

+

<https://github.com/pvateekul/ieat2026>



Advanced ML

Prof. Peerapon Vateekul, Ph.D.
Peerapon.v@chula.ac.th



Outlines

- Deep Learning
- Gen AI
- LAB 5: Object Detection

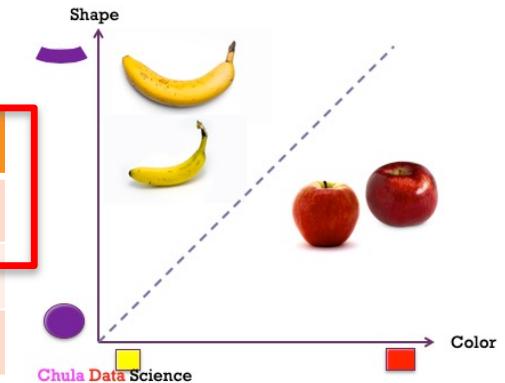
+

Deep Learning



Handcrafted features

Age	Income	Gender	Province	Purchase
25	25,000	Female	Bangkok	Yes
35	50,000	Female	Nontaburi	Yes
32	35,000	Male	Bangkok	No



Can we still tell the features (columns)?





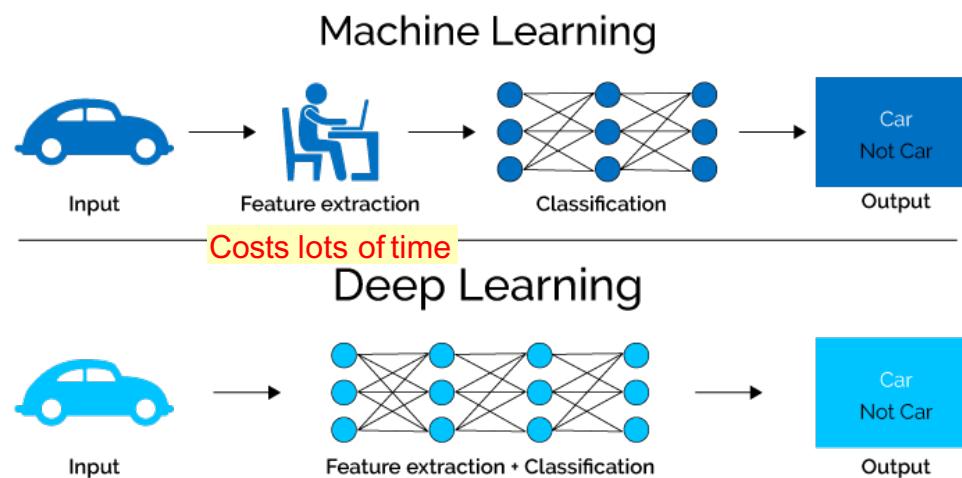
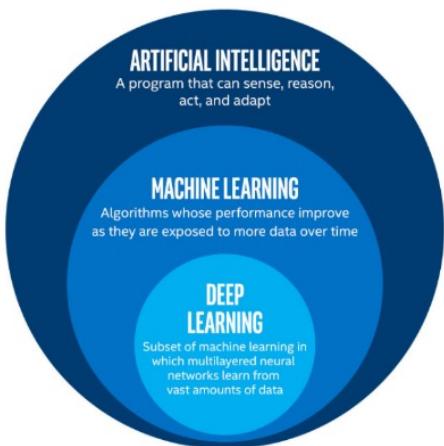
What is Deep Learning (DL)?



Part of the machine learning field of learning representations of data. Exceptional effective at learning patterns.



Utilizes learning algorithms that derive meaning out of data by using a hierarchy of multiple layers that mimic the neural networks of our brain.





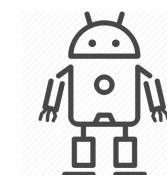
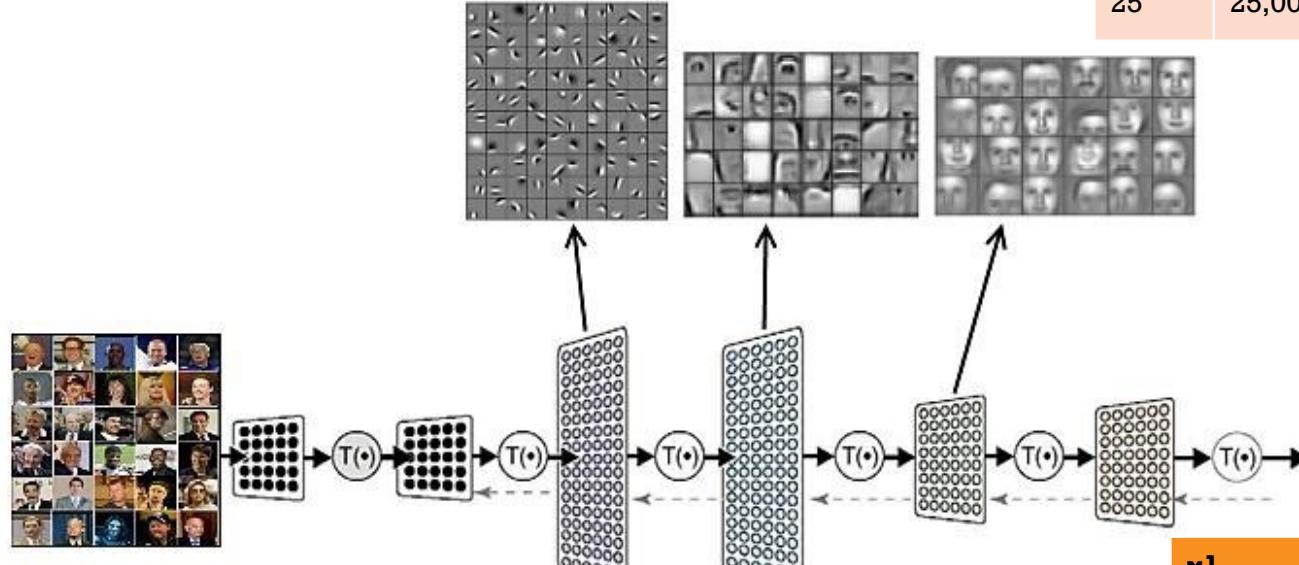
Deep Learning – Basics (cont.)

What did it learn?

A deep neural network consists of a **hierarchy of layers**, whereby each layer **transforms the input data** into more abstract representations (e.g., edge -> nose -> face). The output layer combines those features to make predictions.

