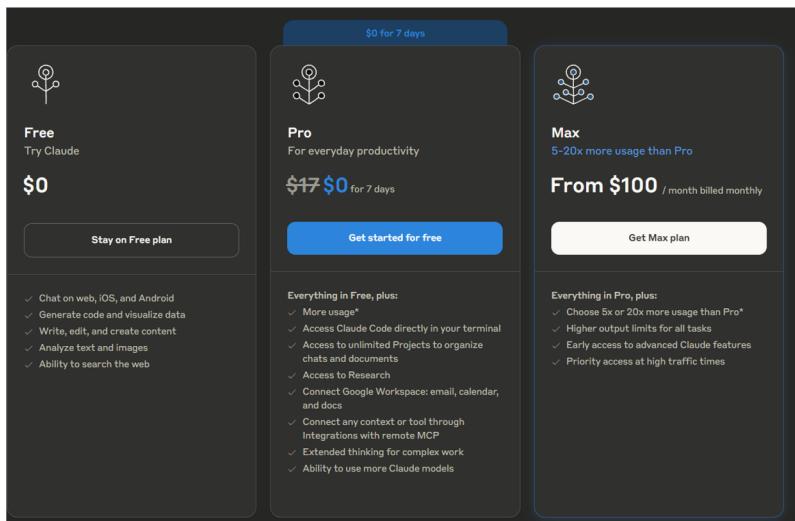
- **Massive Context Window:** Can analyze extremely long documents (up to 200,000 tokens, or over 500 pages) in a single prompt, making it perfect for summarizing books or legal contracts.
- **Artifacts & Claude Code:** Offers a dedicated workspace ("Artifacts") to create and edit code or documents. Paid users can also access "Claude Code" directly in their terminal for programming tasks.
- **Advanced Research:** A feature that can dive deep into both web sources and your uploaded documents to produce thorough, citation-backed reports.



### 3) Claude (cont.)

#### Subscription plans

- **Free:** usage is limited and resets every few hours.
- **Pro (\$17-20/month):** Gives you more usage than the free plan and access to all of Claude's models.
- **Max (\$100 or \$200/month):** For heavy business users. Offers a choice of 5x or 20x more usage than the Pro plan, plus priority access during busy times.



### 3) Claude (cont.)

#### Models

- **Haiku** - The fastest and most cost-effective model. It's best for tasks where speed is critical, like powering a live customer service chatbot.
- **Sonnet** - The balanced model. It offers an ideal mix of intelligence and speed, making it perfect for the majority of enterprise and professional tasks.
- **Opus** - The most powerful and intelligent model in the family. It is designed for the most complex tasks that require deep, multi-step reasoning, such as advanced financial analysis or scientific research.

## 4) Typhoon

### Overview

- A powerful family of AI models from **SCB 10X** that is specifically created and optimized to be an expert in the **Thai language and culture**.

### Features

- **Specialized in Thai Context:** Its main advantage is its deep, native-level understanding of the Thai language, making it more accurate and culturally aware than any global model for Thai-centric tasks.
- **Open-Source for Data Privacy:** Being able to download and run the models locally is a huge benefit for users and businesses who want to ensure their data never leaves their own systems.



**TYPHOON**

## 4) Typhoon

### Subscription plans

- **Free (Open-Source Download):** The models are free to download from platforms like Hugging Face. This is ideal for anyone who wants to run the AI on their own computer for privacy.
- **Free (Online Playground):** You can try out the models for free on the OpenTyphoon website. It's a great way to test its capabilities, but it doesn't save your chat history. <https://playground.opentyphoon.ai/playground>
- **Paid (Production API):** For businesses that need a reliable, production-grade service, Typhoon offers a pay-as-you-go API through a partnership with a platform called **Together AI**. This is a very cost-effective way to use the models without hosting them yourself.

### Models

- **Typhoon 2:** A comprehensive set of models of various sizes (from small to a very large **70B** parameter version)
- **Typhoon 2.1 Gemma:** Newer, **lightweight**, and highly efficient models. fine-tuned on Gemma3 4b and 12b
- **Typhoon OCR:** A bilingual document parsing model built specifically for real-world documents in Thai and English
- **Typhoon T1 & R1:** These are specialized models built for advanced **reasoning**. They are excellent for tackling complex math, coding, and logic problems, especially in a bilingual Thai-English context.
- **Typhoon 2 Audio:** An experimental model that can understand and generate speech, **supporting both speech-to-text and text-to-speech in Thai and English**.



