

+



3099704: AI for Digital Health



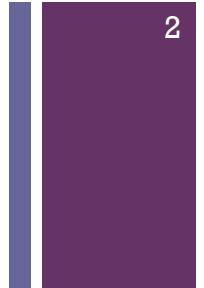
## Advanced Topics: Other Modalities & Model Interpretation

Prof. Peerapon Vateekul, Ph.D.

[Peerapon.v@chula.ac.th](mailto:Peerapon.v@chula.ac.th)



# Outlines



- Introduction
- Speech
- Facial
- Model Interpretation: SHAP value
- Model Interpretation: Grad-CAM

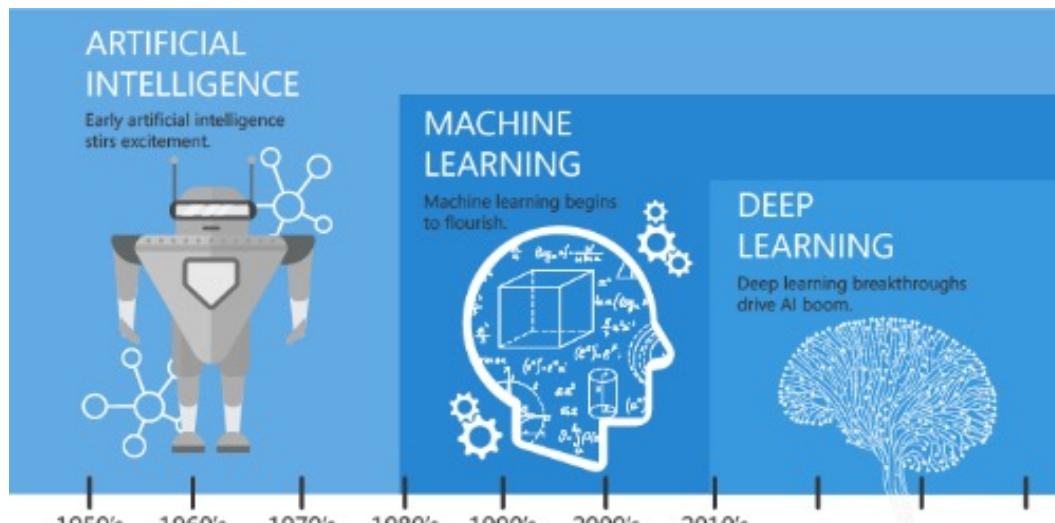


## Introduction

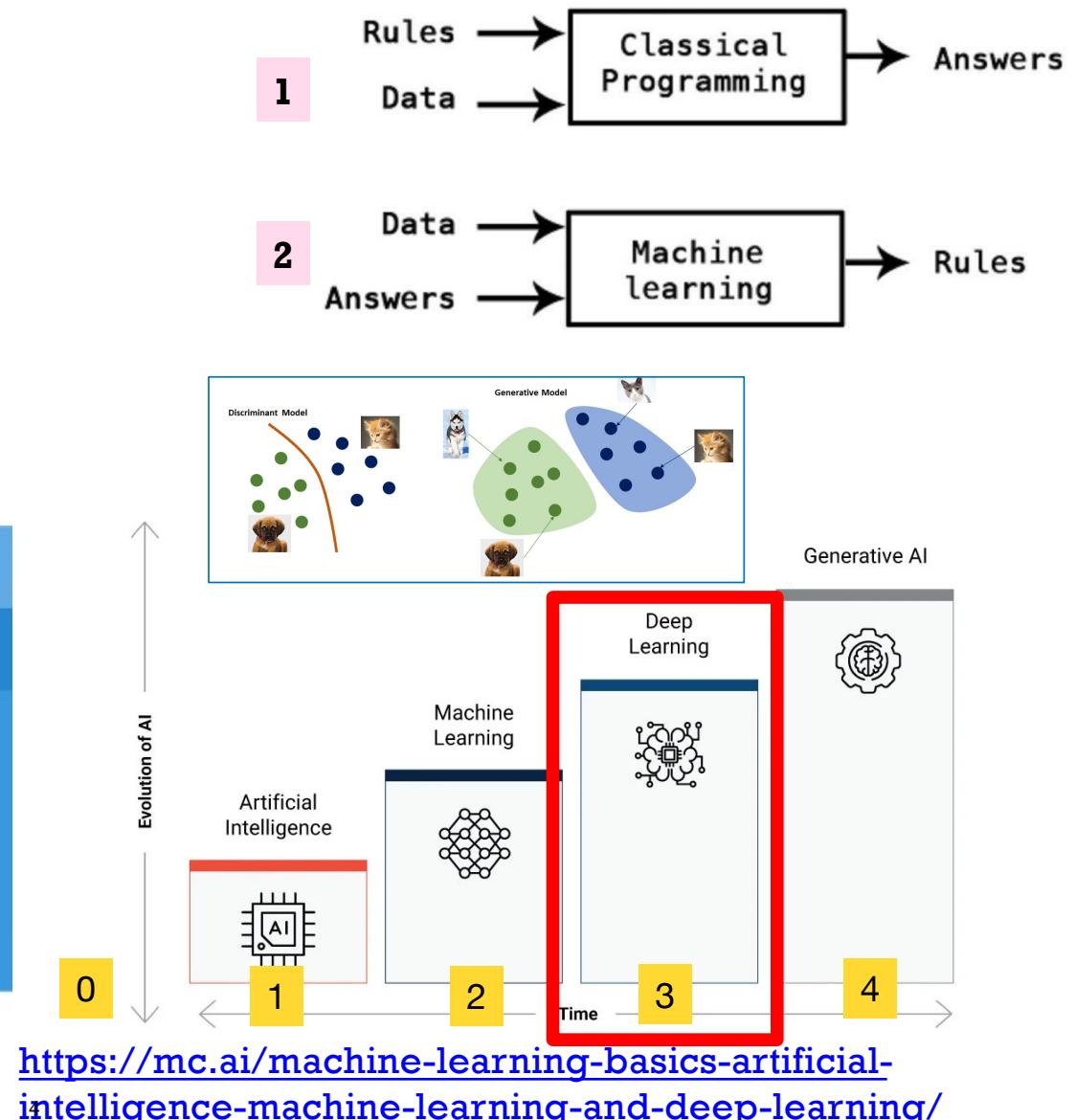


# AI = Automation

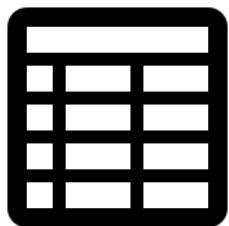
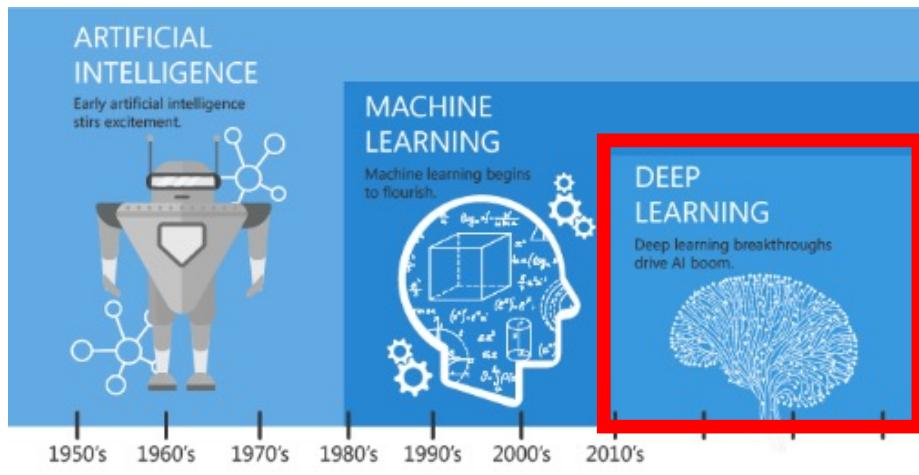
- 0) Not AI Solution (not automatic)
- 1) Rule-based AI
- 2) Machine Learning (ML)



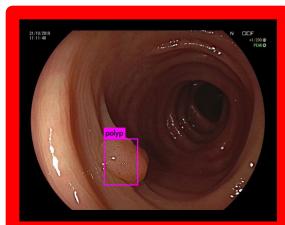
Since an early flush of optimism in the 1950's, smaller subsets of artificial intelligence - first machine learning, then deep learning, a subset of machine learning - have created ever larger disruptions.