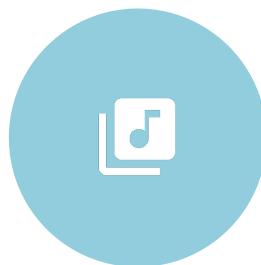


Age	Income	Gender	Province	Corona
25	25,000	Female	Bangkok	Yes

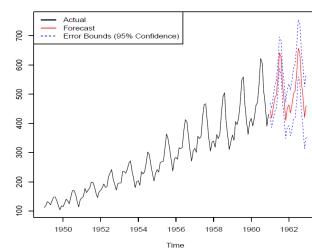


x1	x2	x3	x4	Corona
0.7	0.2	-0.5	-0.1	Yes

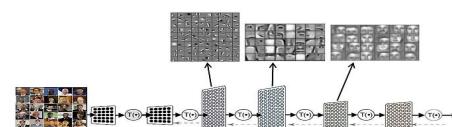
Deep Learning Application



Speech
Recognition



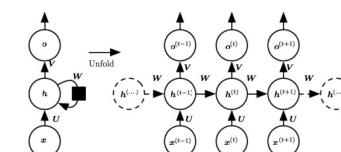
Computer
Vision



CNN



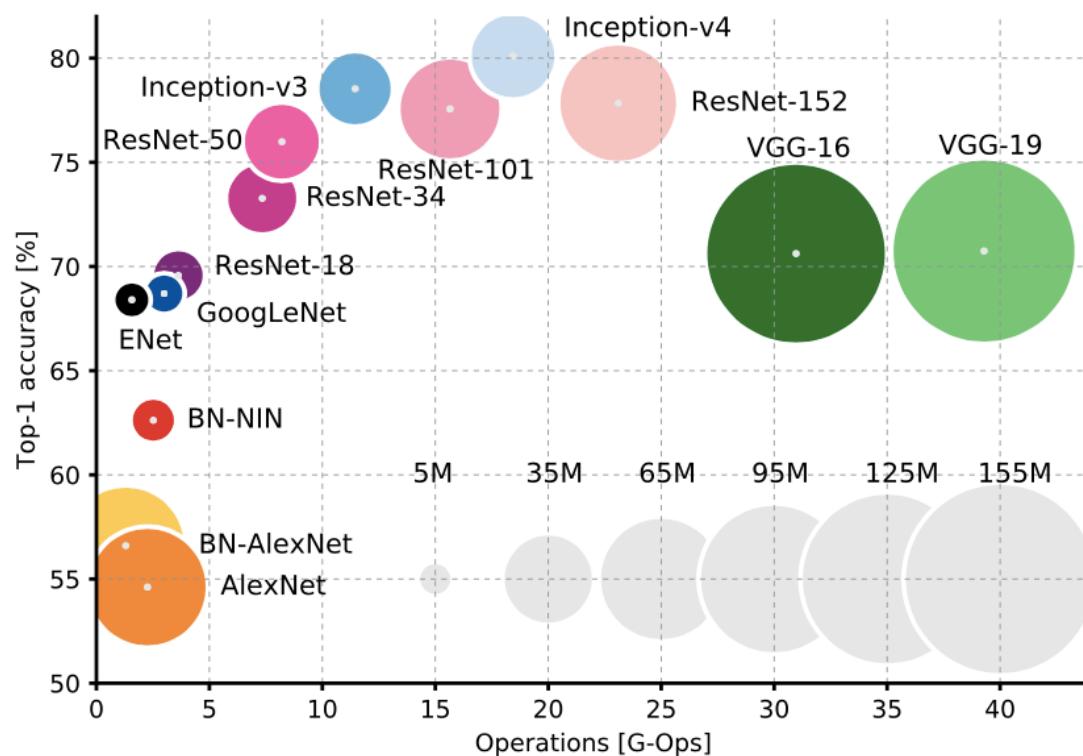
Natural Language
Processing



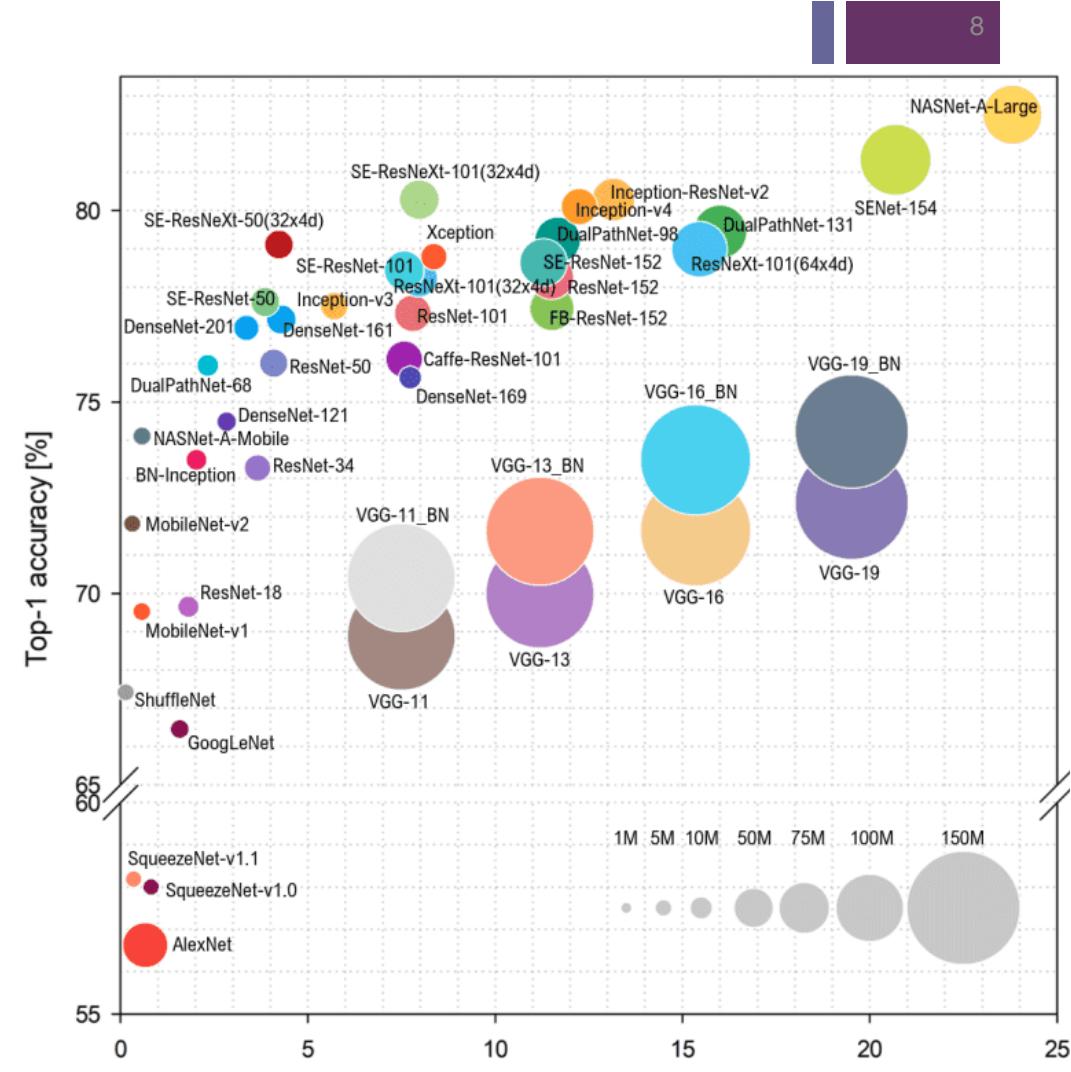
RNN (LSTM)

Spend = $500 + 10 * \text{Age} + 20 * \text{Income1K}$

SOTA of Image Classification



https://blog.csdn.net/qq_34216467/article/details/83061692

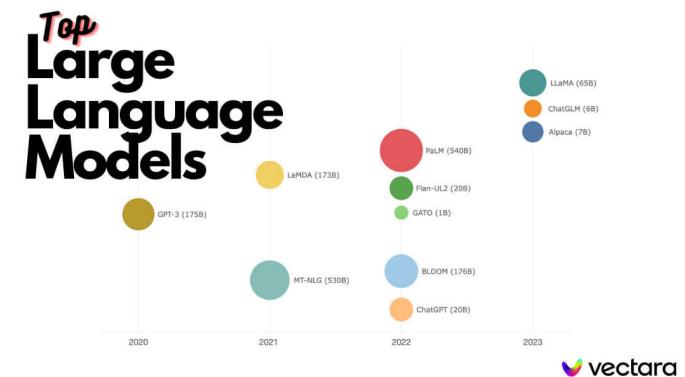
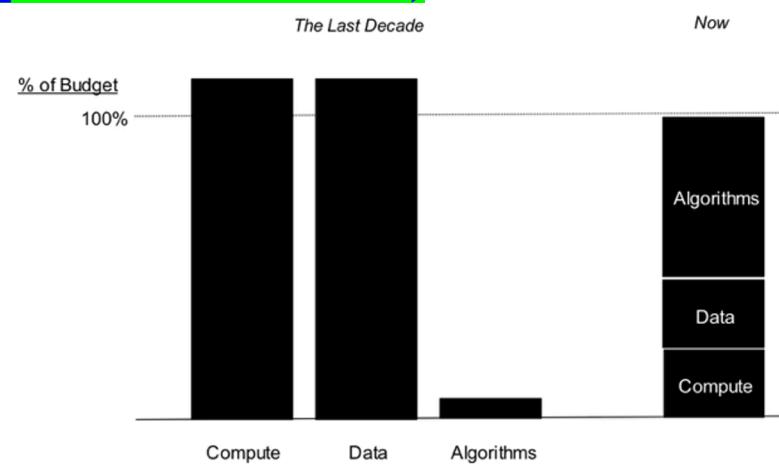


<https://theaisummer.com/cnn-architectures/>



Why now

- Neural Networks has been around since 1990s
- Big data – DNN can take advantage of large amounts of data better than other models
- GPU – Enable training bigger models possible
- Deep (pretrained models) – Easier to avoid bad local minima when the model is large

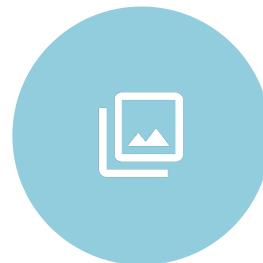
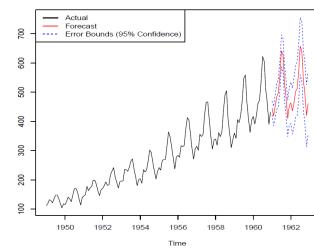


Based on the slide of Aj.Ekapol [/www.kdnuggets.com/2017/06/practical-guide-machine-learning-understand-differentiate-apply.html](http://www.kdnuggets.com/2017/06/practical-guide-machine-learning-understand-differentiate-apply.html)

