



CHULA ENGINEERING COMPUTER
Foundation toward Innovation



Introduction to Deep Learning

Assoc. Prof. Peerapon Vateekul, Ph.D.

Department of Computer Engineering,
Faculty of Engineering, Chulalongkorn University
Peerapon.v@chula.ac.th

www.cp.eng.chula.ac.th/~peerapon/



Outline

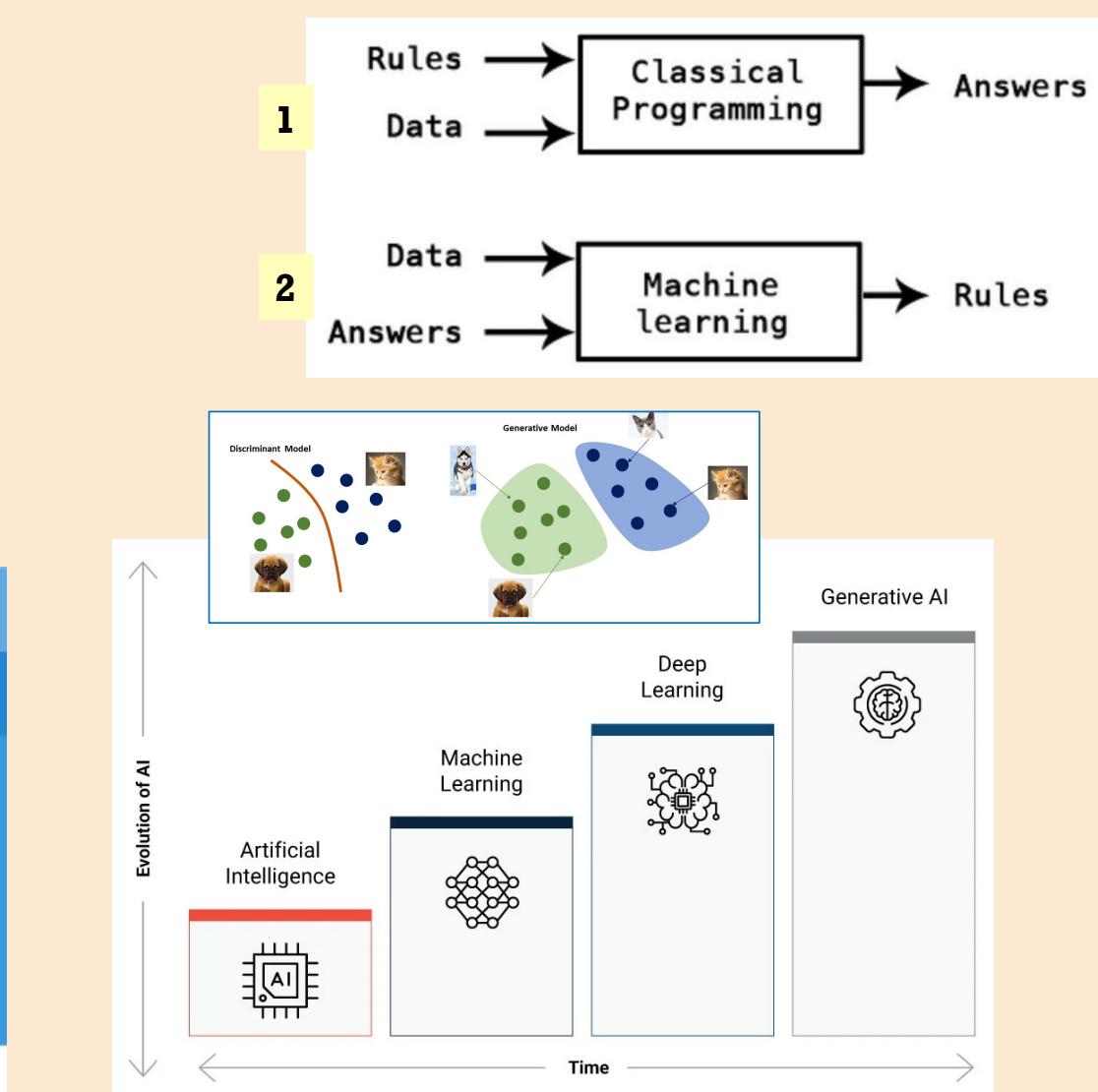
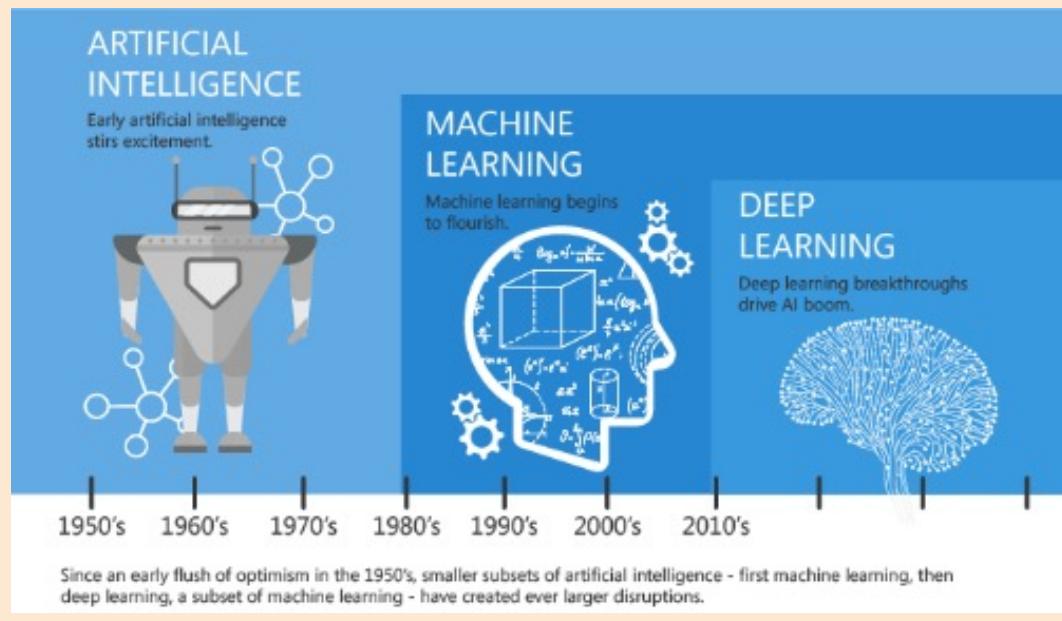
- Introduction to Deep Learning (DL)
- DL applications
 - Image tasks
 - NLP tasks
 - Speech tasks





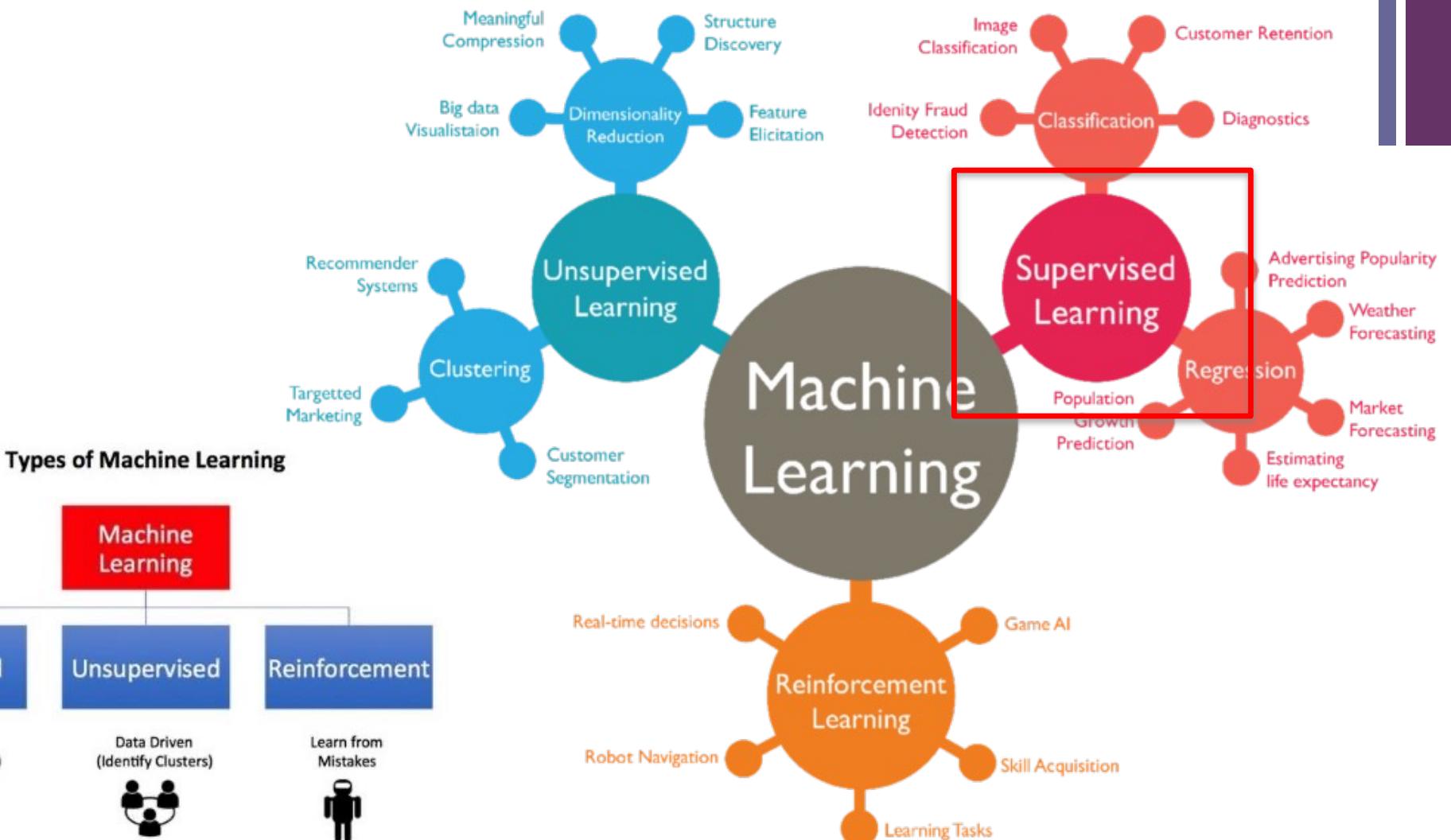
AI = Automation

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



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

+ Machine Learning (ML)