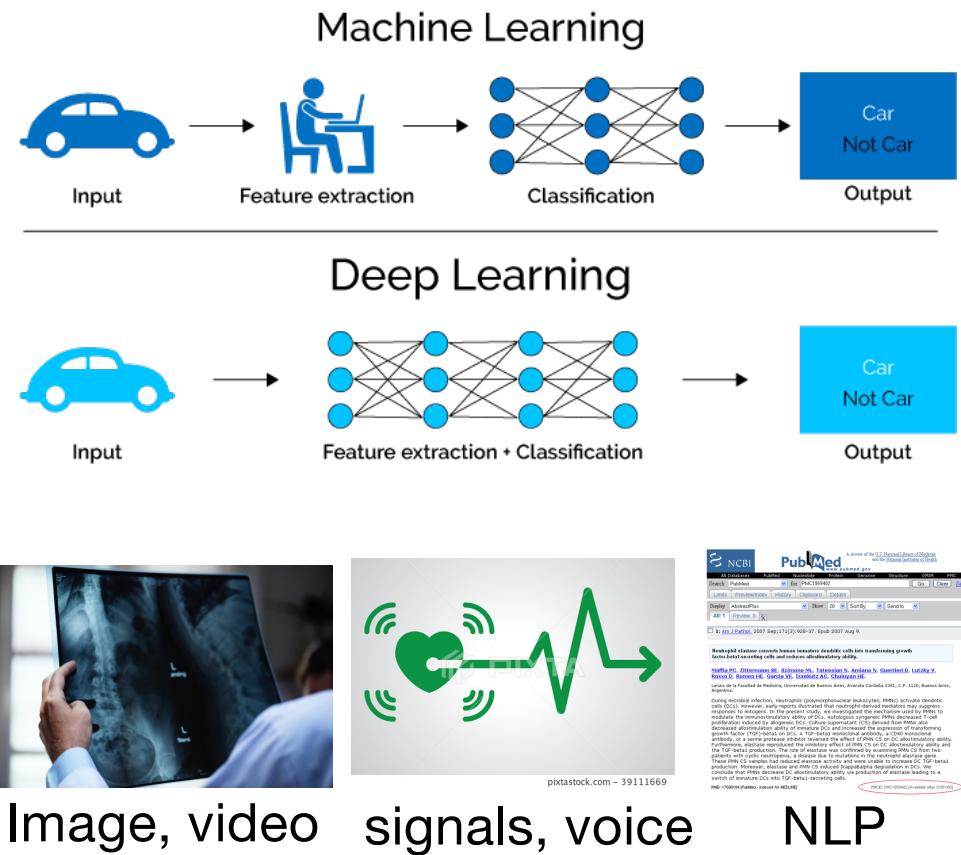
# Arise of Deep Learning



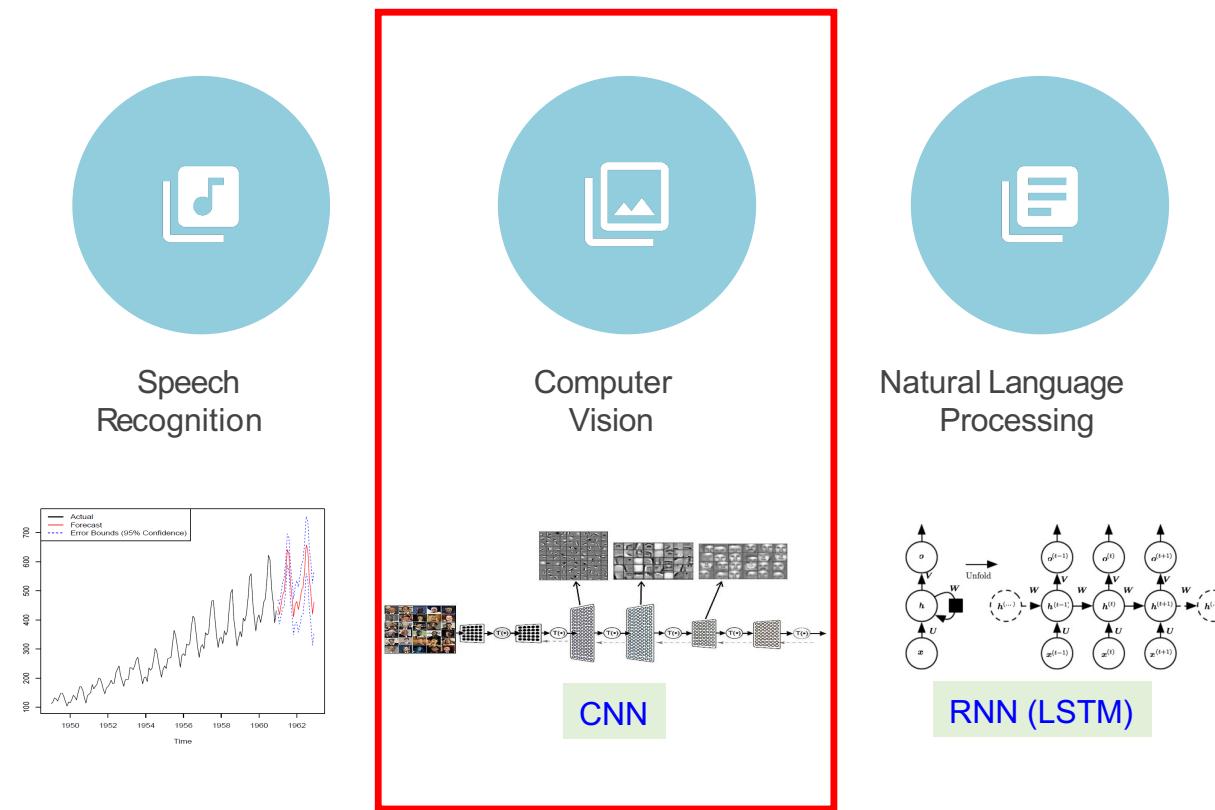
VS



5



# Deep Learning Application



# Type of image tasks

**Semantic Segmentation**



GRASS, CAT,  
TREE, SKY

No objects, just pixels

**Classification + Localization**



CAT

**Object Detection**



DOG, DOG, CAT

Multiple Object

**Instance Segmentation**



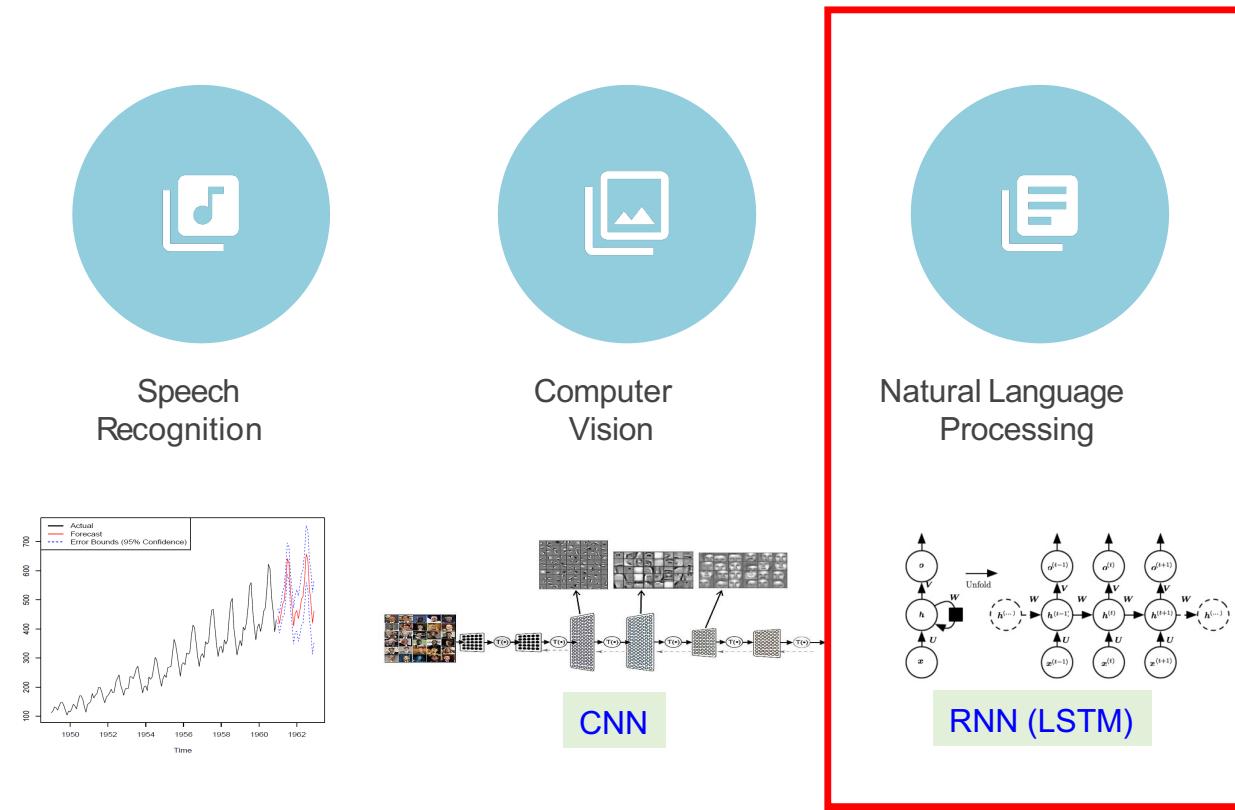
DOG, DOG, CAT

This image is CC0 public domain



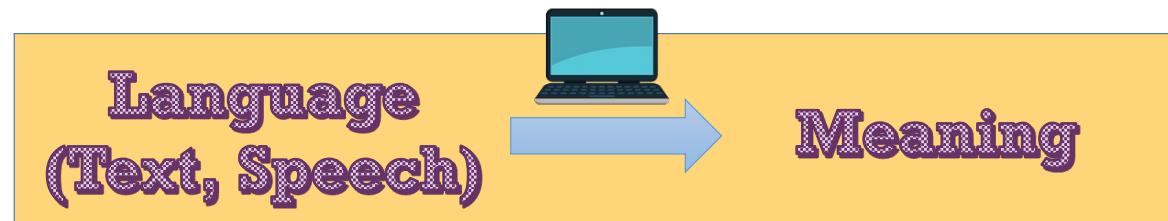
โดยโครงการ D MIND จะใช้เทคนิค Machine Learning