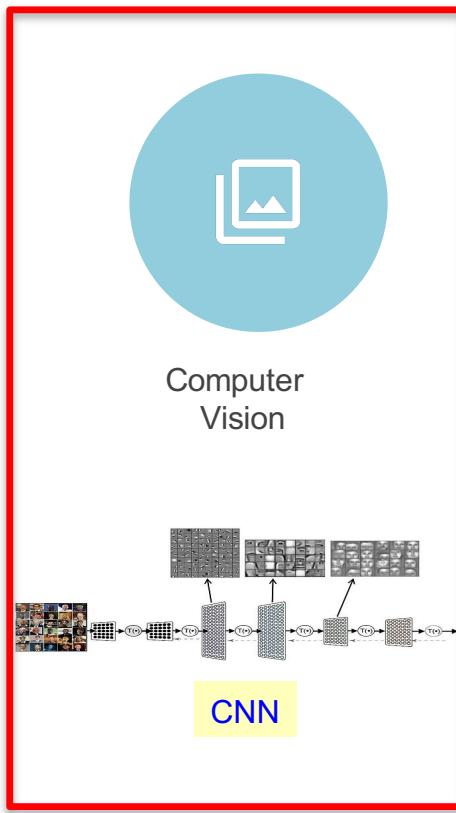
Deep Learning Application



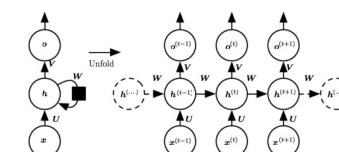
Speech
Recognition



Computer
Vision



Natural Language
Processing



RNN (LSTM)

Type of image tasks

Semantic Segmentation



GRASS, CAT,
TREE, SKY

No objects, just pixels

Classification + Localization



CAT

Single Object

Object Detection



DOG, DOG, CAT

Multiple Object

Instance Segmentation



DOG, DOG, CAT

This image is CC0 public domain



DOWNLOAD THE APP



English | 简体中文



Ultralytics YOLOv8 is a cutting-edge, state-of-the-art (SOTA) model that builds upon the success of previous YOLO versions and introduces new features and improvements to further boost performance and flexibility. YOLOv8 is designed to be fast, accurate, and easy to use, making it an excellent choice for a wide range of object detection and tracking, instance segmentation, image classification and pose estimation tasks.

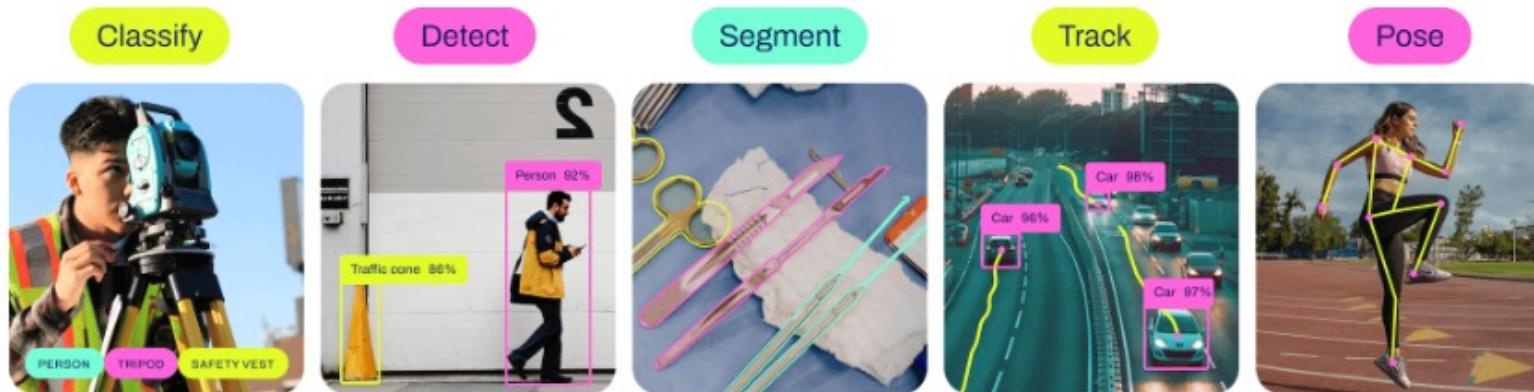
<https://github.com/ultralytics/ultralytics>



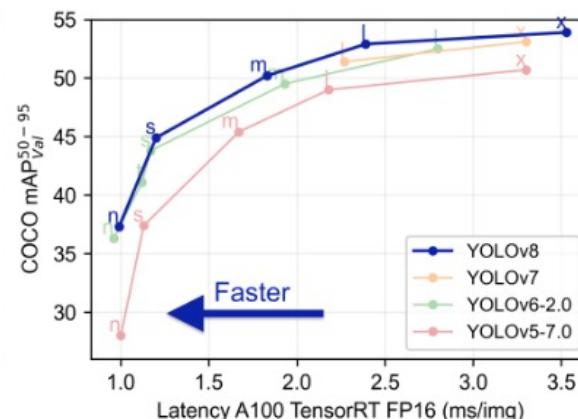
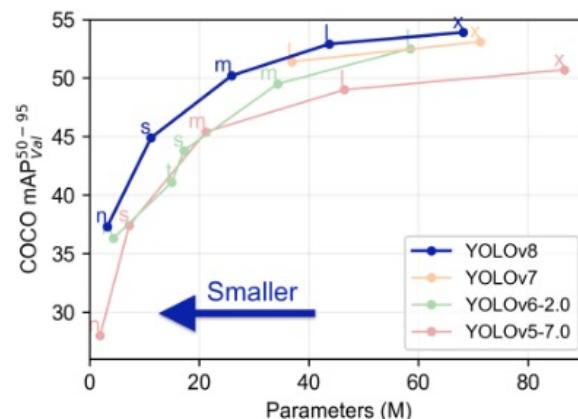
Models

13

YOLOv8 Detect, Segment and Pose models pretrained on the COCO dataset are available here, as well as YOLOv8 Classify models pretrained on the ImageNet dataset. Track mode is available for all Detect, Segment and Pose models.



All Models download automatically from the latest Ultralytics [release](#) on first use.