Task1: Supervised learning

Handcrafted features

Training Data



inputs					target
Age	Gender	BodyTemp	Cough	Corona	
12	Female	37	Yes	Yes	
35	Female	39	No	Yes	
32	Male	38	Yes	No	

Testing Data



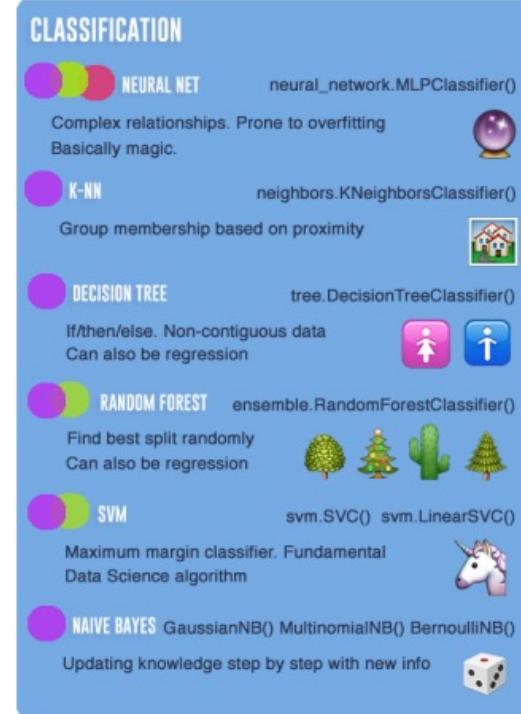
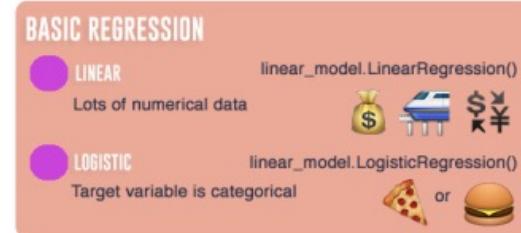
Age	Gender	BodyTemp	Cough	Corona
25	Male	40	No	?

Application: Corona Prediction



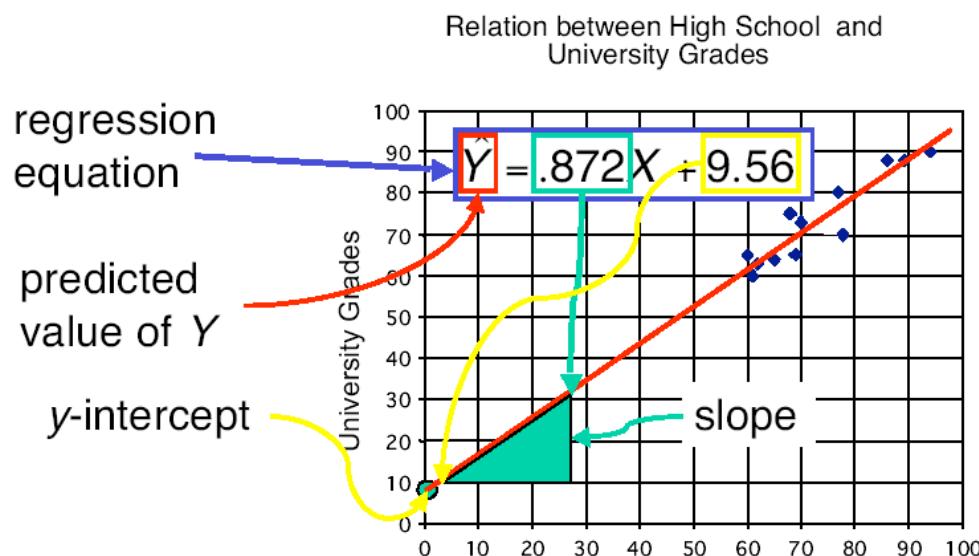
Prediction algorithms

- Decision Tree
- (Logistic) Regression
- kNN
- Support Vector Machine
- Neural Networks (NN)
- Deep Learning





Regression – Linear Relationship



$$\hat{y} = \hat{w}_0 + \hat{w}_1 x_1 + \hat{w}_2 x_2$$

target intercept input

weight, coefficient

- The least square method aims to minimize the following term

$$\sum_{\text{training data}} (y_i - \hat{y}_i)^2$$

+

Logistic Regression (cont.)

Linear Relationship

Training Data



inputs					target
Age	Gender	BodyTemp	Cough	Corona	
12	Female (0)	37	Yes (1)	Yes	
35	Female (0)	39	No (0)	Yes	
32	Male (1)	38	Yes (1)	No	

$$\text{Logit_score} = w_0 + w_1 * \text{Age} + w_2 * \text{Gender} + w_3 * \text{Temp} + w_4 * \text{Cough}$$

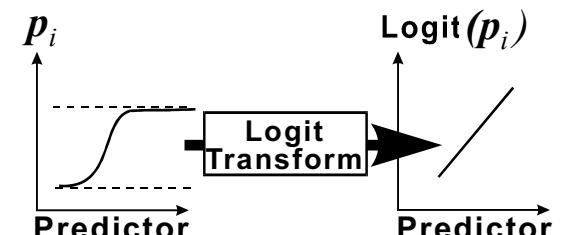
$$\text{Logit_score} = 0.01 - 0.3 * \text{Age} + 0.2 * \text{Gender} + 0.2 * \text{Temp} + 0.9 * \text{Cough}$$

Example

$$\text{Logit_score} = 0.01 - 0.3 * 12 + 0.2 * 0 + 0.2 * 37 + 0.9 * 1 = 4.71$$

prob = 0.9911 → "Yes"

Application: Corona Prediction

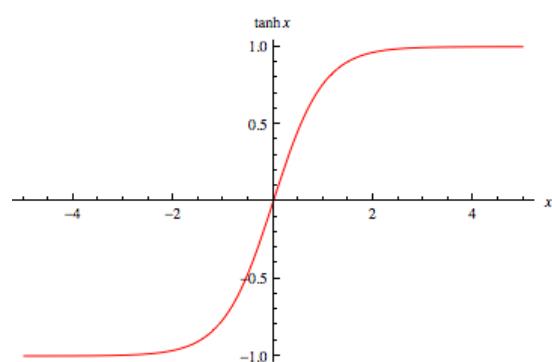
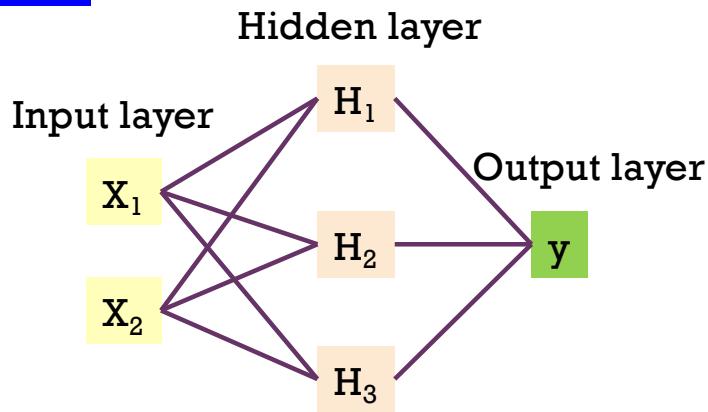


$$\hat{p} = \frac{1}{1 + e^{-\text{logit}(\hat{p})}}$$



Neural Networks (universal approximator)

Non-linear relationship



$$y = \hat{w}_0 + \hat{w}_1 H_1 + \hat{w}_2 H_2 + \hat{w}_3 H_3$$

$$H_1 = \tanh(\hat{w}_{10} + \hat{w}_{11}x_1 + \hat{w}_{12}x_2)$$

$$H_2 = \tanh(\hat{w}_{20} + \hat{w}_{21}x_1 + \hat{w}_{22}x_2)$$

$$H_3 = \tanh(\hat{w}_{30} + \hat{w}_{31}x_1 + \hat{w}_{32}x_2)$$

Stop when?

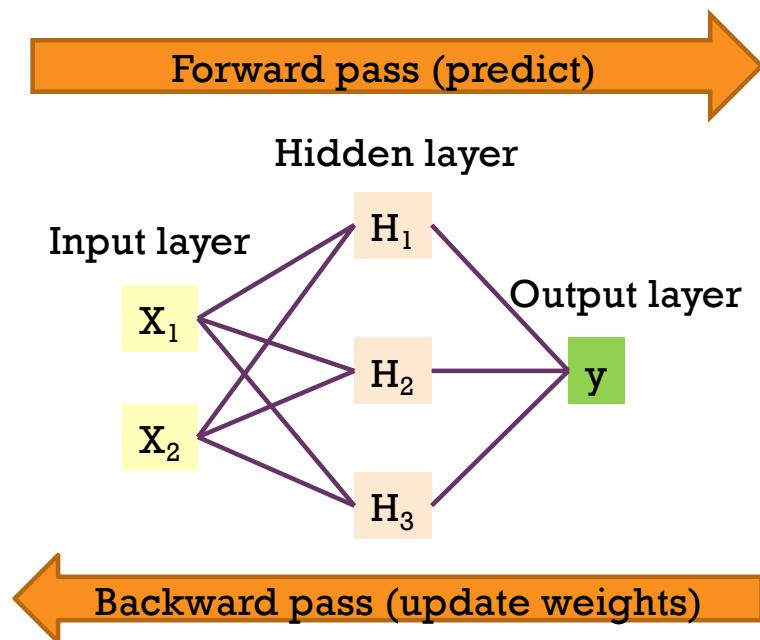
- Converge (no change in loss)
- Max epochs

Important Params:

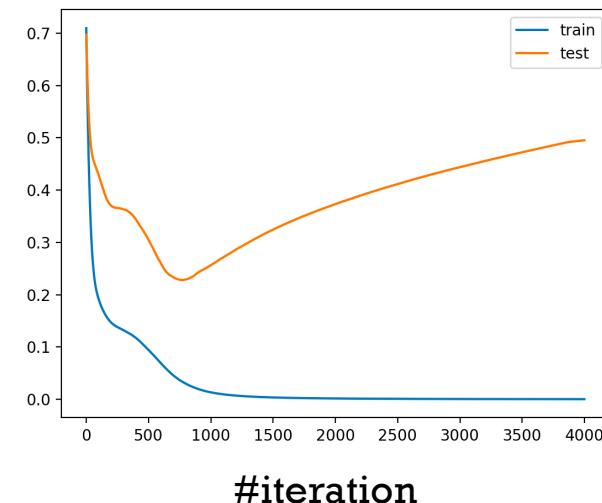
- #hidden units, #hidden layers
- Learning rate, Momentum, decay
- Seed number, etc.