# Deep Learning Application



# 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.

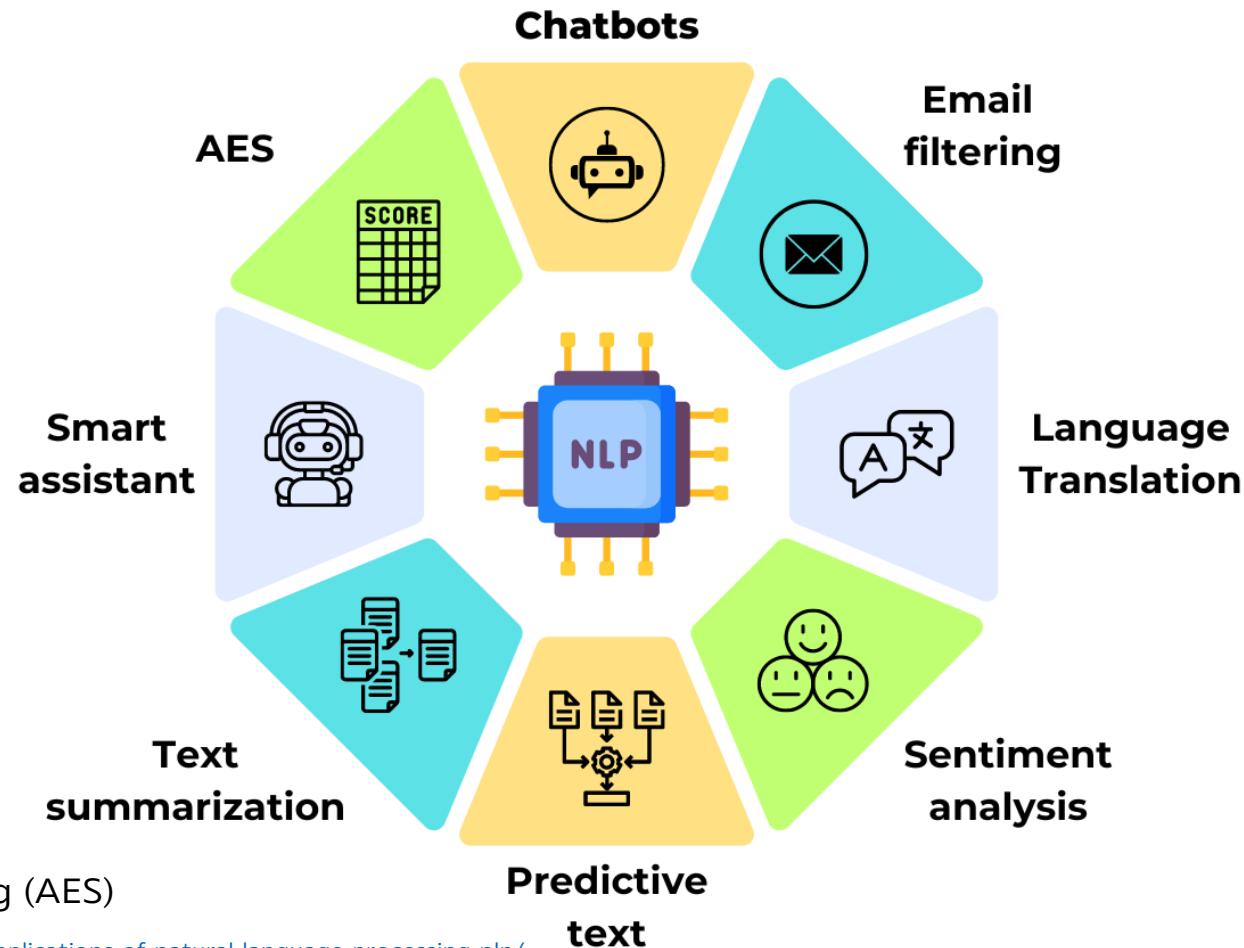


Goal: intelligent processing of human languages

- Not just string matching

## Top 8 Applications of Natural Language Processing (NLP)

# Applications of Natural Language Processing



Automated Essay Scoring (AES)

<https://eastgate-software.com/top-8-applications-of-natural-language-processing-nlp/>

# Tools for Thai NLP



## PyThaiNLP

python 3.6 | pypi v2.1.2 | downloads/month 15k | License Apache 2.0 | license scan | p

code quality A coverage 91% Launch Quick Start Guide on Google Colab DOI

Thai Natural Language Processing in Python.

PyThaiNLP is a Python package for text processing and linguistic analysis.

### News

We are conducting a 2-minute survey to know more about your experience regarding what the library should be able to do. Take part in this survey:

This is a document for development branch (post 2.1). Things will break.

## WangchanBERTa โมเดลประมวล ผลภาษาไทยที่ใหญ่และก้าวหน้า ที่สุดในขณะนี้



VISTEC-depa AI Research Institute of Thailand  
Jan 24 · 5 min read

เปิดให้ทุกคนใช้ฟรีโดยAIResearch.in.th  
BY-SA 4.0



| 28 Dec 2023

### PhayaThaiBERT: Enhancing a Pretrained Thai Language Model with Unassimilated Loanwords

#### Panyut Sriwirote

Department of Linguistics  
Chulalongkorn University  
panyutsriwirote@gmail.com

#### Vasan Timtong

ClickNext  
vasan.t@clicknext.com

#### Abstract

Although WangchanBERTa has become the de facto standard in transformer-based Thai language modeling, it still has shortcomings in regard to the understanding of foreign words, most notably English words, which are often borrowed without orthographic assimilation into Thai in many contexts. We identify the lack of foreign vocabulary in WangchanBERTa's tokenizer as the

Jalinee Thapiang  
ClickNext  
jalinee.t@clicknext.com

Attapol T. Rutherford  
Department of Linguistics  
Chulalongkorn University  
attapol.t@chula.ac.th

(Noiyoo and Thutkawkipin, 2023), and sentiment analysis (Khamphakdee and Seresangtakul, 2023).

Borrowing the architecture from RoBERTa (Liu et al., 2019), WangchanBERTa employs some special techniques aimed to address some characteristics of the Thai language, such as forcing its tokenizer to preserve spaces by replacing them with a special token that will never be included with other tokens, and normalizing sequences of more

# ChatGPT (Generative AI)

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

The screenshot shows a ChatGPT interface with a sidebar of recent chats and a main conversation window.

**Recent Chats (Sidebar):**

- + New chat
- Today: BJC Thailand Details.
- Yesterday: MarTech Conference Largest
- Previous 30 Days: DEQP in Thailand
- April: Mitsubishi Motor Thailand.
- March: Registering with the Department
- Registering Business with DB
- Facial depression analysis.