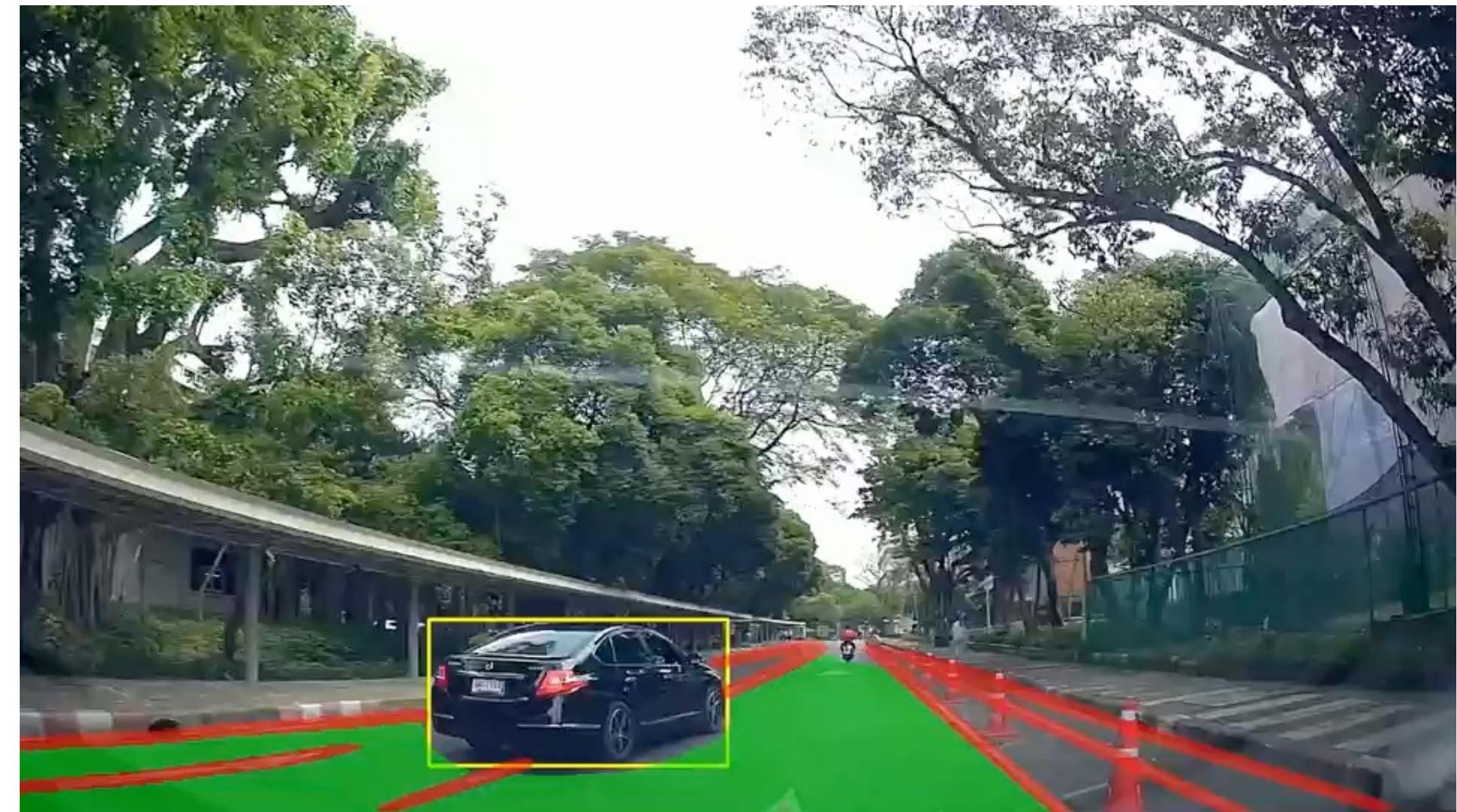


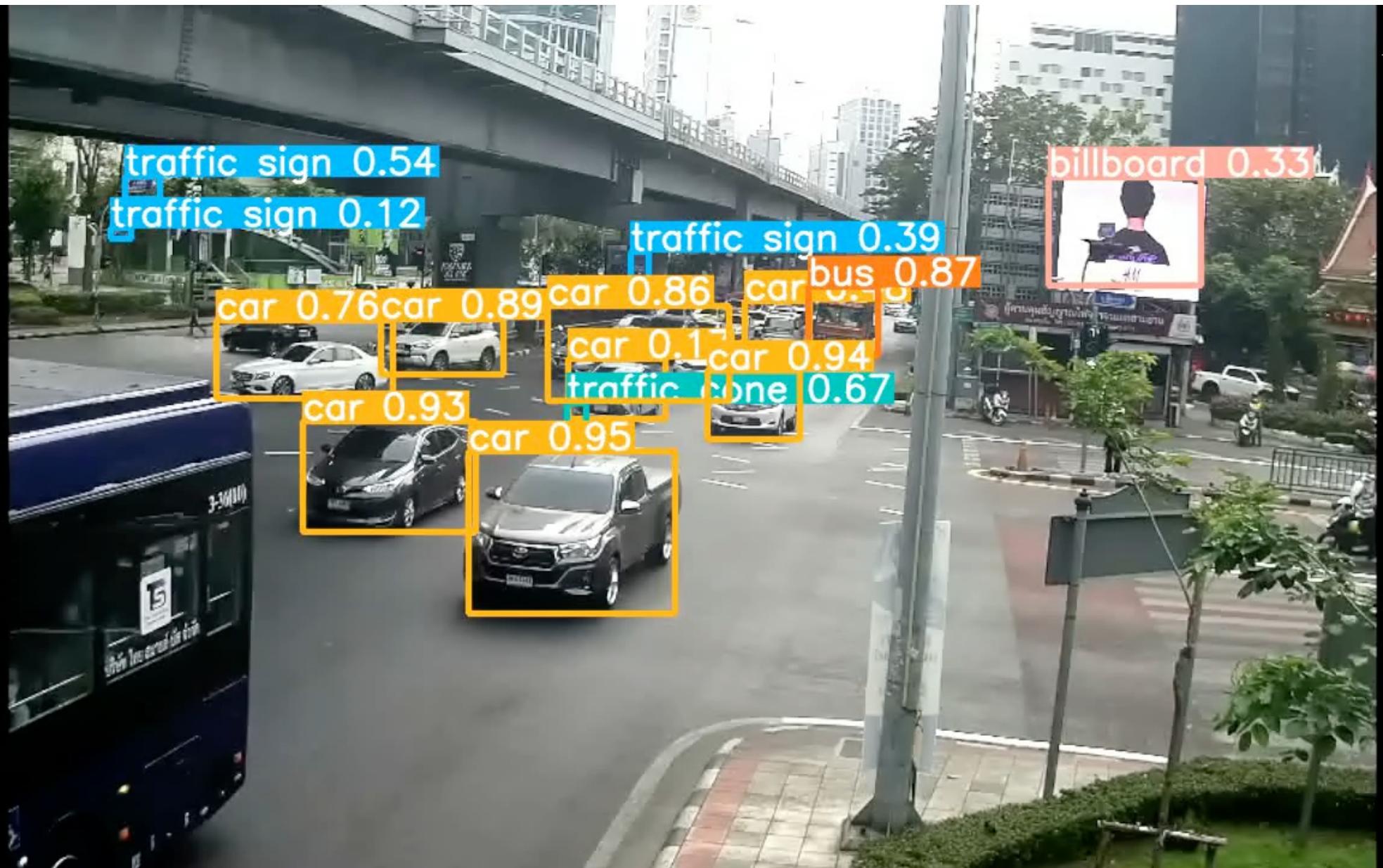
PEOPLE COUNTING

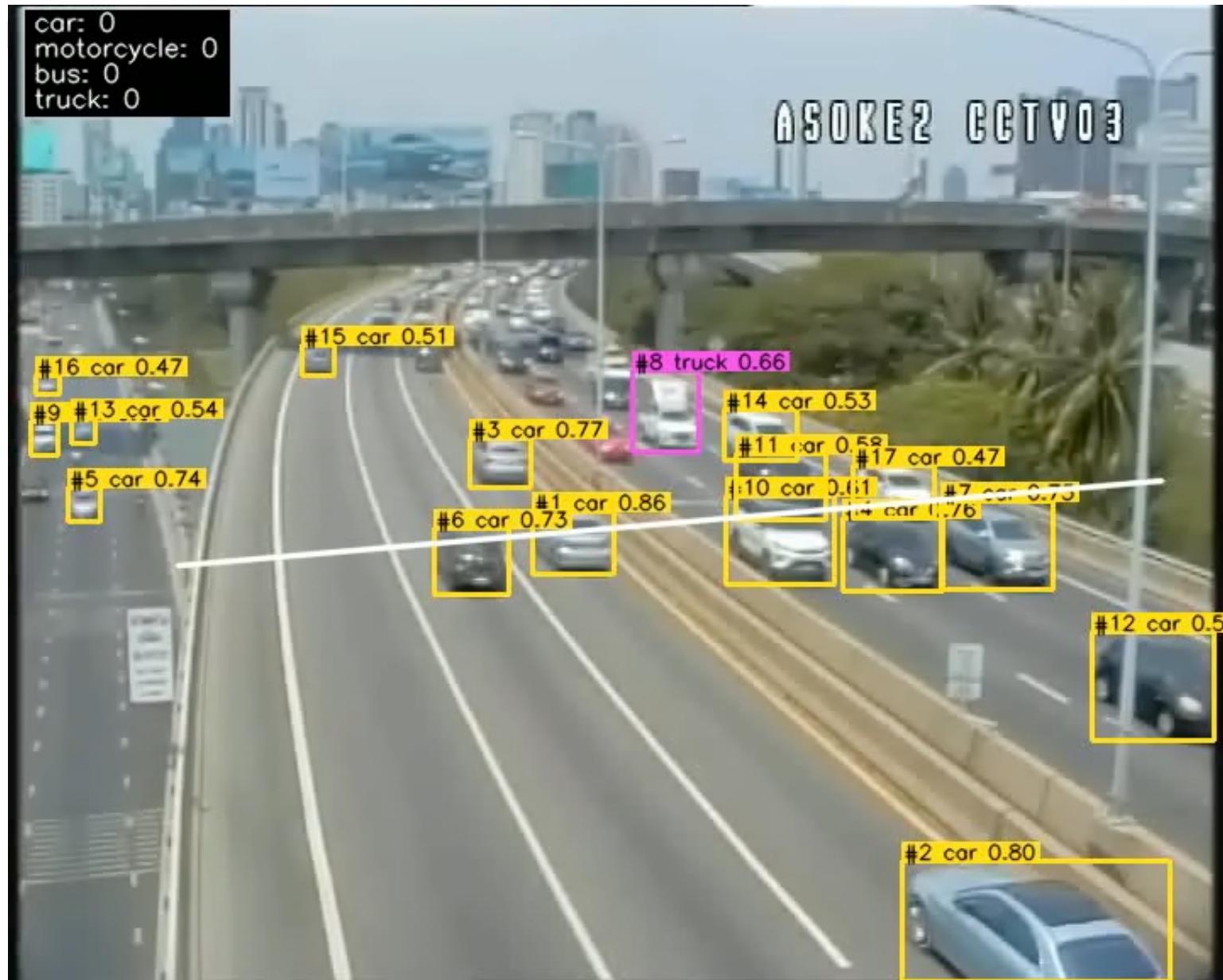


<https://www.hikvision.com/th/newsroom/latest-news/2019/people-counting-for-increased-retail-success/>