Neural Networks (cont.): Training Non-linear relationship



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



$$y = \hat{w}_0 + \hat{w}_1 H_1 + \hat{w}_2 H_2 + \hat{w}_3 H_3$$

$$H_1 = \tanh(\hat{w}_{10} + \hat{w}_{11} x_1 + \hat{w}_{12} x_2)$$

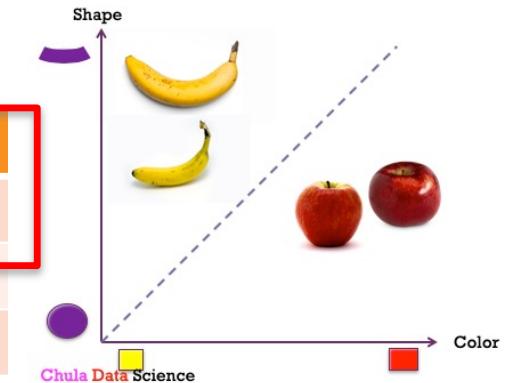
$$H_2 = \tanh(\hat{w}_{20} + \hat{w}_{21} x_1 + \hat{w}_{22} x_2)$$

$$H_3 = \tanh(\hat{w}_{30} + \hat{w}_{31} x_1 + \hat{w}_{32} x_2)$$



Handcrafted features

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



11

Can we still tell the features (columns)?



shutterstock.com - 451802557



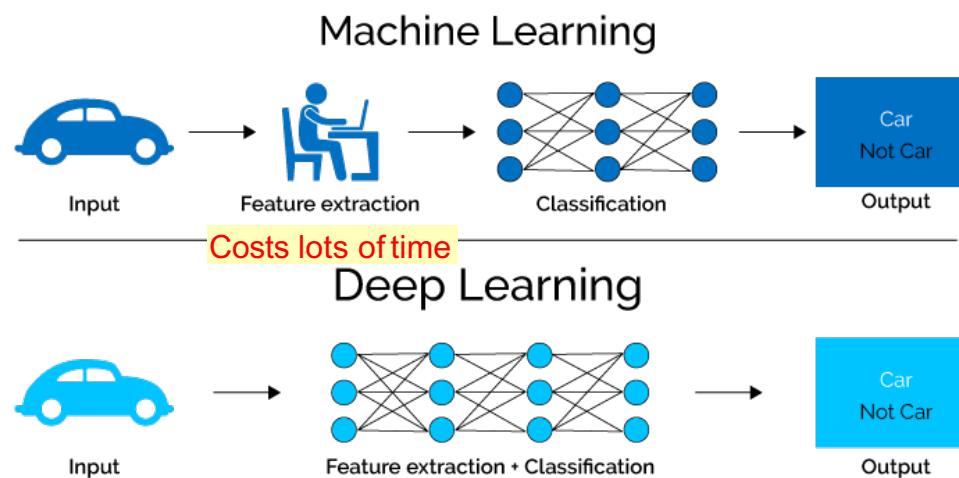
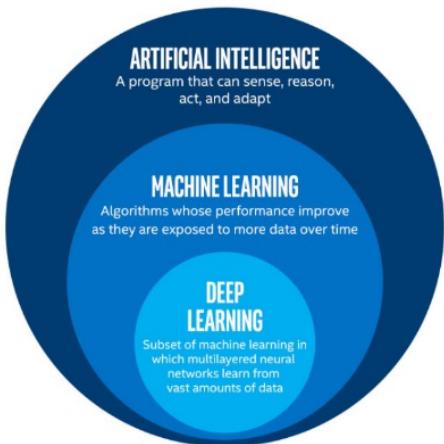
What is Deep Learning (DL)?



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.

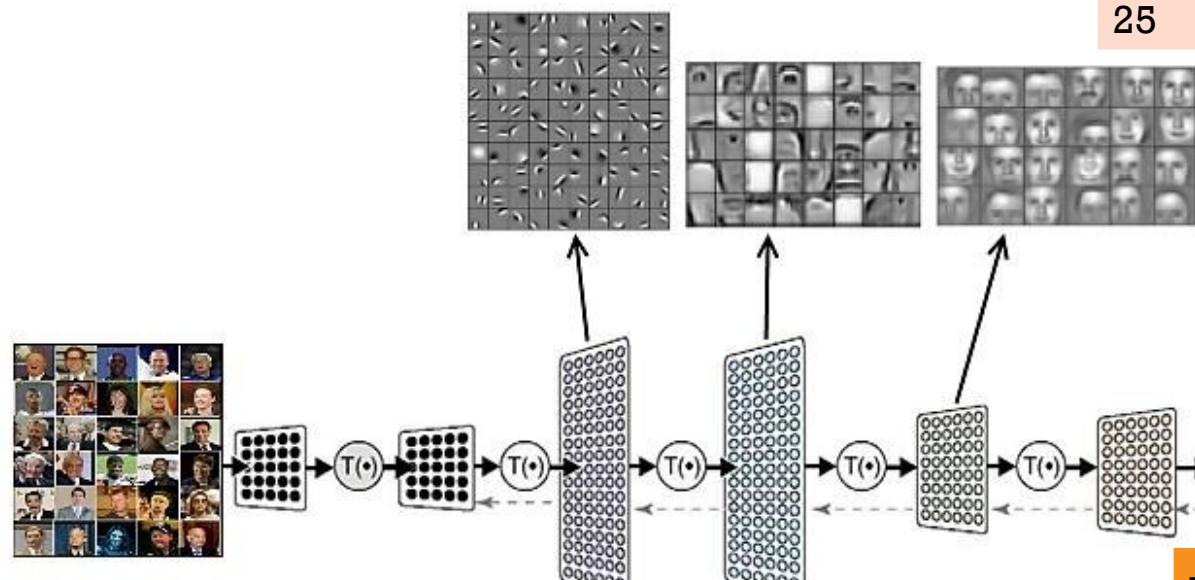




Deep Learning – Basics (cont.)

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.



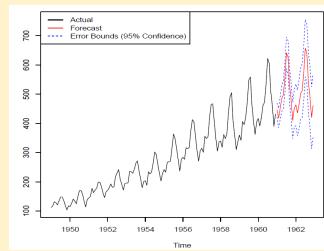
Age	Income	Gender	Province	Corona
25	25,000	Female	Bangkok	Yes