**Conversation Window:**

From: ขอรายละเอียดบริษัท BJC Thailand

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

BJC มีธุรกิจหลัก 4 กลุ่ม ได้แก่

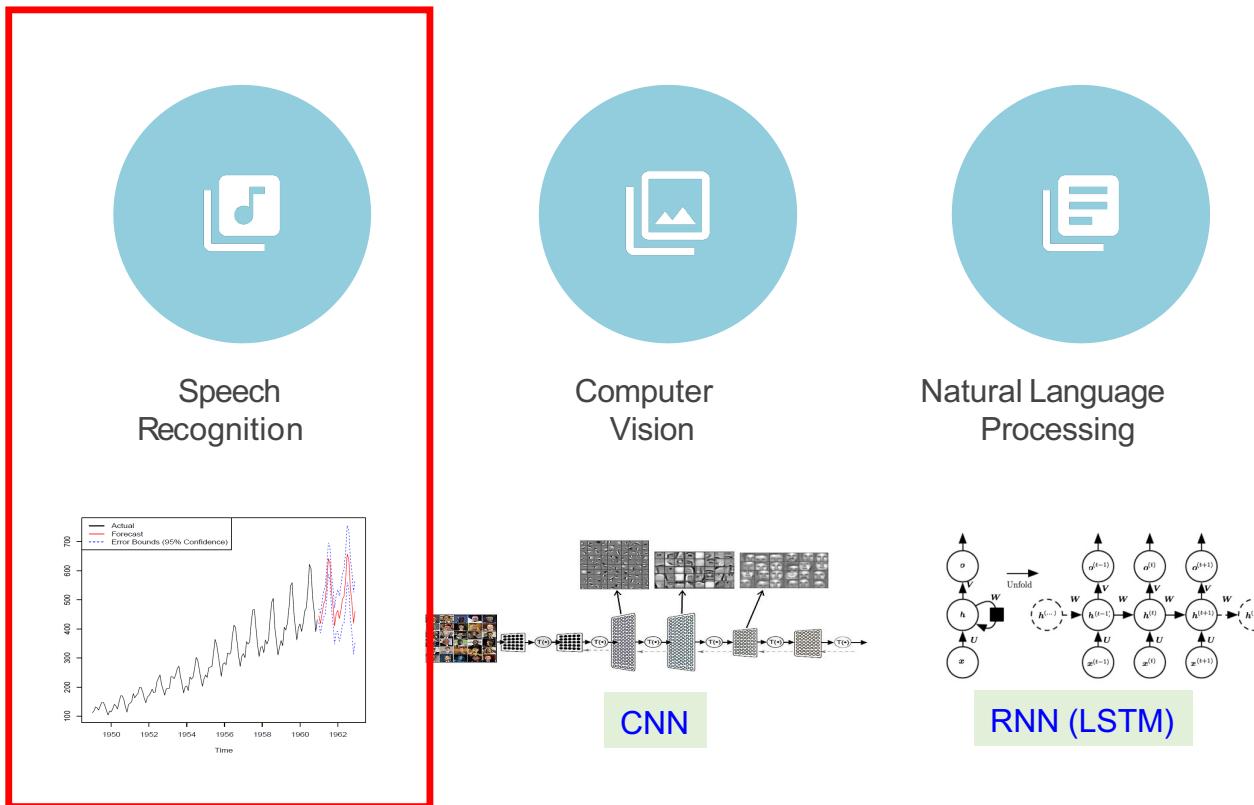
- ธุรกิจค้าปลีก: ชูปเปอร์มาร์เก็ต, ร้านสะดวกซื้อ, ร้านค้าออนไลน์ เช่น Big C, Tesco Lotus, ฯลฯ
- ธุรกิจผลิตอาหารและเครื่องดื่ม: ผลิตและจําหน่ายอาหารและเครื่องดื่ม เช่น น้ำอัดลม, น้ำดื่ม, น้ำผลไม้, อาหารสำเร็จรูป เช่น Oishi, Green Spot, Masita, Malee, Del Monte, UFC
- ธุรกิจโรงแรมและสปา: มีโรงแรมและสปาในหลายจังหวัดของไทย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

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

To the right of the message area, there are buttons for "Stop generating", "Regenerate response", and "Send a message".

ChatGPT<sup>[a]</sup> 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



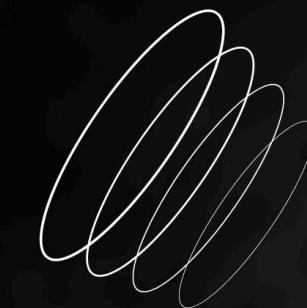
# Auto Speech Recognition (ASR)



Thonburian Whisper



GOWAJEE



GOWAJEE API  
LAUNCH

# Text-to-Speech (TTS)

## Microsoft Edge TTS

The screenshot shows the Microsoft Edge TTS interface. At the top, it says "Edge TTS Text-to-Speech". Below that is a section titled "Text-to-Speech with Microsoft Edge TTS" which includes a note about adjusting speech rate and pitch. A "Try New Version" button is visible. On the right, there's a sidebar with "Turn Your Text Into Professional" features: 40+ languages and 300+ voices supported, custom backgrounds, music, and visual effects, creating engaging video content from simple text, and being perfect for educators and content creators.

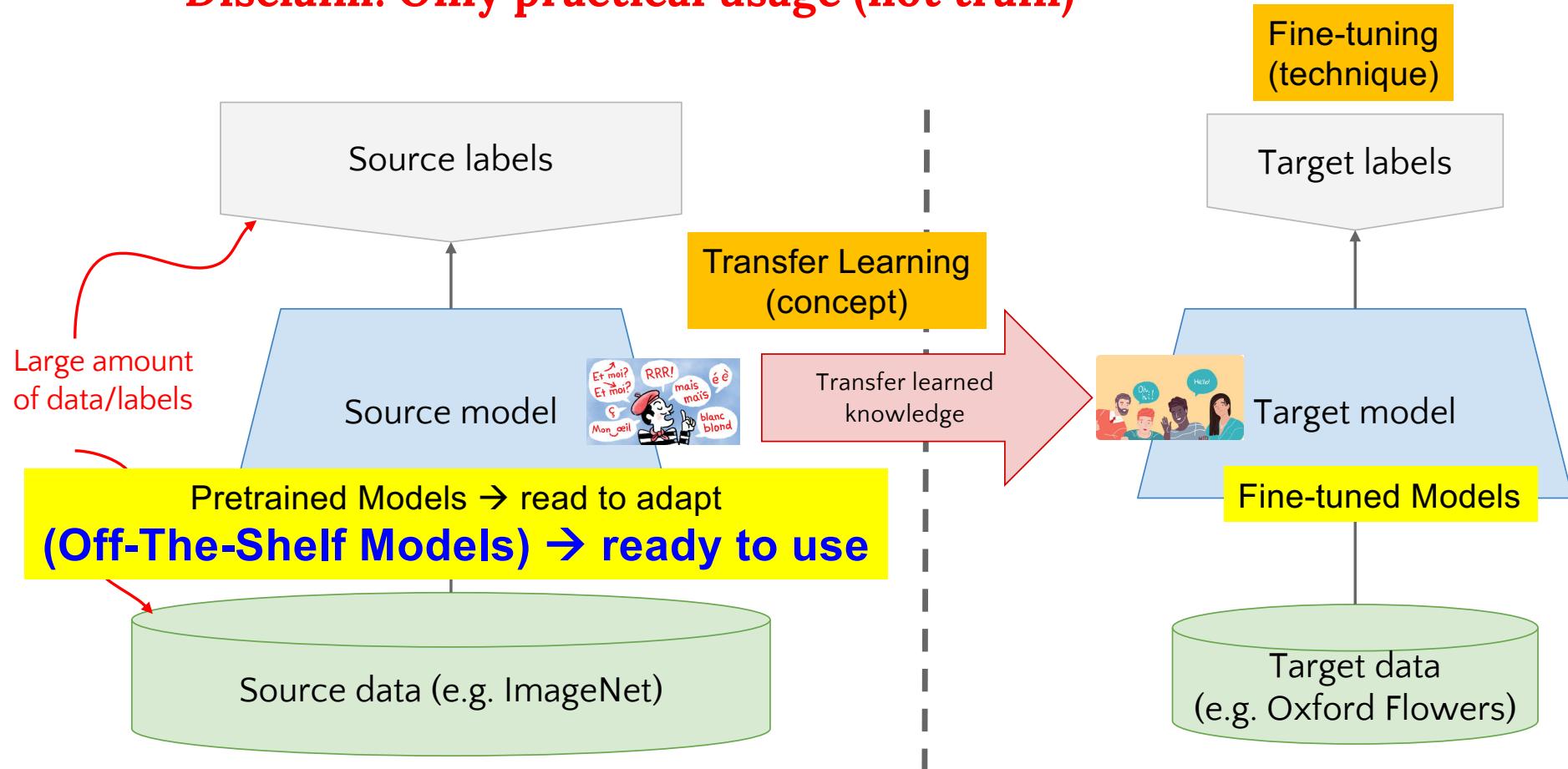
The screenshot shows the BOTNOI Voice website. It features a large "BOTNOI Voice" logo at the top. Below it is a language selection bar with various flags and names: Thai, English, Vietnamese, Polish, Portuguese, Korean, Lao, Chinese, Indonesian, German, French, Spanish, and others. A "กรุณาเลือกภาษาที่คุณต้องการ" (Please select your language) input field is present. At the bottom, there's a "ลองพิมพ์ประโยคที่คุณต้องการที่นี่" (Type the sentence you want here) text input field and a "ส่ง" (Send) button. A note at the bottom says "เครดิตให้คะแนน: 3 ครั้ง / วัน" (Credit limit: 3 times / day).

The screenshot shows the Google Cloud Speech-to-Text interface. At the top, it says "Google Cloud Speech-to-Text". Below that is a "Cloud Speech-to-Text" section with tabs for "Speech-to-Text" (which is selected), "Features", "Options", "How It Works", "Demo", and "Common Uses". To the right, there's a section titled "Turn speech into text using Google AI" with a note about converting audio into text transcriptions and integrating speech recognition into applications with easy-to-use APIs.