boat 0.40.8 boat 0.55

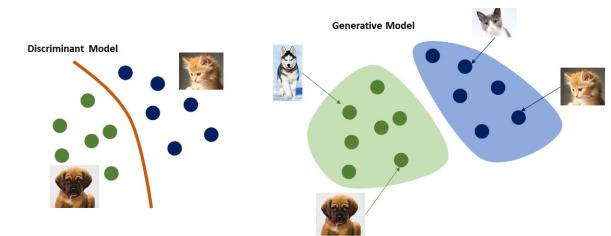




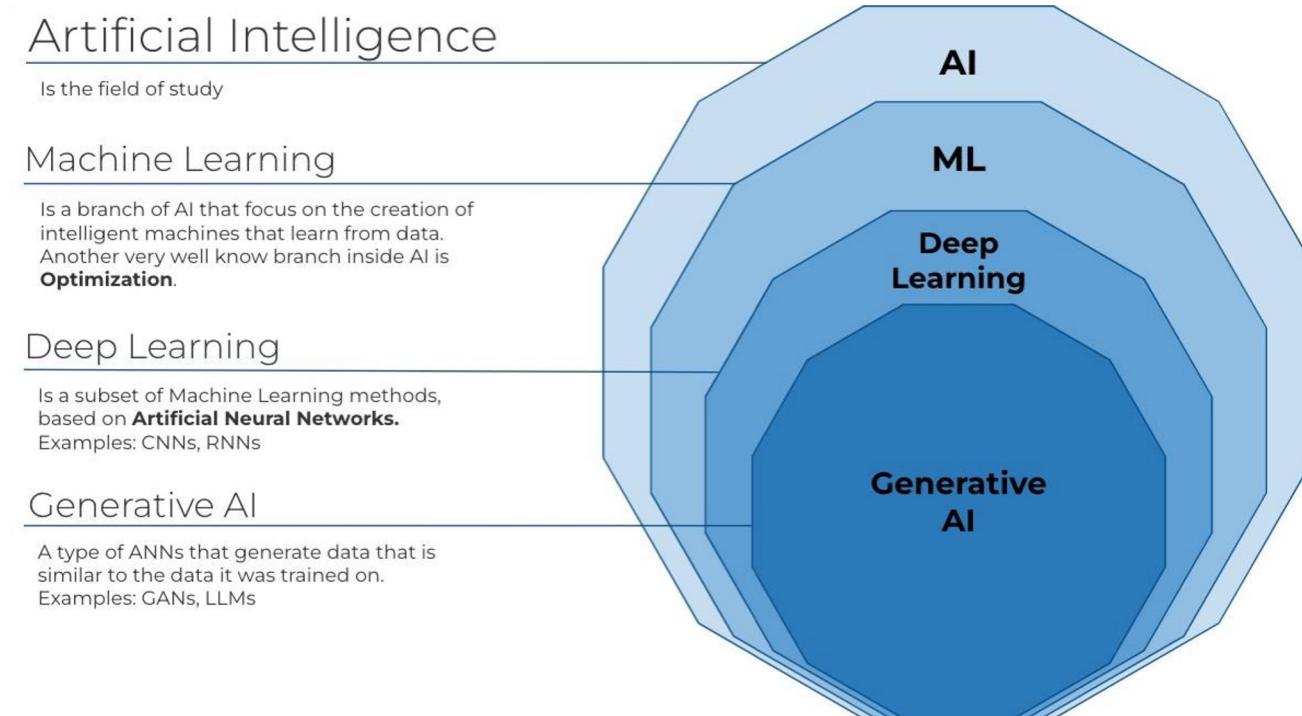


+ Image Classification

What is Generative AI?



Generative artificial intelligence (AI) is a type of AI **that can create new content, such as text, images, music, and videos**. It uses machine learning models to learn from its training data to “generate” new statistically probable outputs.



1) Generative AI - Image

- Midjourney by the Midjourney team
- Imagen 3 by Google
- DALLE 3 by OpenAI

QUESTION #12/12 – WHICH PIECE WAS MADE BY A HUMAN?

Thanks for playing! You got 6 out of 12 correct.

Da Vinci Detective: Not too shabby! You're on your way to becoming an art sleuth.

[Facebook](#) [Twitter](#) [E-Mail](#) [Pinterest](#)

<https://www.foundmyself.com/blog/ai-art-quiz/>



DALL·E 3



2) Generative AI - Text

- ChatGPT by OpenAI
- Claude by Anthropic
- Gemini by Google
- DeepSeek by the Deepseek team



 Claude

Gemini

 deepseek

3) Generative AI - [Video](#)

- Sora by OpenAI
- Veo 2 by Google



4) Multimodal Capability

- Most frontier models are capable of understanding inputs from multiple modalities such as audio and video.
- This ability greatly increases the variety of their use cases such as converting image to text.

Autumnal Country Store Scene ⓘ

Save a copy

System Instructions

Optional tone and style instructions for the model

User



User

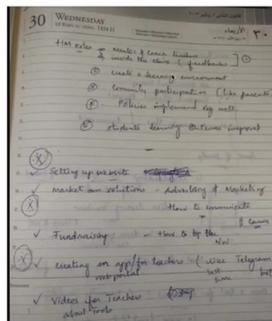
describe this image

Model

This image captures a charming autumn scene at a country store. The store, named "1836 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 ambiance.

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.

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.**

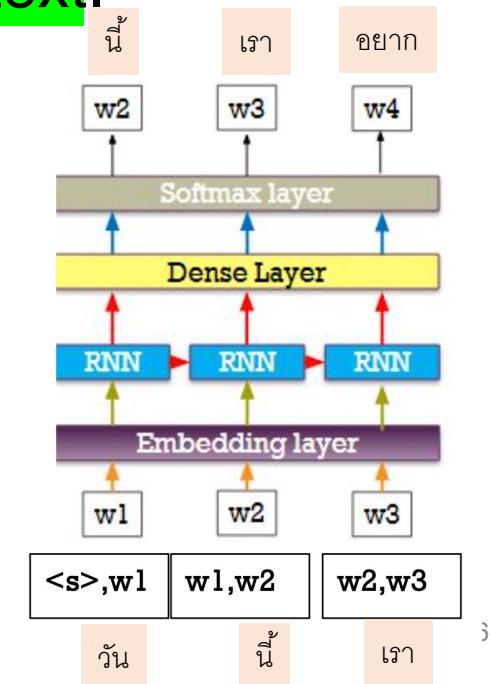


Language Model (LM)