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

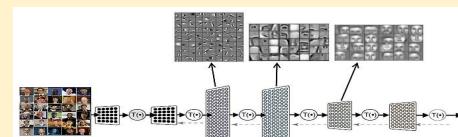
Deep Learning Application



Speech
Recognition



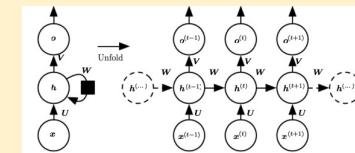
Computer
Vision



CNN



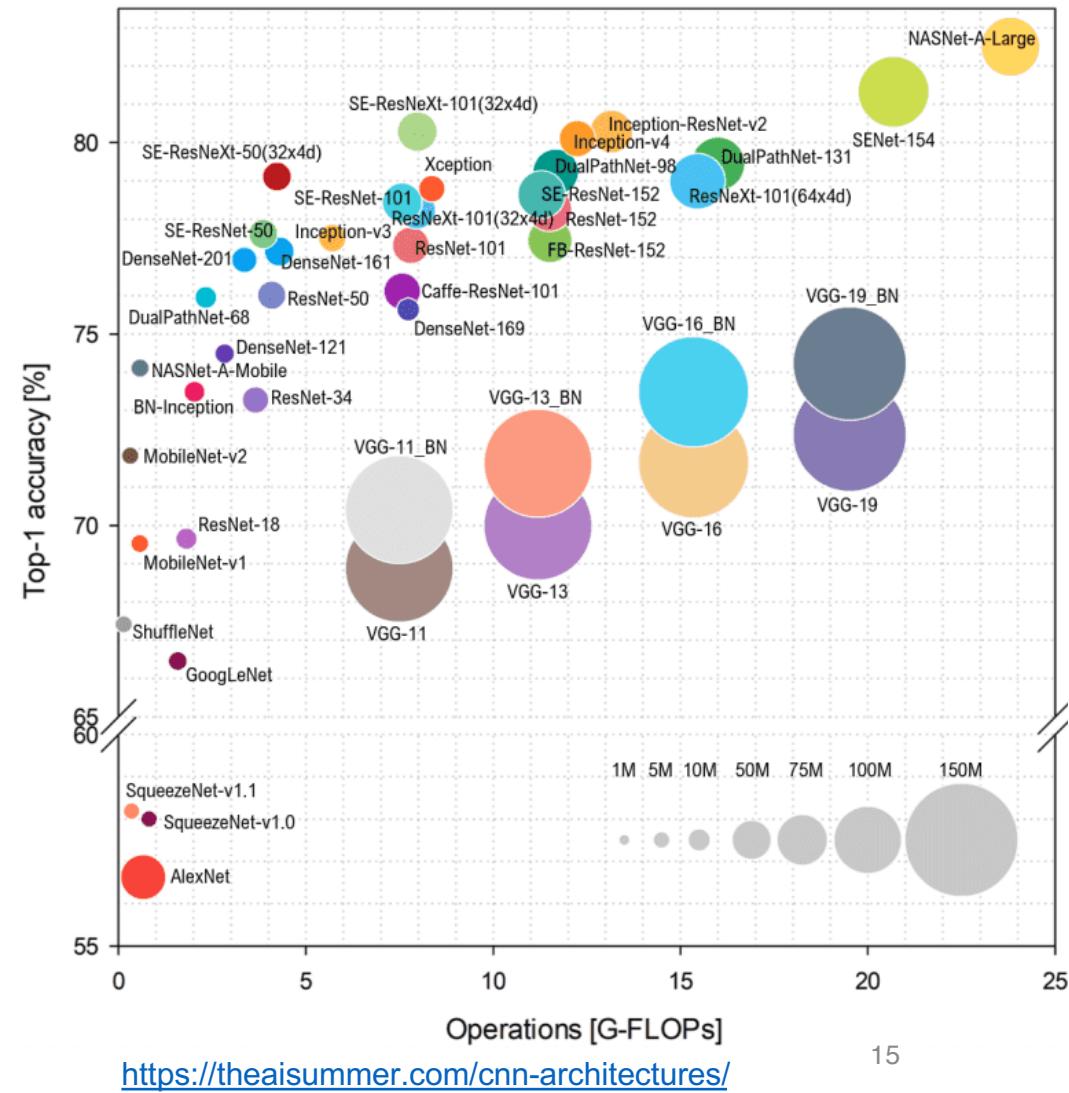
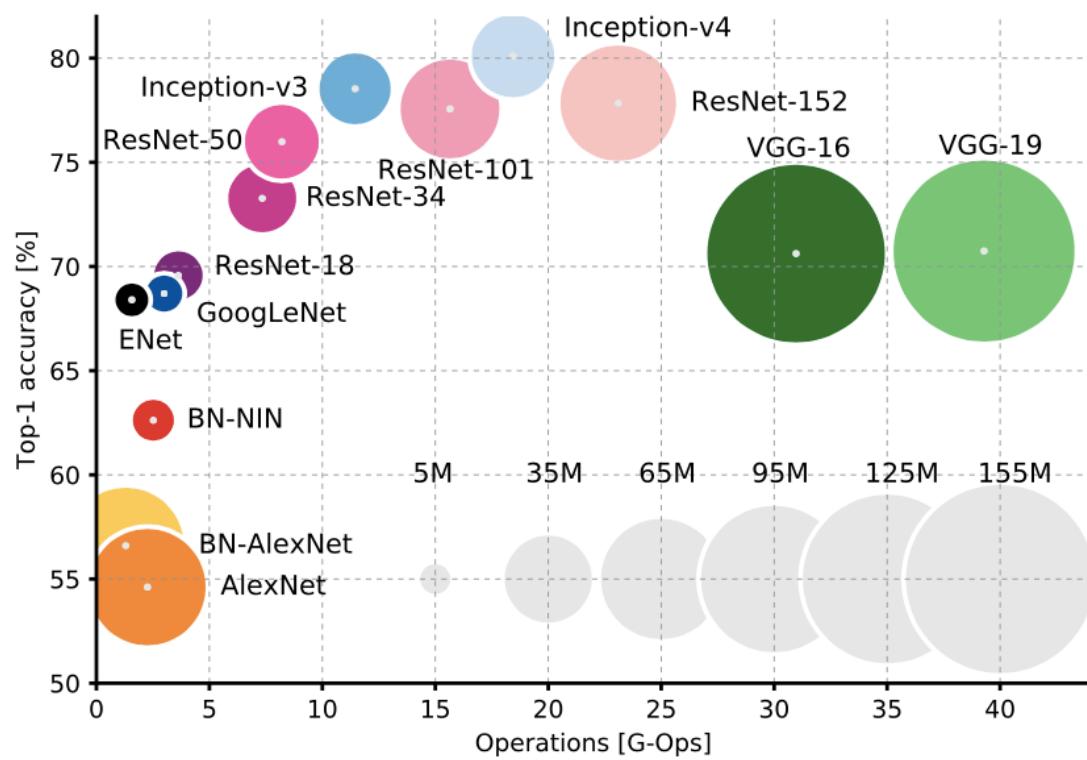
Natural Language
Processing



RNN (LSTM)

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

SOTA of Image Classification

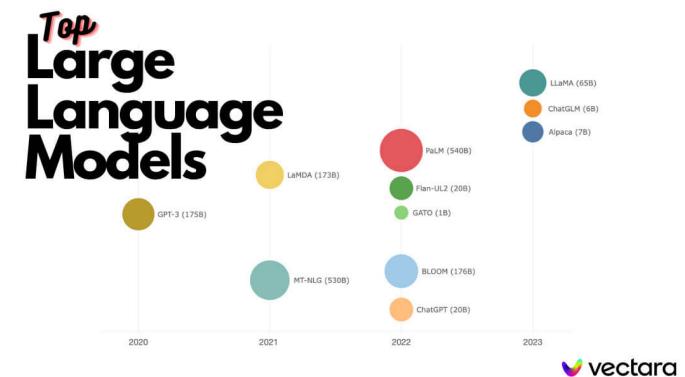
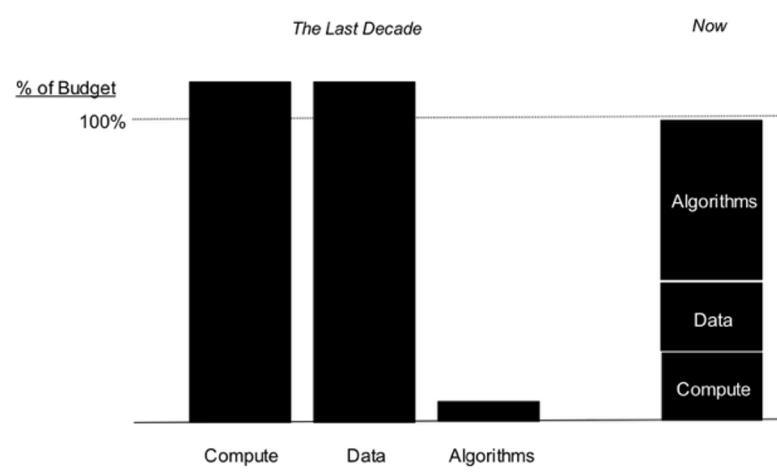


https://blog.csdn.net/qq_34216467/article/details/83061692

<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

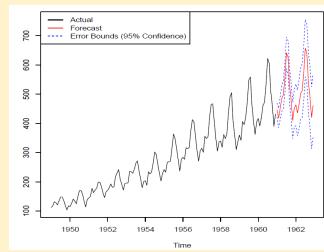


Based on the slide of Aj.Ekapol [/www.kdnuggets.com/2017/06/practical-guide-machine-learning-understand-differentiate-apply.html](http://www.kdnuggets.com/2017/06/practical-guide-machine-learning-understand-differentiate-apply.html)

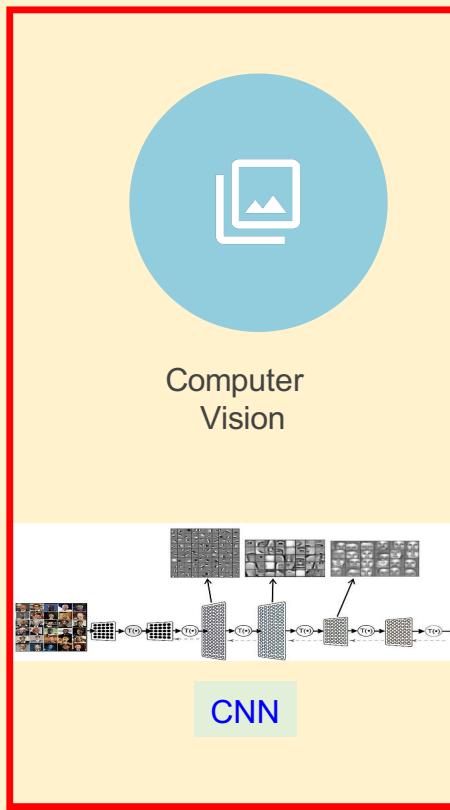
Deep Learning Application



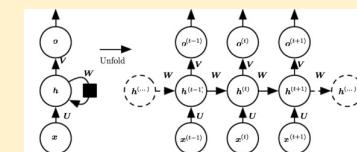
Speech
Recognition



Computer
Vision



Natural Language
Processing



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

YOLO: Release Dates Timeline



YoloV5 (2020) → YoloV8 (2023) → YoloV10 (2024)

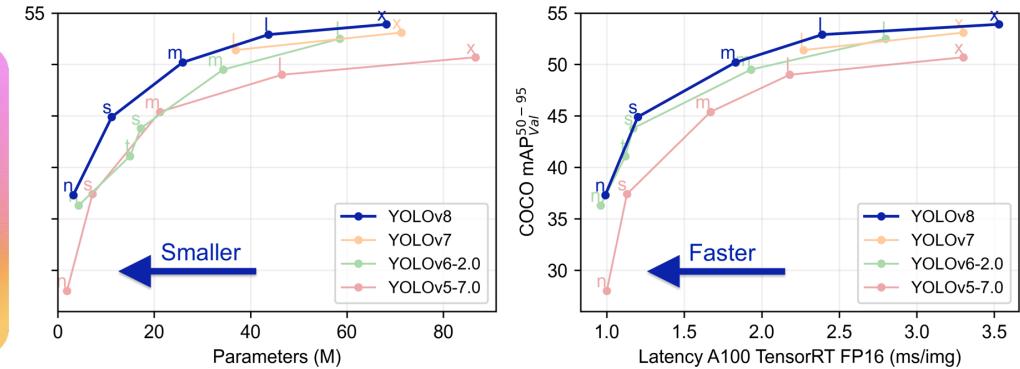


The screenshot shows the Ultralytics YOLOv8 landing page. It features a large logo with three white circles and the text "ultralytics YOLOv8". Below the logo is a QR code with the text "DOWNLOAD THE APP". At the bottom, there are links for "English | 简体中文", "Ultralytics CI passing", "codecov 87%", "DOI 10.5281/zenodo.7347926", "docker pulls 22k", "Run on Gradient", "Open in Colab", and "Open in Kaggle".

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.

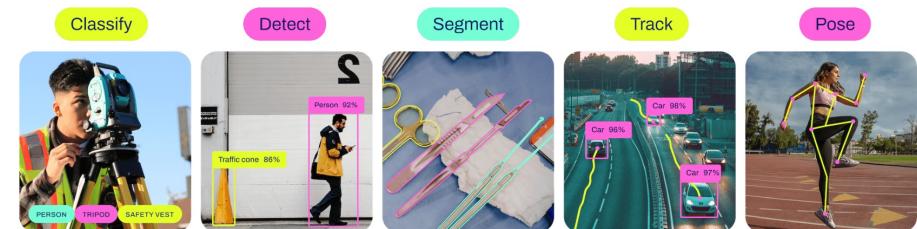
We hope that the resources here will help you get the most out of YOLOv8. Please browse the YOLOv8 [Docs](#) for details, raise an issue on [GitHub](#) for support, and join our [Discord](#) community for questions and discussions!