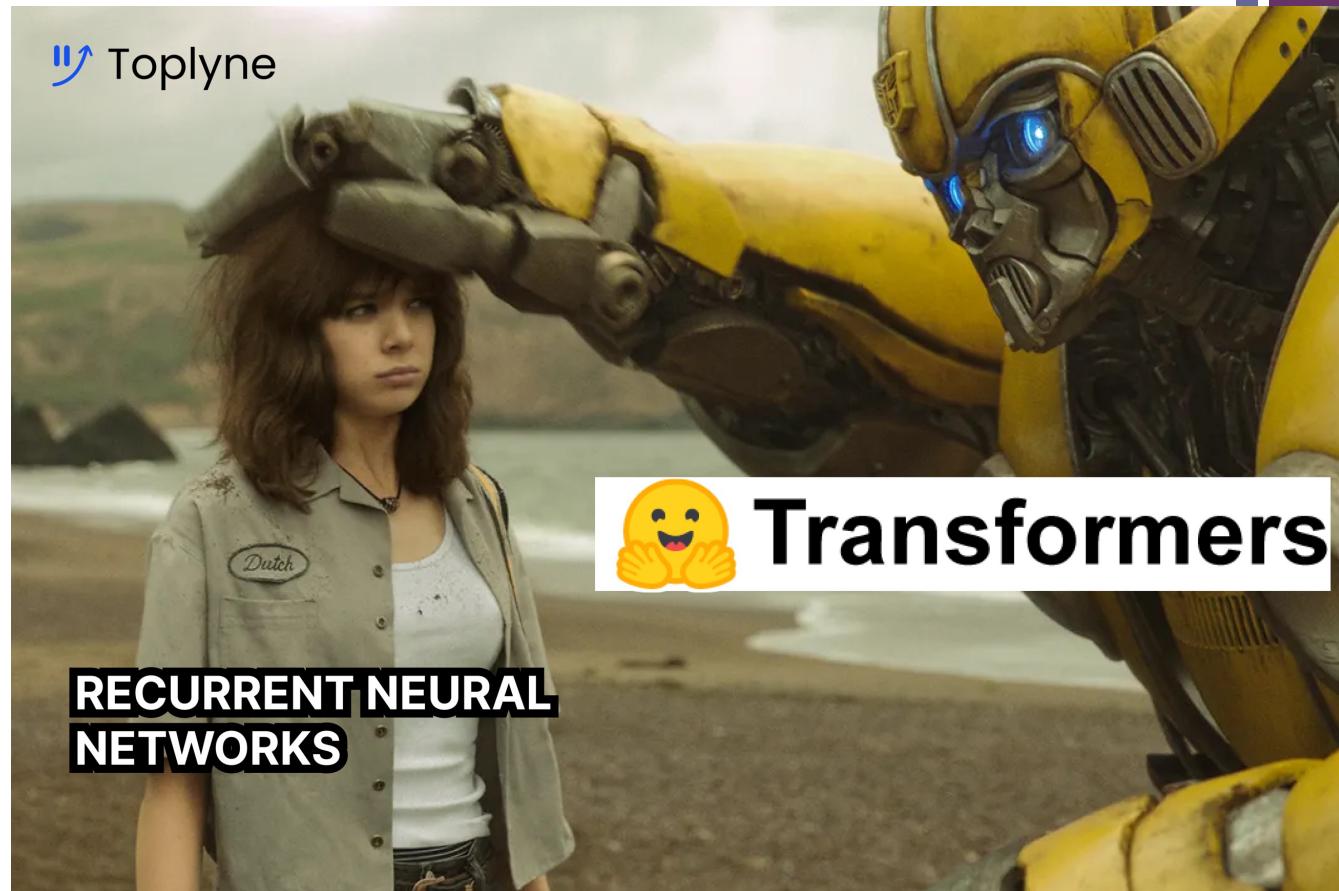
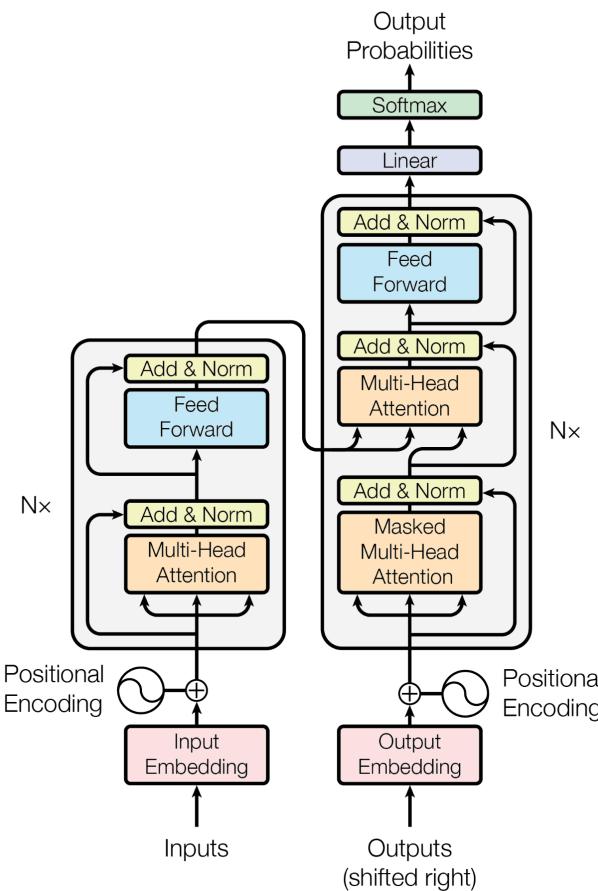
# This lab: Off-the-shelf models

Disclaimer: Only practical usage (not train)





# Rise of Transformer (2017)





# Transformer-Based Models

All Transformer based

1. Encoder-based model: BERT (2018)
  2. Decoder-based model: GPT (2018)
  3. Encoder and Decoder: BART (2019)

## BERT [Devlin, et al, 2018]: Bidirectional Encoder Representation from Transformers

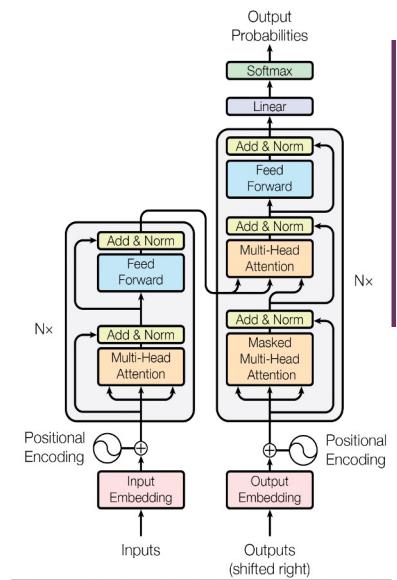
# BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

**Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova**  
Google AI Language  
`jacobdevlin,mingweichang,kentonl,kristout}@google.com`

## Abstract

We introduce a new language representation model called **BERT**, which stands for Bidirectional Encoder Representations from Transformers. Unlike recent language representation models (Peters et al., 2018a; Rad-

There are two existing strategies for applying pre-trained language representations to downstream tasks: *feature-based* and *fine-tuning*. The feature-based approach, such as ELMo (Peters et al., 2018a), uses task-specific architectures that include the pre-trained representations as addi-



OpenAI GPT (Generative Pre-Training) [Radford, 2018]

## Improving Language Understanding by Generative Pre-Training

**Alec Radford**      **Karthik Narasimhan**      **Tim Salimans**      **Ilya Sutskever**  
OpenAI                  OpenAI                  OpenAI                  OpenAI  
alec@openai.com    karthikn@openai.com    tim@openai.com    ilyas@openai.com

## Abstract

Natural language understanding comprises a wide range of diverse tasks such as textual entailment, question answering, semantic similarity assessment, and document classification. Although large unlabeled text corpora are abundant, labeled data for learning these specific tasks is scarce, making it challenging for discriminatively trained models to perform adequately. We demonstrate that large gains on these tasks can be realized by *generative pre-training* of a language model



## Auto-Speech-Recognition (ASR)

Whisper → Thonburian Whisper  
Typhoon ASR Real-Time



# Whisper (OpenAI, 2022)

- Whisper is a multilingual automatic speech recognition (ASR) model
- Audio → Encoder → Decoder → Text
- Trained on **680,000+ hours of diverse audio**
- Supports:
  - Speech recognition
  - Speech translation
  - Language identification
  - Designed to be robust to noise, accents, and domain variation
- Transformer Encoder–Decoder (Seq2Seq)
  - Encoder: processes audio (log-Mel spectrogram)
  - Decoder: generates text tokens autoregressively



## Multitask training data (680k hours)

### English transcription

- 🗣 "Ask not what your country can do for ..." (blue)
- 📝 Ask not what your country can do for ... (green)

### Any-to-English speech translation

- 🗣 "El rápido zorro marrón salta sobre ..." (blue)
- 📝 The quick brown fox jumps over ... (green)