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

วันนี้เราอยากจะกินข้าวมัน _____

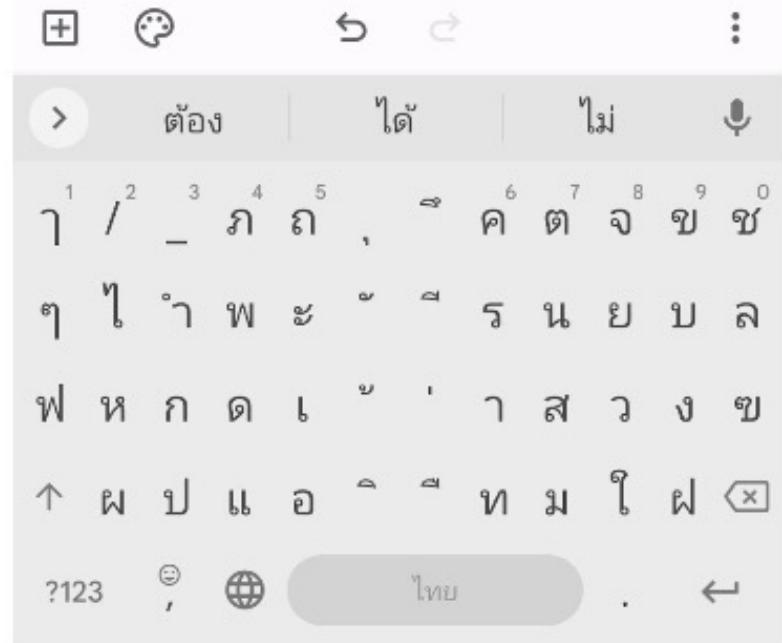
วัน นี่ เรา อยา ก จะ กิน ข้า ว มัน



สร้างข้อความด้วย AI



สวัสดีวันเป็นผ่านการคัดสรรร่วมกับของท่านจะ

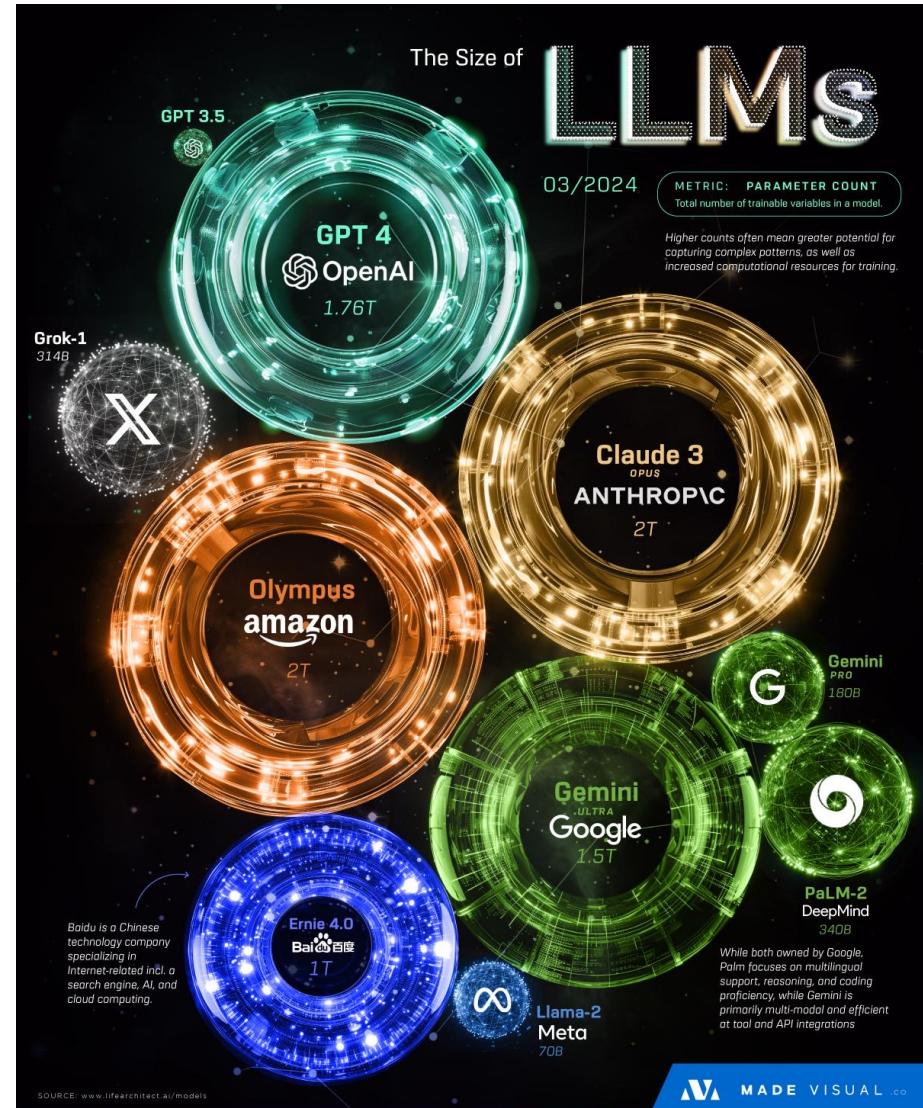


Large + LM = LLM

Model	Year	Parameters (approx)	Tokens Seen (approx)
GPT-1	2018	117M	~0.1B
GPT-2	2019	1.5B	~10B
GPT-3	2020	175B	~300B
GPT-4	2023	~1.7T (MoE est.)	~13T
GPT-5	2025	300B	~114T

Amazon now develops its own family of **foundation models** called **Amazon Nova** (launched late 2024):

- **Nova Micro** → lightweight text-only model (fast, low cost)
- **Nova Lite** → multimodal (text, images, video), efficiency-focused
- **Nova Pro** → stronger multimodal, balance between accuracy & speed
- **Nova Premier** → top-tier multimodal model (teacher model for distillation)
- **Nova Canvas** → image generation
- **Nova Reel** → video generation
- **Nova Sonic** → speech-to-speech (voice)
- **Nova Act** → agent model that can operate tools and browsers



Popular LLM

Currently, large language models are very popular and there are many models available for use. They are categorized by their application as follows:

1. Open-source models: [Llama-4](#), [Mistral](#), [DeepSeek-V3](#), [Gemma-3](#), [Qwen-3](#)
2. Closed-source models: [GPT-5](#), [Claude](#), [Gemini](#)
3. Specific models: [Typhoon-2.5](#), [SEA-LION](#)



LLM Application



ChatGPT

1) Assistant Chatbot

A screenshot of the ChatGPT interface. A user message says: "Explain what ChatGPT is to me, but imagine I'm a little high." The AI response is partially visible below.

2) Text Summarization

The SCISPACE platform interface. On the left, there's a sidebar with various icons. In the center, a paper titled "CLLMate: A Multimodal Benchmark for Weather and Climate Events Forecasting" by Haobo Li, Zhouwei Wang, Jiachen Wang, and YueYa Wang from Hong Kong University of Science and Technology is displayed. Below the paper, there's a "Chat" section where a user has asked "Explain Abstract of this paper". A summary of the abstract is shown, followed by a list of bullet points under "Introduction of Forecasting". At the bottom, there's a text input field for asking questions and a "High Quality" button.

3) Content Creation

The Jasper AI Script interface. It shows a "Video Script for Jasper AI for Marketing" document. The script includes sections like "Opening (0:00 - 0:15)", "Introduction to Jasper AI (0:15 - 0:45)", and "Key Features (0:45 - 1:30)". To the right, there's a "Style Guide" panel with several grammar and punctuation errors highlighted, such as "Style Guide has identified a typo" and "Your Style Guide states Revolutionize as a violation".

4) Data Analytics

The Gemini in Looker Conversational Analytics interface. It shows a "Hello, Sean" greeting and a "Chat with your data" section. Below, there's a "Select data to begin:" dropdown menu with options like "[GOLD] Looker Coffee - BigQuery", "[GOLD] Looker Coffee - Looker Core", and "[0] Orders, Products, and Customers - Looker Coffee...". A "Connect to data" button is at the bottom.

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
32

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

Prompt Engineering Techniques



Zero-shot
prompting



Few-shot prompting
or in-context
learning



Chain-of-thought
prompting



Tree-of-thought
prompting



Iterative
refinement



Feedback
loops



Prompt
chaining



Role-playing



Maieutic
prompting

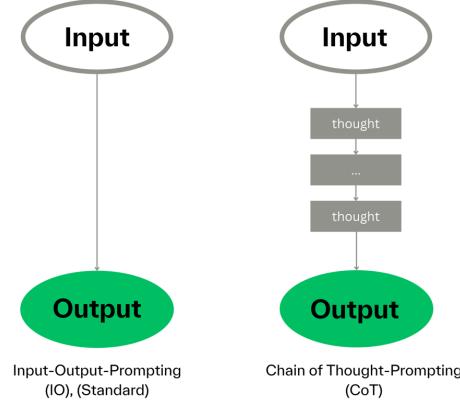


Complexity-based
prompting

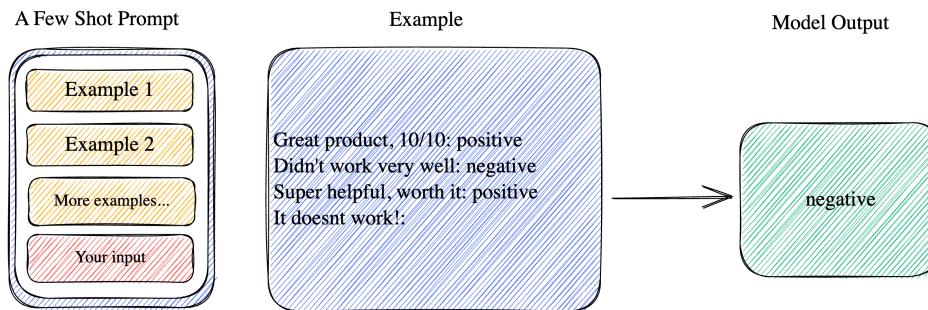
servicenow.

<https://www.servicenow.com/ai/what-is-prompt-engineering.html>

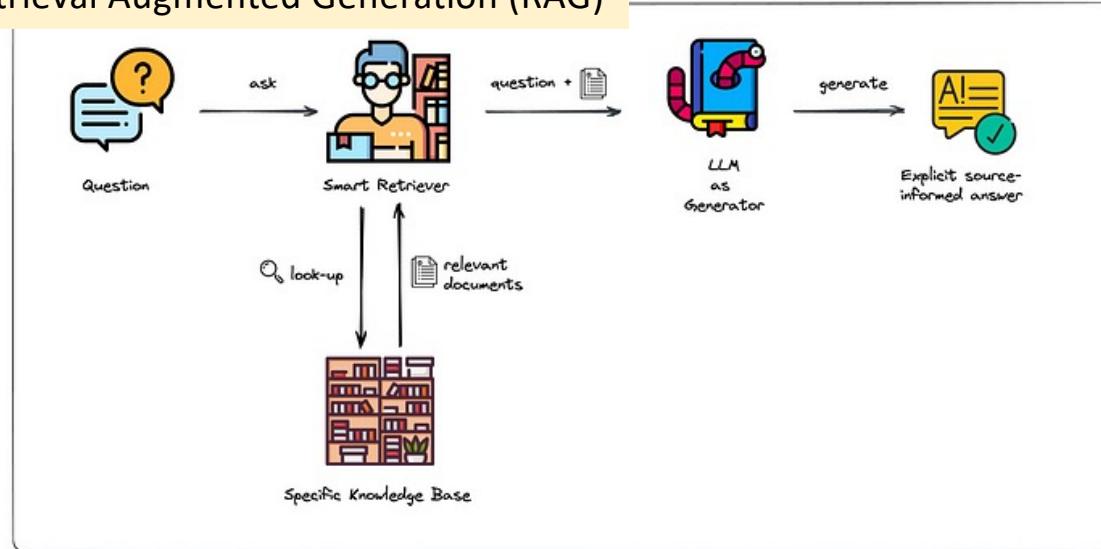
2) Chaing-of-Thoughts (CoT)