To request an Enterprise License please complete the form at [Ultralytics Licensing](#).



Models

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.



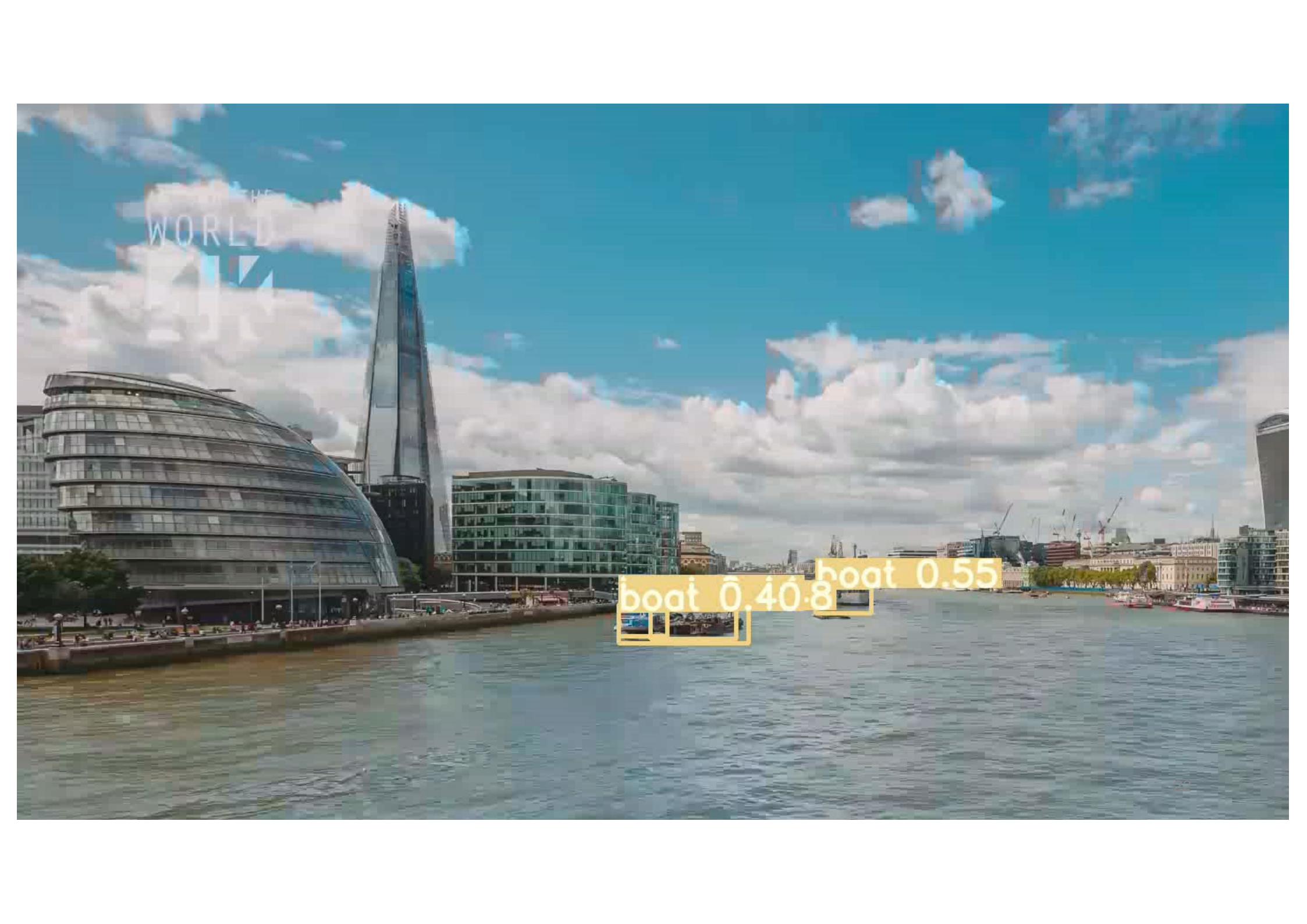
All [Models](#) download automatically from the latest Ultralytics [release](#) on first use.



PEOPLE COUNTING



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



WORLD

boat 0.40.8 boat 0.55

In Frame: 3
Identified: 3



ID: Orn



DB: 88 (img)

ID: Cherprang



DB: 176 (img)

ID: Pun



DB: 48 (img)