Data Mining Group  
Machine Intelligence and Knowledge Discovery Lab  
Chulalongkorn University

---

# Conclusion

# Why do need to learn Gen AI? Everyone has been using it.

---

## Google Disputes News That Search Engine Use Is Falling

Google pushed back on a news report suggesting Search traffic is declining, saying query volume is still growing.



SEJ STAFF

Roger Montti

May 8, 2025 · 3 min read

70 SHARES  
4.9K READS



<https://www.searchenginejournal.com/google-disputes-news-that-search-traffic-is-falling/546233/>

## ChatGPT Surges in Popularity: Now Boasting Over 200 Million Weekly Users

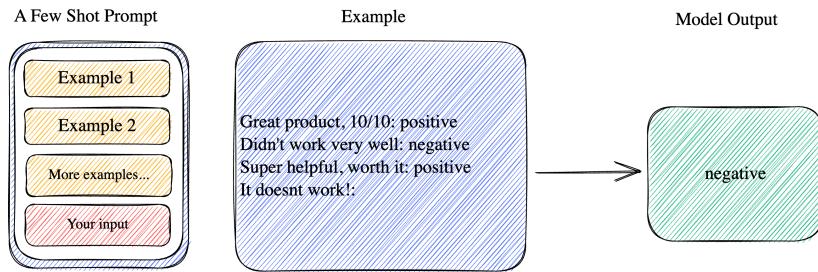
Onome · September 2, 2024



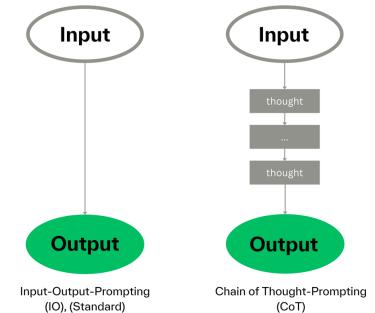
43

# Prompting is important!

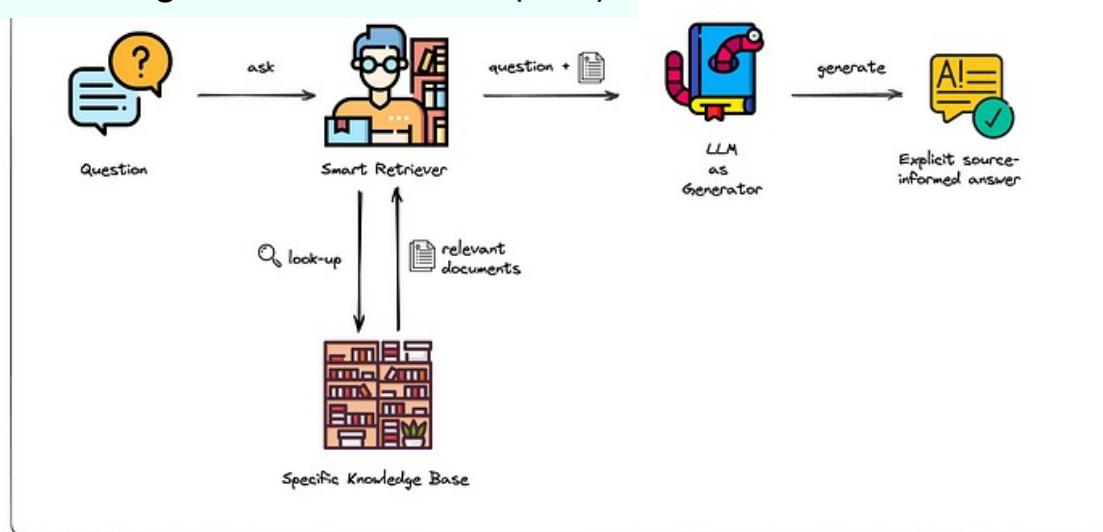
## 1) Few-Shot Prompting



## 2) Chain-of-Thoughts (CoT)



## 3) Retrieval-Augmented Generation (RAG)



---

**Q&A**

**Peerapon.v@chula.ac.th**