1) Few-Shot Prompting

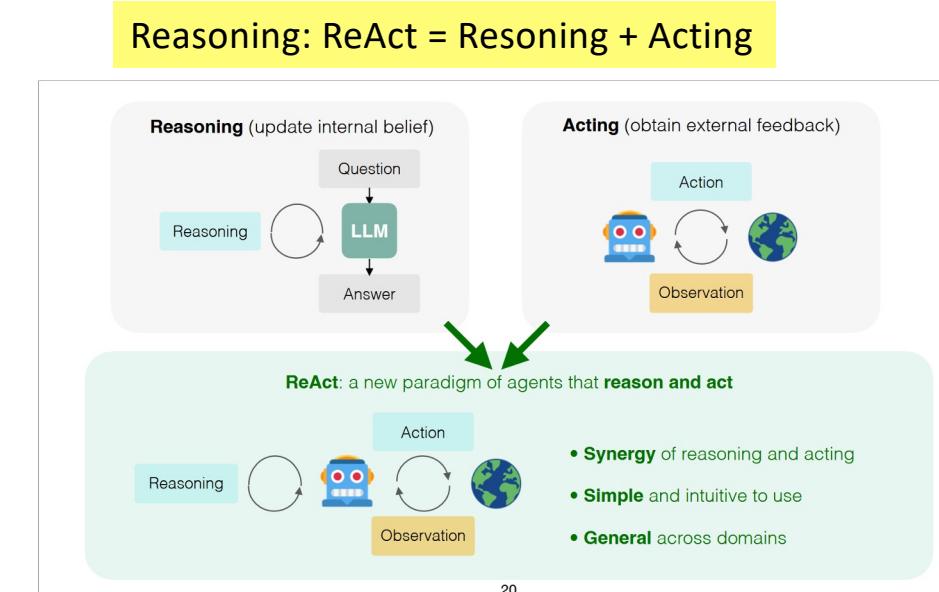
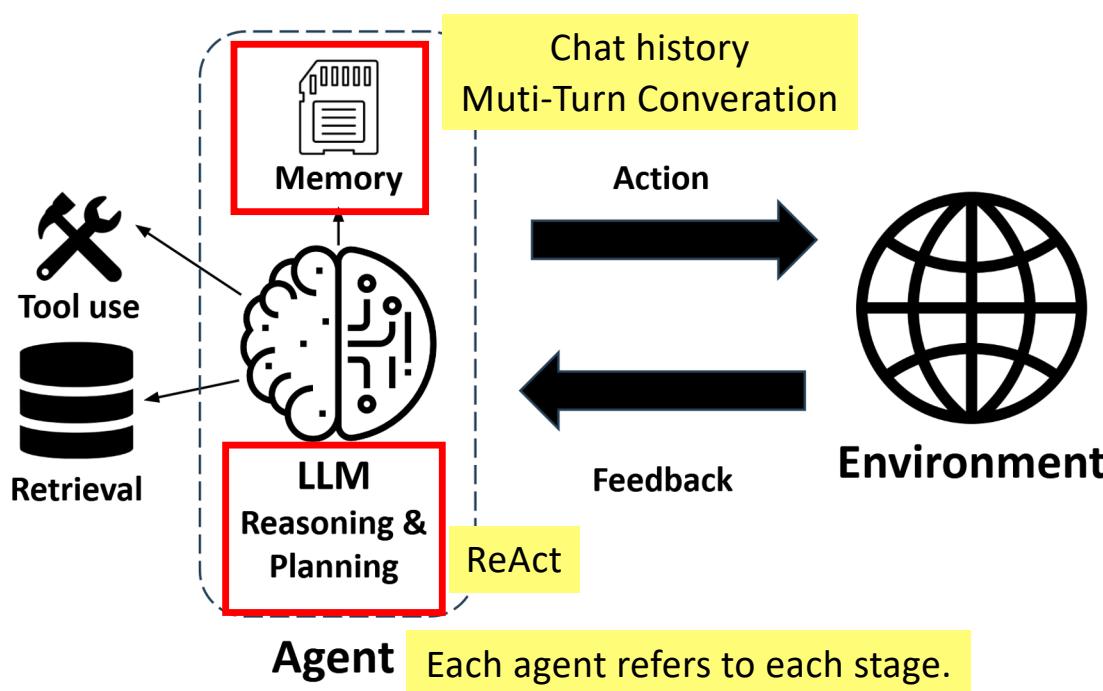


3) Retrieval Augmented Generation (RAG)



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.



From Just Generative AI to **Agentic AI** From Chatbot to **Personal Assistant**

- User: “Book me a round-trip flight from Bangkok to Tokyo next weekend, preferably morning flights, and send me the itinerary.”
- Agentic AI plans, searches, books, and delivers the itinerary automatically.

The screenshot shows a news article from LARA (LaraNews) titled "Agentic AI booking tool launched by Kiwi.com". The article is by Rob Munro, dated August 8, 2025, and features the "Featured, Ticketing" category. The headline is prominently displayed over a background image of a computer monitor showing the Kiwi.com logo.

Online travel agency Kiwi.com has launched a new system that allows AI-powered airline direct booking, a trend rapidly emerging as a significant industry disruptor.

The Czech company, in partnership with technology firm Alpic, is the first in the industry to release a Model Context Protocol (MCP) server, which acts as a direct interface between its flight inventory and major AI platforms.

Model Context Protocol (MCP) is an open standard that creates a unified way for large language models (LLMs) to interact with external services, tools, and data. The protocol has been described as the “USB-C for AI” because it provides a standardized communication method, much like how a USB-C cable connects various devices.

<https://www.laranews.net/agentic-ai-booking-tool-launched-by-kiwi-com/>



LLM Use Cases: Domain Specific Chatbot

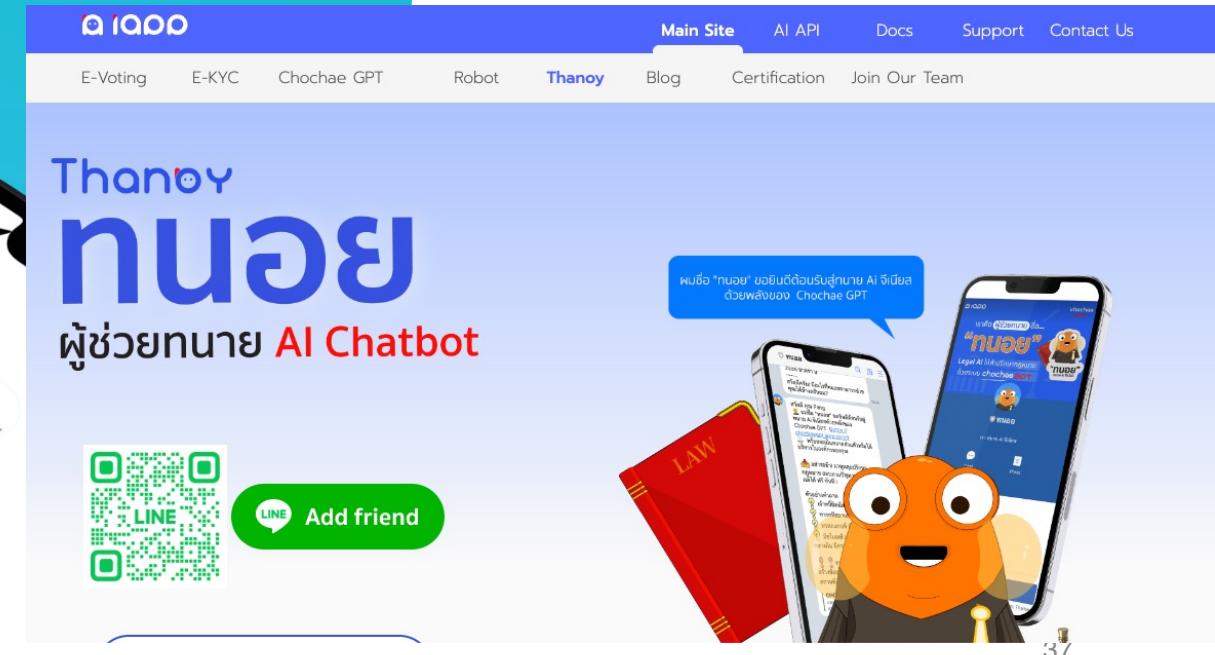


Overview Our Goals Partners Explore Try it

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 left, there's a QR code for LINE and a green "Add friend" button. On the right, there's a cartoon character of a yellow fish-like creature with a tie, standing next to a smartphone displaying the Thanoy app interface. A speech bubble above the phone says: "พบว่า 'กันอย' ของคุณต้องการสุ่มนำ้ด้วย AI จึงได้อ่านข้อมูลเพิ่มเติม". The page has a light blue background.