ตัวอย่างความลับบุกวันนี้



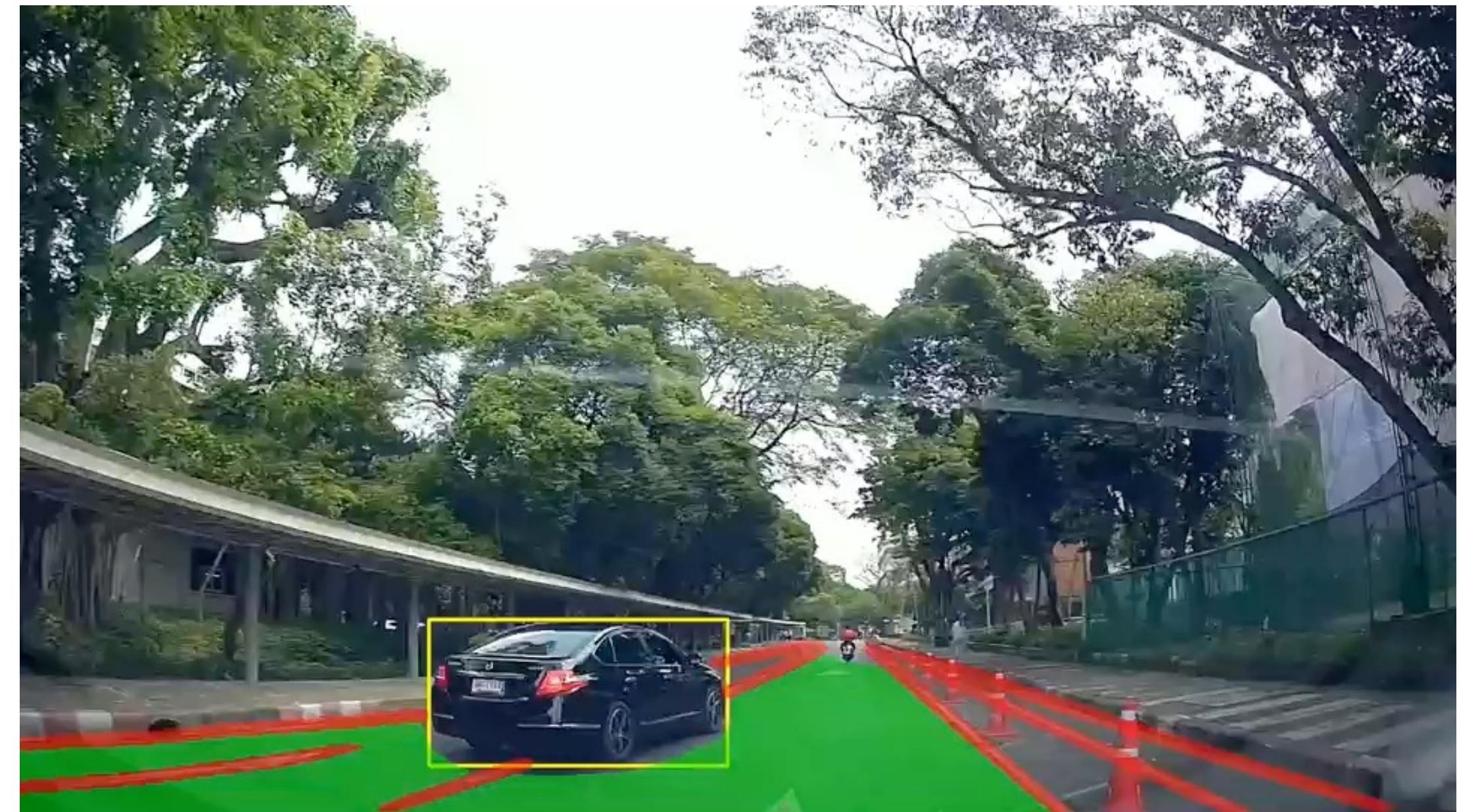
www.youtube.com/workpointofficial

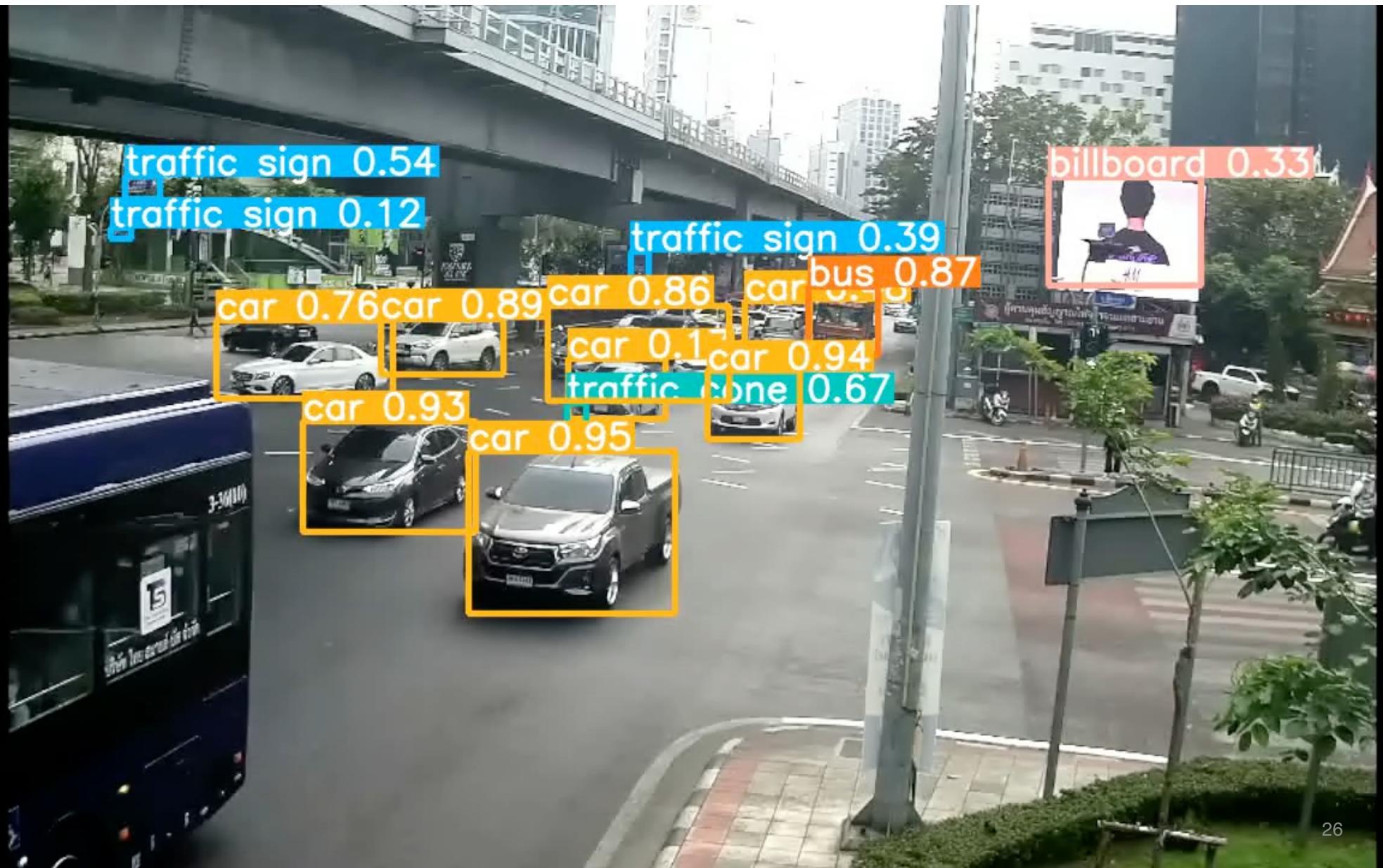


traffic cone 0.87

traffic cone 0.85

traffic cone 0.89





Meta เพชร Segment Anything โมเดล AI สำหรับแยกและวัดตุ่นภาพ-วิดีโอ แม้ไม่เคยเห็นมาก่อน

By: arjin  on 6 April 2023 - 07:16

Tags: Meta Computer Vision Artificial Intelligence



Meta เพชรโครงการ Segment Anything โมเดล AI สำหรับงานแยกและวัดตุ่นภาพและวิดีโอ (Segmentation) มีจุดเด่นคือความสามารถในการแยกและวัดตุ่นต่าง ๆ แม้จะไม่เคยเห็นให้รู้สึกวัดถูกน้ำก่อน และมาพร้อมเครื่องมือที่วัดเส้นขอบตุ่นให้อัตโนมัติ

ในงานที่เพชรนี้ Meta นำเสนอดังข้างใต้แก่ Segment Anything Model (SAM) โมเดลสำหรับการแยกและวัดตุ่น เพย์แพร์ภายใน Apache 2.0 และชื่ออยู่ 1 พันล้านดอลลาร์ เช่น สำหรับงาน Segmentation (SA-1B) อนุญาตให้ใช้สำหรับงานวิจัย

Meta นอกจาก Segmentation เป็นงานด้านซึ่งของ Computer Vision ที่ต้องอาศัยการเรียนรู้ข้อมูลพื้นฐานก่อน แต่ด้วยชุดโมเดล SAM นี้ จะสามารถถูกนำไปประยุกต์ลงพัฒนาให้หลากหลายกันที่ เช่น การแยกและวัดตุ่นแบบเรียลไทม์ใน AR/VR, การนำระบบจับวัดตุ่นไปใช้กับแอปตัดต่อวิดีโอ หรือในงานวิจัยอื่น ๆ

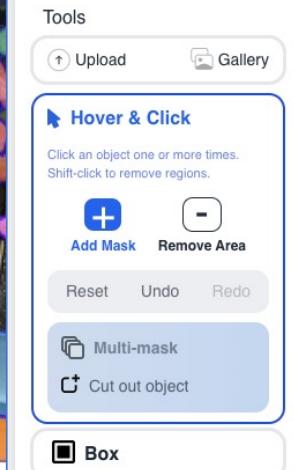
Meta ยังสร้างเว็บสำหรับทดลองใช้โมเดล Segment Anything ที่อยู่ใน

ที่มา: Meta



Segment Anything

Research by Meta AI



Home Demo Dataset Blog Paper

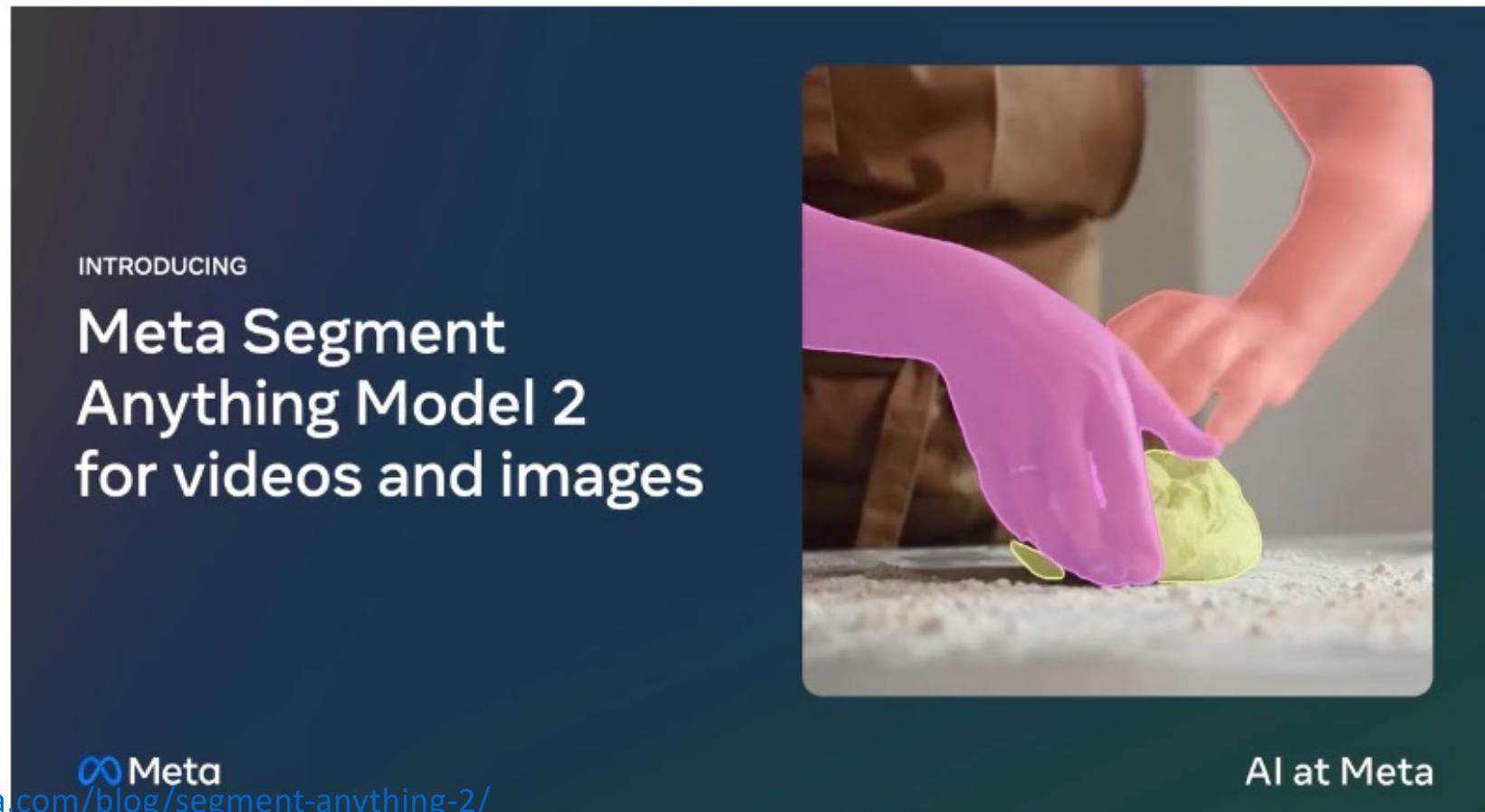
<https://segment-anything.com/demo#>

Cut out the selected object, or try multi-mask mode.



Introducing SAM 2: The next generation of Meta Segment Anything Model for videos and images

July 29, 2024 • 15 minute read



INTRODUCING

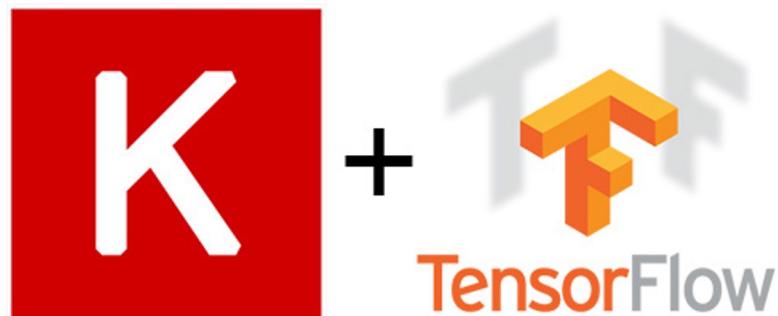
Meta Segment Anything Model 2 for videos and images

 Meta

<https://ai.meta.com/blog/segment-anything-2/>

AI at Meta

Deep Learning tools



Tools for Computer Vision

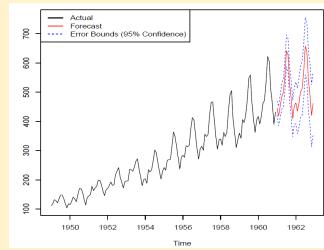
The screenshot shows the TensorFlow API documentation for Keras applications. The left sidebar lists various models under the 'applications' category, with 'Overview' selected. The main content area is titled 'Modules' and lists several modules: `densenet`, `efficientnet`, `imagenet_utils`, `inception_resnet_v2`, `inception_v3`, `mobilenet`, `mobilenet_v2`, `nasnet`, `resnet`, `resnet50`, `resnet_v2`, `vgg16`, and `vgg19`. To the right, a code snippet demonstrates how to build a VGG19 model with two additional dense layers and an output layer.

```
1 base_model = VGG19(weights=None, include_top=False, input_shape=(224, 224, 3))
2
3 for layer in base_model.layers:
4     layer.trainable = True
5
6 x = base_model.output
7 x = Flatten()(x)
8 x = Dense(1024)(x)
9 x = Dropout(0.5)(x)
10 x = Dense(512)(x)
11 x = Dropout(0.5)(x)
12 output = Dense(num_class, activation='softmax')(x)
13
```