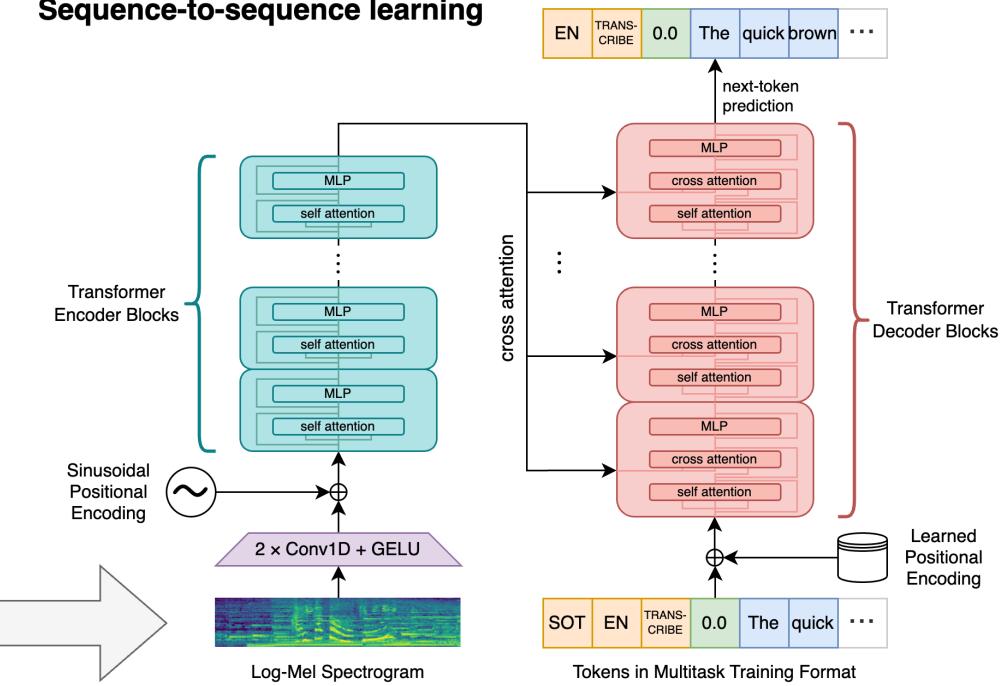
### Non-English transcription

- 🗣 "언덕 위에 올라 내려다보면 너무나 넓고 넓은 ..." (blue)
- 📝 언덕 위에 올라 내려다보면 너무나 넓고 넓은 ... (green)

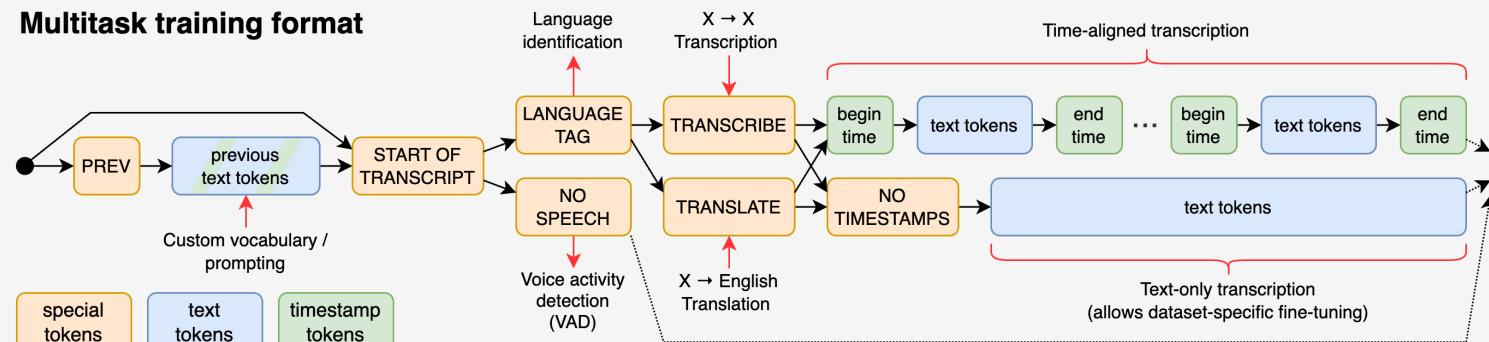
### No speech

- 🎧 (background music playing) (blue)
- ∅ (green)

## Sequence-to-sequence learning



## Multitask training format

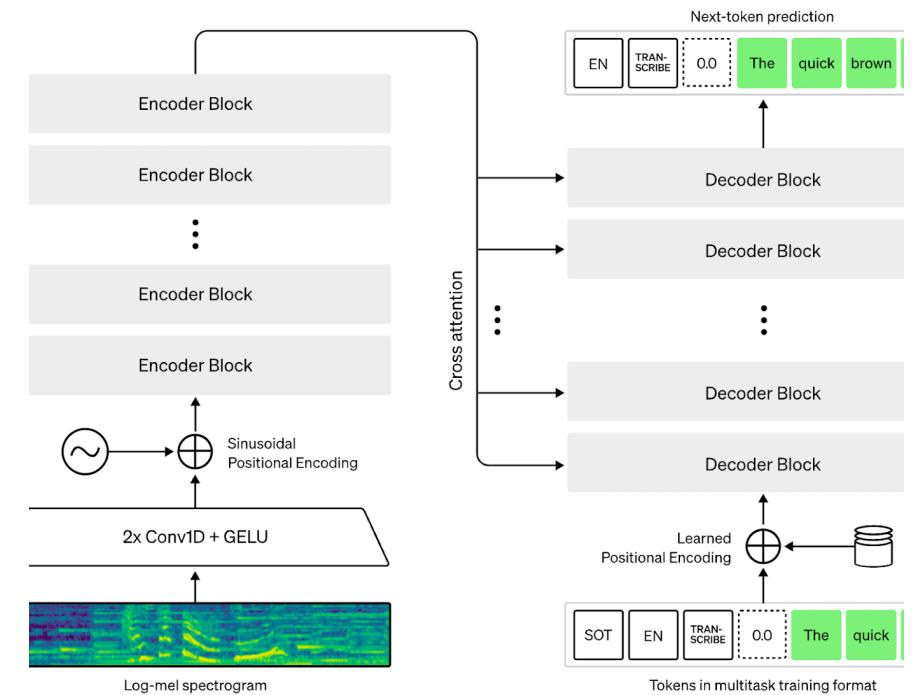




## Thonburian Whisper (looloo, 2022)

- A Thai ASR model based on OpenAI Whisper, fine-tuned with Thai speech datasets (e.g., **Commonvoice 13, Gowajee corpus, Thai Elderly Speech, Thai Dialect datasets**).
- Not officially disclosed; estimated at ~500–3,000 hours of Thai speech.
- Designed to work well in real-world conditions (background noise) and in **specialized domains like healthcare/medical and finance**.

Aspect	Whisper	Thonburian Whisper
Training	Multilingual	Thai-focused fine-tuning
Thai accuracy	Moderate	Higher
Accents	General	Thai-specific
Use case	Global ASR	Thai ASR applications



**THAIRATH Money**



**Expertpool**  
ธุรกิจยุค AI :  
ส่องการใช้  
Speech-to-Text  
ในอุตสาหกรรมชั้นนำ

朔พัฒน์ ล้ำสมบัติ  
CEO บริษัท เวิร์ดเซ็นส์ จำกัด

**PresScribe** | **looloo Health**



**AI ‘PresScribe’**  
แปลงเสียงการสนทนาระหว่างแพทย์และผู้ป่วย  
เพื่อเป็นแบบกีดการทำงานแพทย์โดยอัตโนมัติ

จาก **Looloo Health**  
“คืนเวลาให้หมอ ได้ทำหน้าที่หมอ”

นพ.วิรภัทร บุนนาค  
Head AI Healthcare Specialist  
Looloo Health

<https://loolootech.com/wordsense-speech-to-text/>



# Typhoon

- Typhoon is a family of open AI models developed for speech and language understanding and developed by SCB10x.
- Focuses on Asian and regional languages

- Designed for:

- Speech recognition (ASR)
- Language models (LLMs)
- Real-world deployment

Model	Architecture
Typhoon LLM	Decoder-only Transformer (GPT-style)
Typhoon ASR	Encoder–Decoder Transformer
GPT	Decoder-only Transformer
Whisper	Encoder–Decoder Transformer

- Goal: High-quality, open, and practical AI models

<https://opentyphoon.ai/>



LATEST



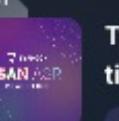
**Typhoon-Si-Med-Thinking-4B**

Text

First-of-its-kind ranked-list medical reasoning model (Research Preview)

[Learn More →](#)

LATEST



**Typhoon Isan ASR Real-time**

Audio

Real-time general Thai and Isan speech-to-text model with fast, accurate, and affordable deployment from laptops to enterprise.

[Learn More →](#)

LATEST



**Typhoon Isan ASR Whisper**

Audio

Speech recognition model that delivers state-of-the-art accuracy for transcriptions of both Isan and general Thai speech

[Learn More →](#)

LATEST



**Typhoon OCR**

Image

Next-generation bilingual vision-language model for document parsing with superior Thai document understanding.

[Learn More →](#)

LATEST



**Typhoon Translate**

Text

a Thai-English translation model that brings superior instruction accuracy and fine-grained control—letting users customize tone, terminology, and formatting while...

[Learn More →](#)

LATEST



**Typhoon 2.5**

Text

Typhoon 2.5 brings agentic intelligence, ultra efficiency, and natural Thai fluency to real-world workflows.

[Learn More →](#)

LATEST



**Typhoon ASR Real-time**

Audio

Real-time Thai speech-to-text, fast, accurate, and affordable—from laptops to enterprise.

[Learn More →](#)

LATEST



**Typhoon 2.1 Gemma**

Text

Lightweight, high-performance models with improved Thai alignment and controllable thinking mode for efficient deployment.