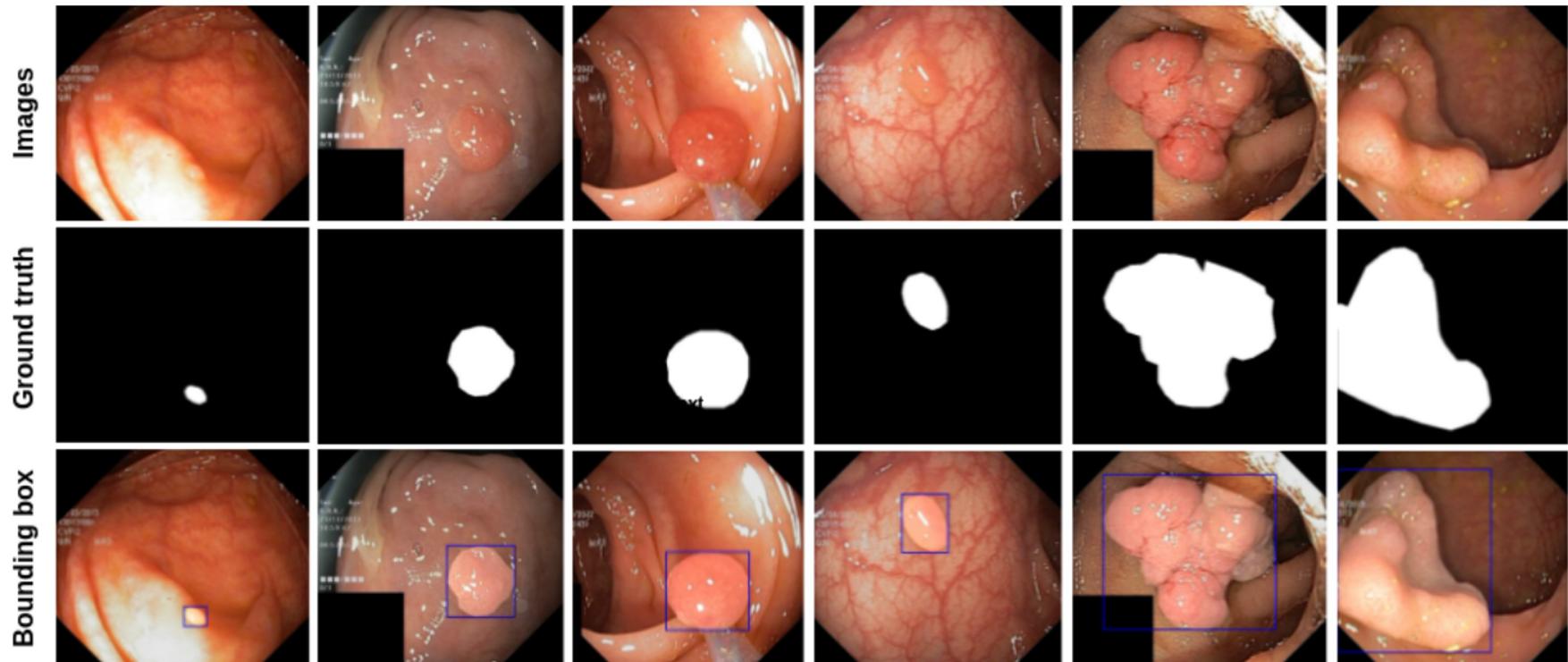


LAB 5: Object Detection

Dataset: kvasir dataset (2017)

- **Kvasir dataset** consists of 4,000 images, annotated, including 8 classes showing anatomical landmarks, pathological findings or endoscopic procedures in the GI tract.
- The dataset consist of the images with different resolution from 720x576 up to 1920x1072 pixels in JPG format and documents in JSON format
- The dataset was released in 2017 by the **Simula Research Laboratory, Norway**.
- To simplify the experiment, we selected only **500 images** containing polyps and prepared the dataset in a format compatible with YOLO training.

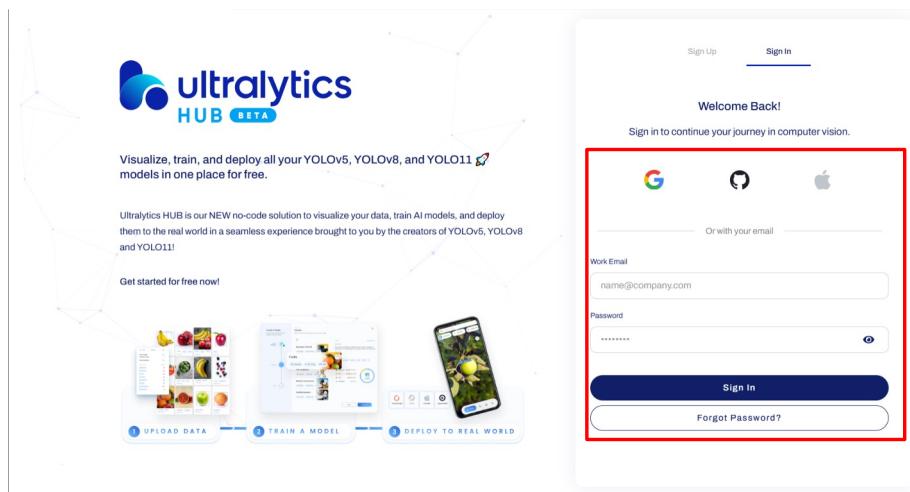
Dataset: [kvasir dataset \(2017\)](#)



The figure shows the example images, bounding box, and mask from Kvasir-SEG. The white mask shows the area covered by the polyp region, and the background regions contain non-polyp tissue pixels.

YOLOv8n (Ultralytics Hub)

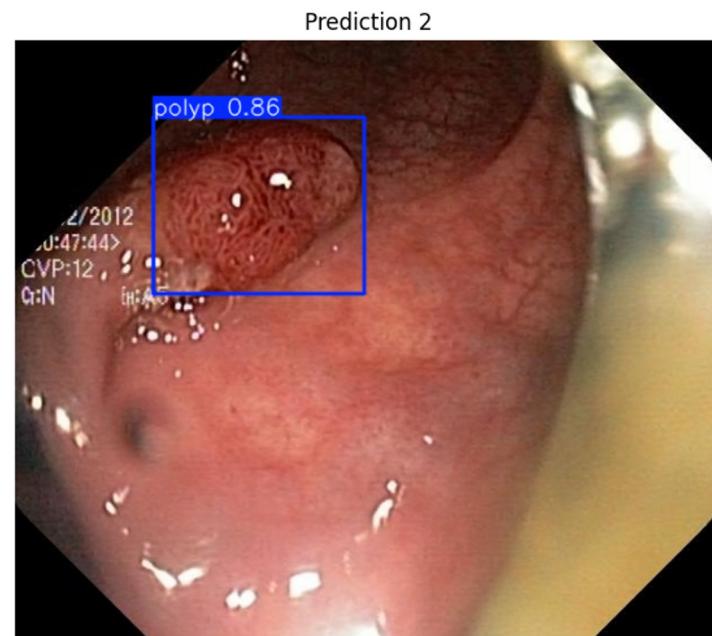
Sign in [Ultralytic hub](#)



The image shows the Ultralytics Hub home dashboard. On the left, a sidebar menu lists 'Home', 'Datasets', 'Projects', 'Models', 'Integrations', 'Support', and 'Trash'. A user profile icon 'N Natapon' is at the bottom. The main area has a 'Get Started' section with a 'Start Welcome Tutorial' button and a 'Safety helmets' thumbnail. Below it are sections for 'Datasets' (with 'Upload Dataset' button), 'Projects' (with 'Create Project' button), and 'Models' (with 'Train Model' button). On the right, there's a 'Recent' section with a search bar and a list of recent files ('test_segment_33', 'test_32', 'test_segment_32'). A large circular profile picture for 'Hello Natapon!' is shown, along with storage information ('Storage 61.9 MB / 20.0 GB') and a 'Feedback' button. A small 'v1.1.87' watermark is in the bottom right corner.

YOLOv8n (Ultralytics Hub)

Results may vary between runs due to random seed initialization and hyperparameter tuning; however, the overall performance should be similar to the results shown on this page.



+

Thank you & any questions