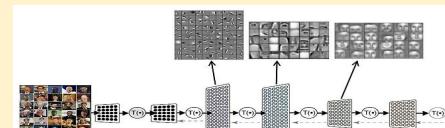
Deep Learning Application



Speech
Recognition



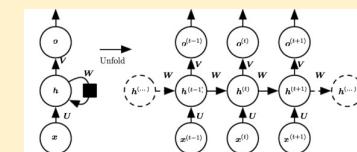
Computer
Vision



CNN



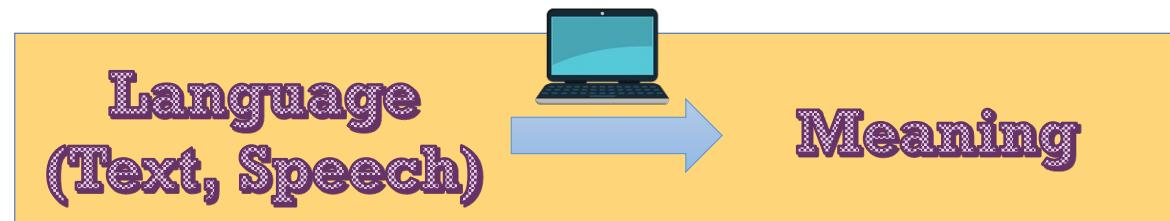
Natural Language
Processing



RNN (LSTM)

Introduction: Natural Language Processing (NLP)

- Subfield of AI
- GOAL:



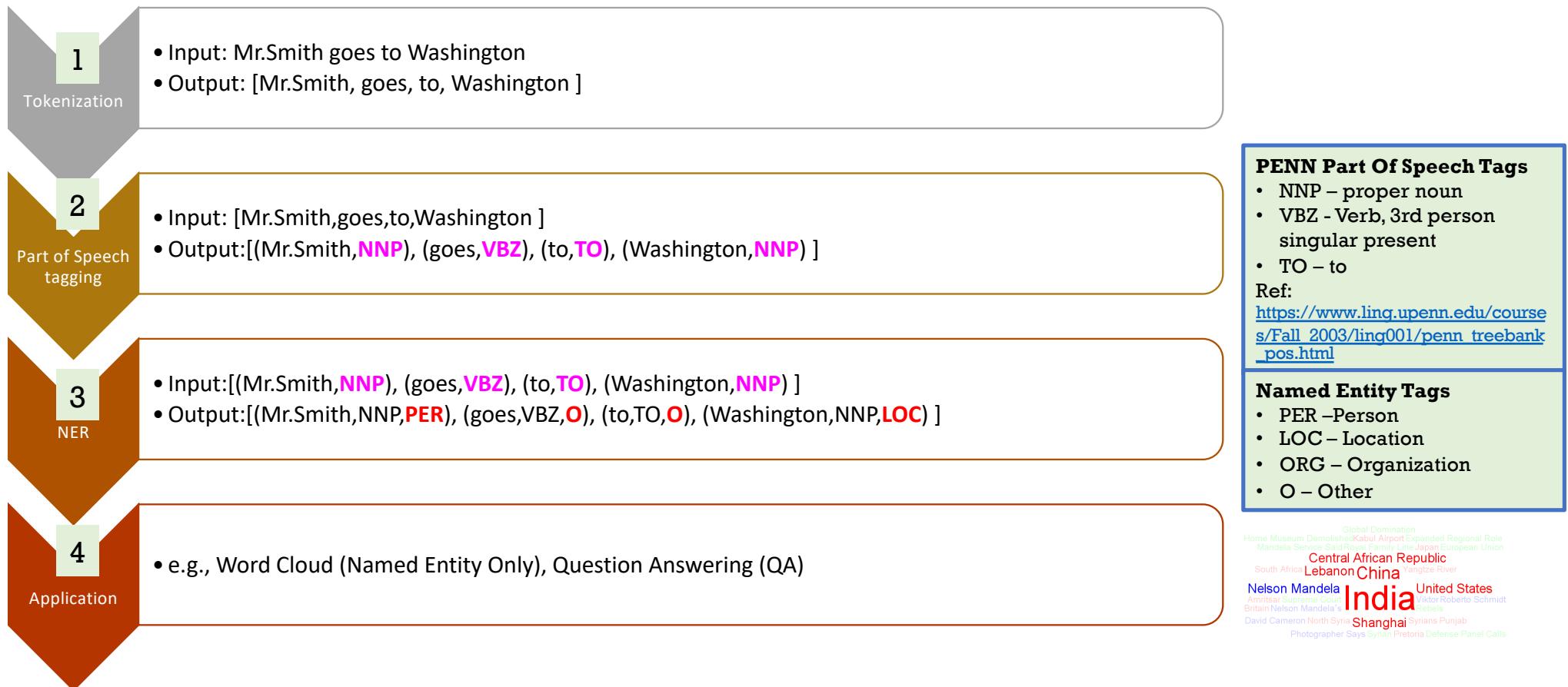
- Bridge the gap between **how people communicate** and **what machines understand** in order to perform useful tasks, e.g.
 - Making appointments, buying things, question answering, etc.



Goal: intelligent processing of human languages

- Not just string matching

Intro (cont.): NLP Pipeline Example (English)



NLP today: Machine Translation (MT)

Google google translate

All Images Maps News Videos More Settings Tools

About 1,180,000,000 results (0.39 seconds)

English ▾ Thai ▾

As the new year gets underway, expert commentators give their view on what 2018 holds in store.

Here are three big themes to watch out for over the next 12 months.

Can the stock market rally go on? The new year has begun with stock markets in the UK and US hitting new record highs.

The Dow Jones Industrial Average rose above 25,000 points for the first time this week, while the broader S&P 500 is also at historic highs.

เป็นปีใหม่ที่กำลังได้รับการแสดงความคิดเห็นของผู้เชี่ยวชาญให้มุมมองของพากษาเกี่ยวกับสิ่งที่ 2018 เก็บไว้ในร้าน

ต่อไปนี้เป็นหัวข้อใหญ่สามข้อที่ควรระวังในช่วง 12 เดือนข้างหน้า

การซัมมูมตลาดหุ้นสามารถดำเนินต่อไปได้หรือไม่?

ปีใหม่เริ่มมีตลาดหุ้นในสหรัฐอาณาจักรและสหราชอาณาจักรสูงเป็นประวัติการณ์

ดัชนีเฉลี่ยอุตสาหกรรมดาวโจน斯ปรับตัวสูงขึ้นกว่า 25,000 จุดเป็นครั้งแรกในสัปดาห์นี้ขณะที่ดัชนี S & P 500 ที่ใหญ่ขึ้นก็อยู่ในระดับสูงเป็นประวัติการณ์

Markets, Brexit and Bitcoin: 2018's themes

By Chris Johnston
Business reporter

5 January 2018

f t m Share

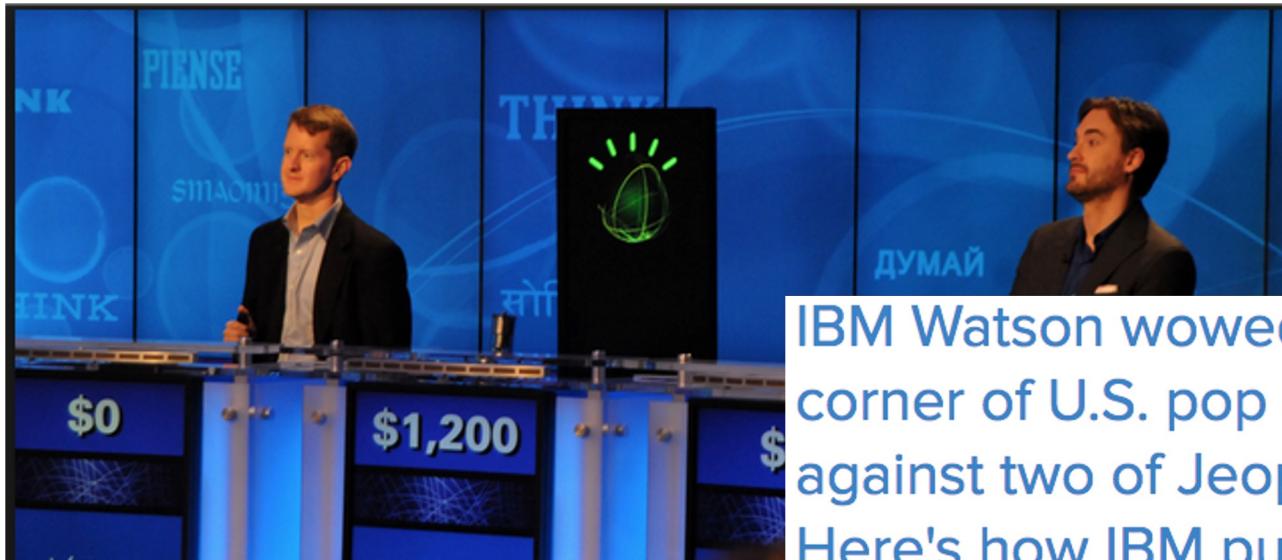


GETTY IMAGES

As the new year gets underway, expert commentators give their view on what 2018 holds in store.

<http://www.bbc.com/news/business-42581934>

NLP today: Question Answering (QA)



IBM Watson wowed the tech industry and a corner of U.S. pop culture with its 2011 win against two of Jeopardy's greatest champions. Here's how IBM pulled it off and a look at what Watson's real career is going to be.