[Learn More →](#)

LATEST



**Typhoon 2**

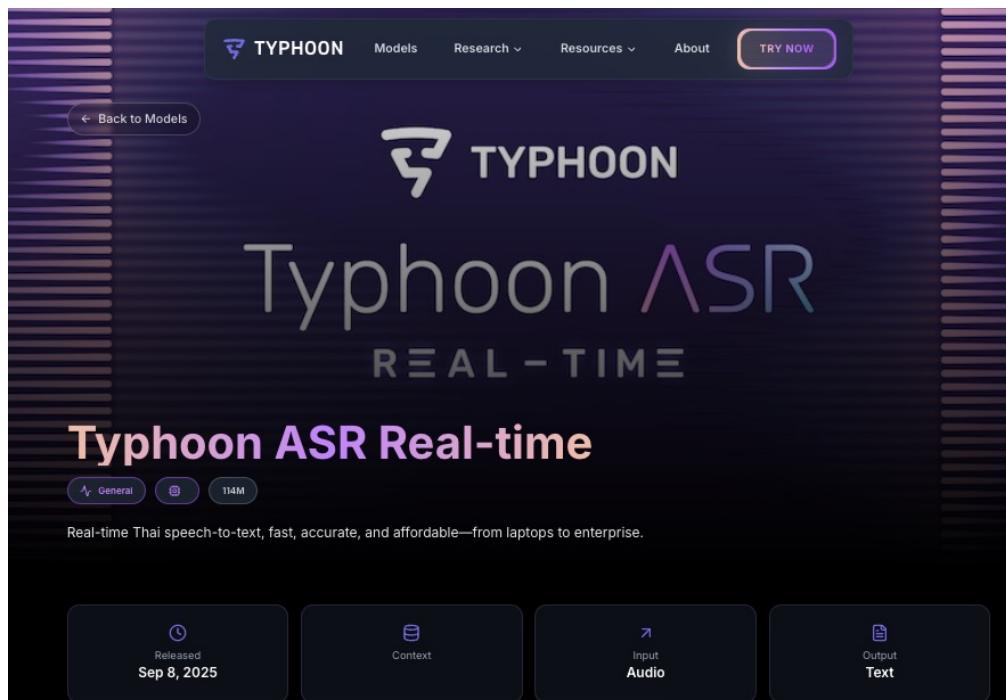
Text

Comprehensive Thai-engineered model lineup from 1B to 70B parameters with 128K context and function-calling capabilities.

[Learn More →](#)



# Typhoon ASR Real-Time & Isan



The screenshot shows the Typhoon ASR Real-time model page. At the top, there's a navigation bar with the Typhoon logo, 'Models', 'Research', 'Resources', 'About', and a 'TRY NOW' button. Below the navigation is a large 'Back to Models' button. The main title 'Typhoon ASR REAL - TIME' is displayed prominently. A sub-section titled 'Typhoon ASR Real-time' is shown with a 'General' tab selected, showing a release date of 'Sep 8, 2025'. Other tabs include 'Context', 'Input Audio', and 'Output Text'. A brief description below the tabs states: 'Real-time Thai speech-to-text, fast, accurate, and affordable—from laptops to enterprise.'



The screenshot shows the Typhoon Isan ASR Real-time model page. It features a similar layout to the first page, with the Typhoon logo and 'TRY NOW' button at the top. The main title 'Typhoon Isan ASR REAL - TIME' is displayed. A sub-section titled 'Typhoon Isan ASR Real-time' is shown with a 'General' tab selected, showing a release date of 'Nov 27, 2025'. Other tabs include 'Context', 'Input Audio', and 'Output Text'. A brief description below the tabs states: 'Real-time general Thai and Isan speech-to-text model with fast, accurate, and affordable deployment from laptops to enterprise.'

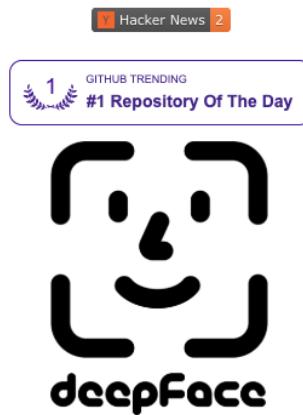


## Facial

DeepFace  
MediaPipe  
OpenFace



# 1) DeepFace (Meta)



Hacker News 2

1 GITHUB TRENDING  
#1 Repository Of The Day

DeepFace is a lightweight [face recognition](#) and facial attribute analysis ([age](#), [gender](#), [emotion](#) and [race](#)) framework for python. It is a hybrid face recognition framework wrapping state-of-the-art models: [VGG-Face](#), [FaceNet](#), [OpenFace](#), [DeepFace](#), [DeepID](#), [ArcFace](#), [Dlib](#), [SFace](#), [GhostFaceNet](#), [Buffalo\\_L](#).

A modern face recognition pipeline consists of 5 common stages: [detect](#), [align](#), [normalize](#), [represent](#) and [verify](#). While DeepFace handles all these common stages in the background, you don't need to acquire in-depth knowledge about all the processes behind it. You can just call its verification, find or analysis function with a single line of code.

- It is built on CNN-based architectures, offering a variety of model options.

- Features:

- Face embedding (vector)
- Face verification (are the same person?)
- Face analysis



<https://github.com/serengil/deepface>

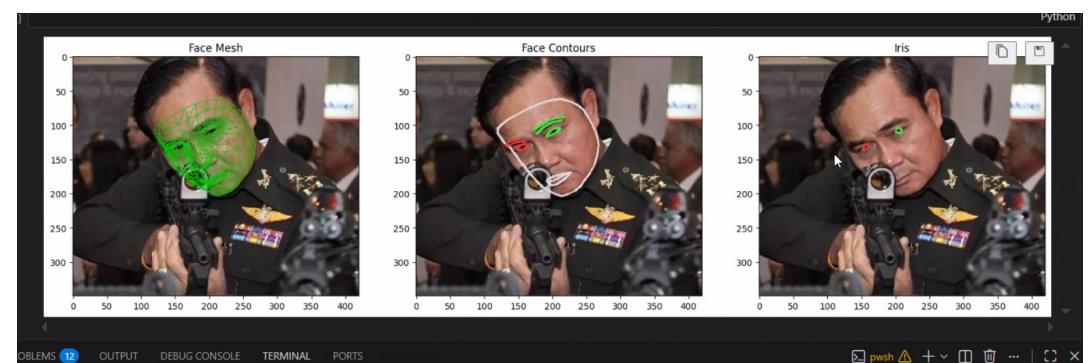




## 2) MediaPipe (Google)

- An open-source cross-platform framework by Google
- Designed for **real-time perception pipelines**
- Model: a combination of classical CV & Lightweight deep learning models (CNN-based)
- Output:
  - Face landmarks (468 points)
  - Face bounding box
  - Face mesh (3D-like structure)
  - Head pose estimation

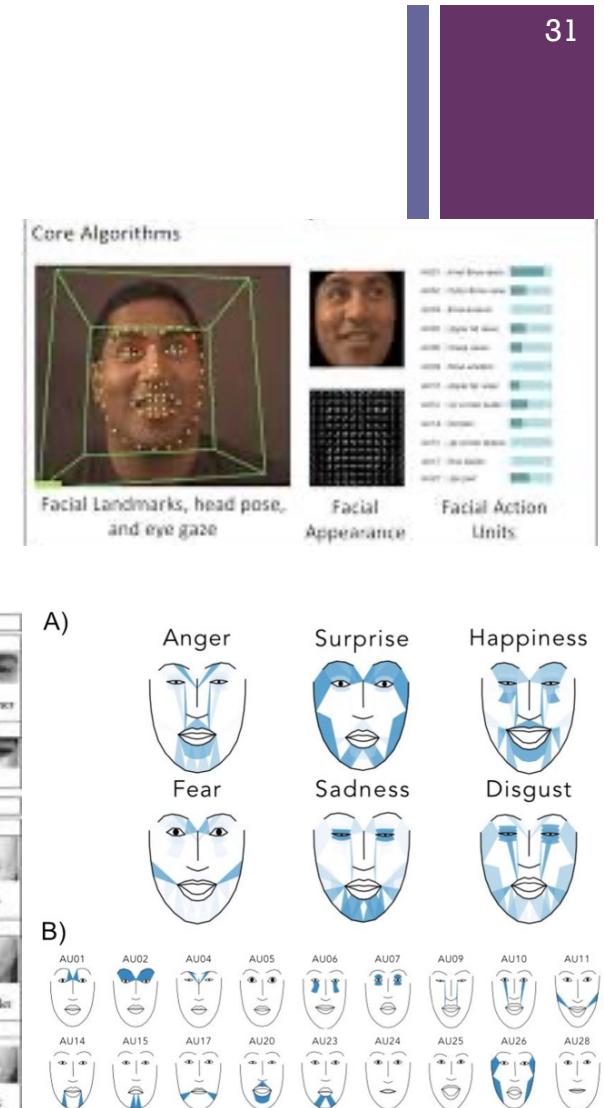
```
100%|██████████| 5.98M/5.98M [00:00<00:00, 124MB/s]
Action: emotion: 100%|██████████| 4/4 [00:32<00:00,  8.24s/it]
[{"age": 45,
 "region": {"x": 115,
 "y": 113,
 "w": 396,
 "h": 396,
 "left_eye": (378, 273),
 "right_eye": (241, 262)},
 "face_confidence": 0.9,
 "gender": {"Woman": np.float32(0.47134966), "Man": np.float32(99.528656)},
 "dominant_gender": 'Man',
 "race": {"asian": np.float32(0.0056317714),
 "indian": np.float32(0.00031789645),
 "black": np.float32(3.2617245e-06),
 "white": np.float32(99.27343),
 "middle eastern": np.float32(0.3446174),
 "latino hispanic": np.float32(0.3760018)},
 "dominant_race": 'white',
 "emotion": {"angry": np.float32(11.738707),
 "disgust": np.float32(0.002389708),
 "fear": np.float32(5.2961206),
 "happy": np.float32(0.823493),
 "sad": np.float32(22.521772),
 "surprise": np.float32(0.091061),
 "neutral": np.float32(59.526455)},
 "dominant_emotion": 'neutral'}]
```