<https://www.techrepublic.com/article/ibm-watson-the-inside-story-of-how-the-jeopardy-winning-supercomputer-was-born-and-what-it-wants-to-do-next/>

Ref: Prof. Regina Barzilay, NLP @MIT

ChatGPT (Generative AI)

<https://chat.openai.com/>

The screenshot shows a ChatGPT session with the following exchange:

ข้อรายละเอียดบริษัท BJC Thailand

BJC Thailand เป็นบริษัทค้าปลีกและผู้ผลิตอาหารและเครื่องดื่มชั้นนำในประเทศไทย ก่อตั้งขึ้นในปี พ.ศ. 2517 (1974) โดยมีกลุ่มธุรกิจที่เกี่ยวข้องกับกลุ่มธุรกิจ Charoen Pokphand Group เป็นผู้ถือหุ้นส่วนสูงสุดในบริษัท

BJC มีธุรกิจหลัก 4 กลุ่ม ได้แก่

- ธุรกิจค้าปลีก: ชูปเปอร์มาร์เก็ต, ร้านสะดวกซื้อ, ร้านค้าออนไลน์ เช่น Big C, Tesco Lotus, ฯลฯ
- ธุรกิจผลิตอาหารและเครื่องดื่ม: ผลิตและจ้าหาน้ำของอาหารและเครื่องดื่ม เช่น น้ำอัดลม, น้ำดื่ม, น้ำผลไม้, อาหารสำเร็จรูป เช่น Oishi, Green Spot, Masita, Malee, Del Monte, UFC
- ธุรกิจโรงแรมและสปา: มีโรงแรมและสปาในหลายจังหวัดของไทย รวมถึงเกาะสมุยและเชียงใหม่ ซึ่งมีธุรกิจเกี่ยวกับการจัดตั้งงานประชุม และงานอิเวนท์ต่างๆ
- ธุรกิจที่เกี่ยวข้องกับการพัฒนาอสังหาริมทรัพย์: เช่น การลงทุนในโครงการอสังหาริมทรัพย์ที่เกี่ยวข้องกับธุรกิจของบริษัท

ในปี 2563 (2020) BJC มีรายได้จากการขายสุทธิ 408,695 ล้านบาท และมีพนักงานประมาณ 140,000 คน โดยมีการขยายธุรกิจไปยังต่างประเทศ

Stop generating

แล้วเกี่ยวกับ Big C ยังไง?

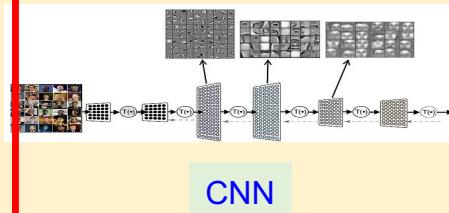
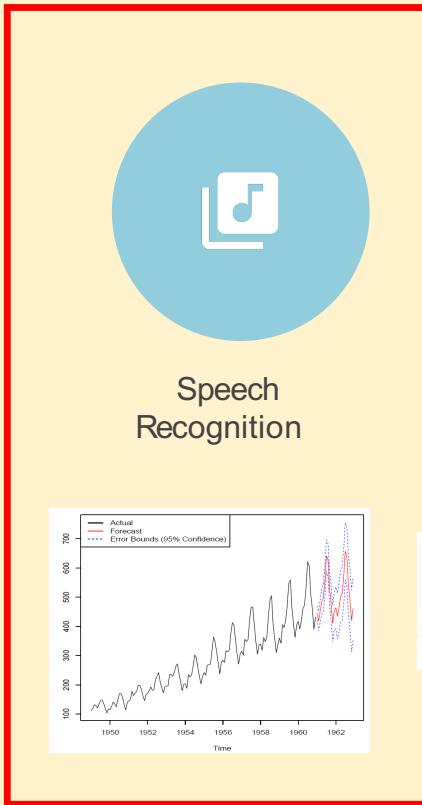
Big C เป็นธุรกิจค้าปลีกของ BJC Thailand ซึ่งเป็นชูปเปอร์มาร์เก็ตในลักษณะของ "hypermarket" โดยจำหน่ายสินค้าหลากหลายประเภท เช่น เครื่องใช้ไฟฟ้า, เครื่องดื่มน้ำ, เครื่องสำอาง, เครื่องแต่งกาย, ผลิตภัณฑ์เพื่อสุขภาพ, ผลิตภัณฑ์อาหารสัตว์, ผลิตภัณฑ์ไอที, รถยนต์, รถจักรยานยนต์, อุปกรณ์กีฬา, อุปกรณ์เดินทาง และสินค้าอื่นๆ อีกมากมาย โดย Big C มีสาขาทั้งในประเทศไทยและต่างประเทศ และมีการขยายธุรกิจออนไลน์ของบริการจัดส่งสินค้าด้วย ในปี 2563 (2020) ธุรกิจค้าปลีกของ BJC Thailand รายการ Big C มีรายได้จากการขายสุทธิ 165,464 ล้านบาท และมีจำนวนสาขาทั้งหมด 147 สาขา โดยมีรายการค้าปลีกอื่นๆ เช่น Tesco Lotus, Makro, และธุรกิจค้าปลีกออนไลน์อย่าง Lazada และ JD Central อยู่ในกลุ่มธุรกิจค้าปลีกของ BJC Thailand ด้วย

Regenerate response

Send a message.

ChatGPT^[a] is an artificial intelligence (AI) chatbot developed by OpenAI and released in November 2022. It is built on top of OpenAI's GPT-3.5 and GPT-4 foundational large language models (LLMs) and has been fine-tuned (an approach to transfer learning) using both supervised and reinforcement learning techniques.

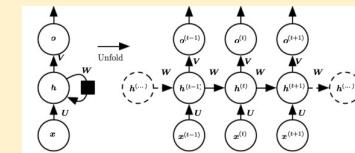
Deep Learning Application



CNN



Natural Language Processing

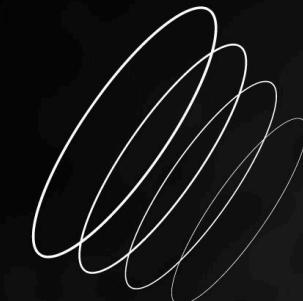


RNN (LSTM)

Thonburian Whisper



 GOWAJEE



GOWAJEE API
LAUNCH



Spaces

biodatlab/whisper-thai-demo

like 11

Running on T4

Thonburian Whisper Demo 🇹🇭

Thonburian Whisper



Transcribe Audio

Transcribe YouTube

Transcribe long-form microphone or audio inputs with the click of a button! Demo uses the fine-tuned checkpoint [biodatlab/whisper-th-medium-combined](#) and 😊 Transformers to transcribe audio files of arbitrary length.

Microphone Input

Record

Default - MacBook ...

Transcription Output

Audio File Upload



Drop Audio Here

- or -

Click to Upload

https://huggingface.co/spaces/biodatlab/whisper-thai-demo?fbclid=IwZXh0bgNhZW0CMTEAAROLKUeVwIMoRPCpepgMA00j0jtvrVfLBEWetTzZQshg8a7_oq1UARafqC4_aem_hgR19w4x48qUtGcwUFRIiQ



Speech To Text

WER: 10.6% CER: 4.7% DER: 13.5%

TH

AI can transcribe speech into text, supporting multiple file types such as mp3, wav, and flac.

Publisher: Gowajee

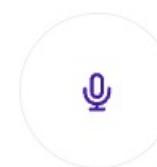
[API Document](#)[Add to Library](#)[Let's try demo](#)[How to call the API?](#)[Full evaluation](#)

We utilized a model from our partner, Gowajee. The model is trained by using over 1,000 hours of annotated data collected online by our partner from various sources. This model can perform in general topics but specializes in call center. The model could be worsened if audio contains code switching, low-quality speech, and overlapping speech data. The model's performance in the term of transcribe speech was evaluated on around 60 hours of speech data, which was collected from the same source as the training data. For diarization, it was evaluated on artificial conversations created from audio pools containing around 50 speakers. Each speaker appears only once in a conversation, ensuring there is no overlap between speakers across conversations.

For more information, visit [Guide Book](#)

Pricing

<https://console.visai.ai/ai-marketplace/ready-to-use/c7ba44f3-0de6-4457-8743-75438e3eac60/speech-to-text>

[Let's try demo](#)[Record audio](#)

You can record audio up to 1 minute

OR

[Click to upload](#) or drag and drop

Text-to-Speech (TTS)

The screenshot shows the GitHub repository page for PyThaiNLP/PyThaiTTS. The top navigation bar includes links for Product, Solutions, Open Source, Pricing, and a search bar. Below the header, the repository name 'PyThaiNLP / PyThaiTTS' is shown with a 'Public' badge. The main content area has tabs for Code (selected), Issues (2), Pull requests, Actions, Projects, Security, and Insights. Under the Code tab, there are buttons for 'dev' (3 branches), 'tags' (2 tags), 'Go to file', and 'Code'. A commit list is displayed, starting with a commit from 'wannaphong' for 'PyThaiTTS v0.1.1' on Aug 27, 2022, with 10 commits. Other files listed include .github/workflows, docs, notebook, pythaitts, .gitignore, LICENSE, README.md, requirements.txt, and setup.py. At the bottom, there is a link to 'README.md'.

<https://github.com/PyThaiNLP/PyThaiTTS>

The screenshot shows the Narakeet website for Thai Text-to-Speech. The header features the Narakeet logo and navigation links for Tools, Voices, Help, Pricing, and News. The main section is titled 'Text to Speech Thai Language' in large red text. Below the title, a sub-headline states: 'Easily convert text to speech in Thai, and 90 more languages. Try our Thai text to speech free online. No registration required.' A 'CREATE AUDIO' button is prominently displayed. The text explains that Thai text to speech voices make it easy to create audio materials and videos with Thai voiceover, faster and cheaper than hiring Thai voices. It notes that Thai voices sound natural, similar to a native speaker, and can create Thai TTS MP3, WAV and M4A files. The text also mentions that Thai, or Central Thai (historically also called Siamese) is the official language of Thailand. It's a tonal language with about 30 million native speakers. Narakeet has 4 Thai text to speech male and female voices. A video player is shown with the text 'Play the video below (with sound) for a quick demo.' Below the text, there is a photograph of a street scene in Thailand with people under colorful umbrellas.

<https://www.narakeet.com/languages/thai-text-to-speech/>