### 3) OpenFace (CMU)

- OpenFace was developed by Carnegie Mellon University (CMU).
- An open-source facial behavior analysis toolkit.
- Widely used in affective computing and **psychology**.
- Model: a combination of CNN + classical ML (SVM/regression)
- Output:
  - 68 facial landmarks
  - 17 action Units (AUs) (FACS-based)
    - AU presence & intensity
  - Head pose
  - Eye gaze
  - **NOT directly support emotion recognition**

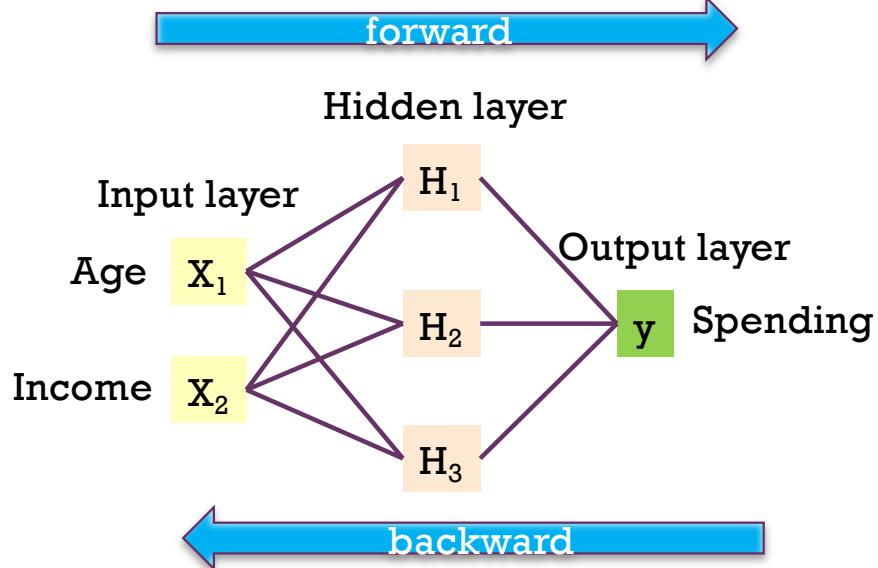


+

SHAP Value

$$\text{Spending} = 500 + 10 \cdot \text{Income} / 10K + 2 \cdot \text{Age}$$

# Neural Networks (universal approximator)

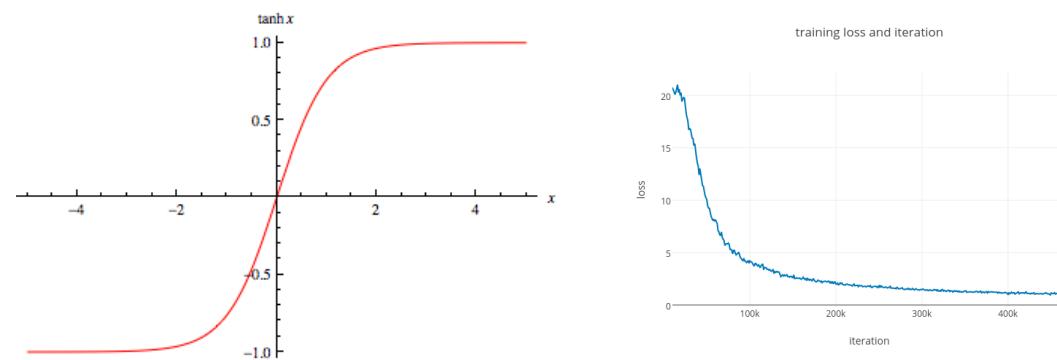


$$\text{Spending} = \hat{w}_0 + \hat{w}_1 H_1 + \hat{w}_2 H_2 + \hat{w}_3 H_3$$

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

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

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



How to update weight

<https://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example/>

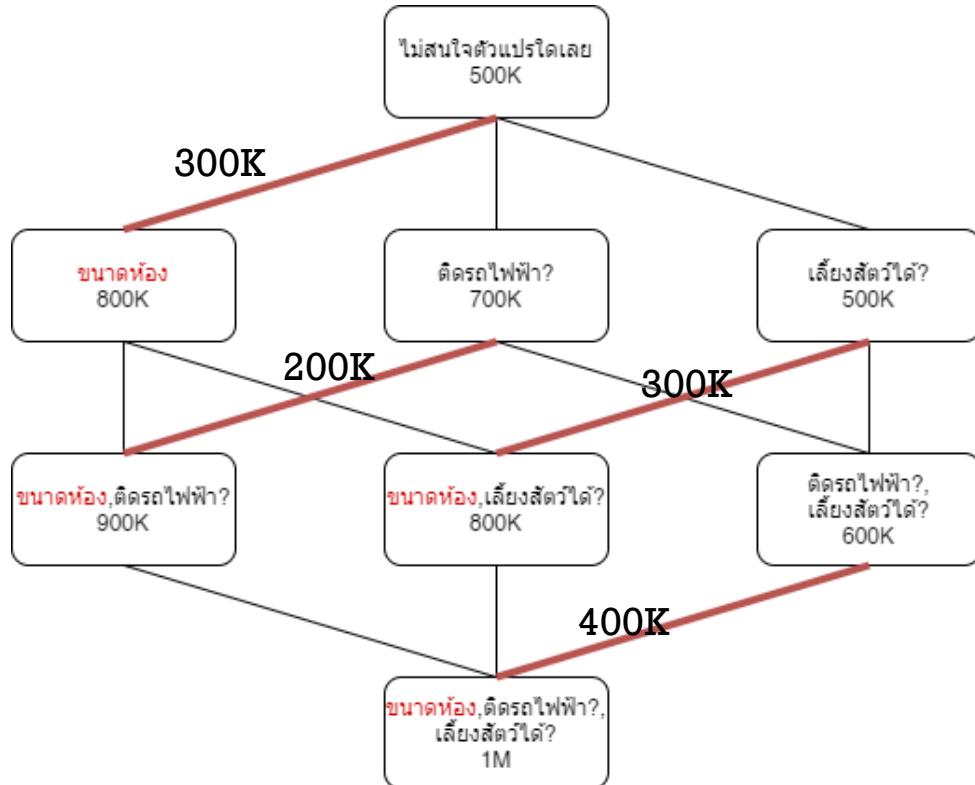


# What is SHAP Value?

- SHAP (SHapley Additive exPlanations) is a method for explaining individual predictions of machine learning models by assigning each feature a contribution value. It has been inspired from [the game theory](#).
  - SHAP explains why a model makes a prediction
  - [Widely used in Explainable AI \(XAI\)](#), especially in healthcare and finance
- For ML, it was proposed by Scott Lundberg, et. al., in 2017 in the paper “A Unified Approach to Interpreting Model Predictions”



## ตัวอย่างการคำนวณ SHAP Value



[Link](#)

## อธิบายโมเดลให้เข้าใจมากขึ้นด้วย SHAP



In touch Kunakorntum

Follow

2 min read · Sep 20, 2020

35

### กำหนดใหม่ 3 ตัวแปร

- ขนาดของห้องในคอนโด
- คอนโดติดรถไฟฟ้าหรือไม่
- สามารถเลี้ยงสัตว์ได้หรือไม่



### SHAP(ขนาดห้อง) = average(300K, 200K, 300K, 400K) = 300K บาท

- การคำนวณ SHAP Value จะใช้เวลาที่มาก ดังนั้นจึงมีการนำหลักการ Sampling และ Approximation มาร่วมกับการคำนวณค่าอีกด้วย + Surrogate Model (NN → Tree)



# SHAP Additive Formula

## Baseline + all SHAP values = prediction

For a single instance  $x$ :

$$f(x) = \mathbb{E}[f(X)] + \sum_{j=1}^M \phi_j$$

Where:

- $f(x)$  = model prediction
- $\mathbb{E}[f(X)]$  = **baseline** (expected model output)
- $\phi_j$  = SHAP value of feature  $j$

### Example Model

Prediction task: Disease risk

Features:

- Age
- Glucose

Baseline:

$$\mathbb{E}[f(X)] = 0.20$$

### Marginal Contributions

Feature	Contribution
Age	+0.15
Glucose	+0.35

So:

$$0.20 + 0.15 + 0.35 = 0.70$$

Prediction for a patient:

$$f(x) = 0.70$$

# Expected Output [Each example]

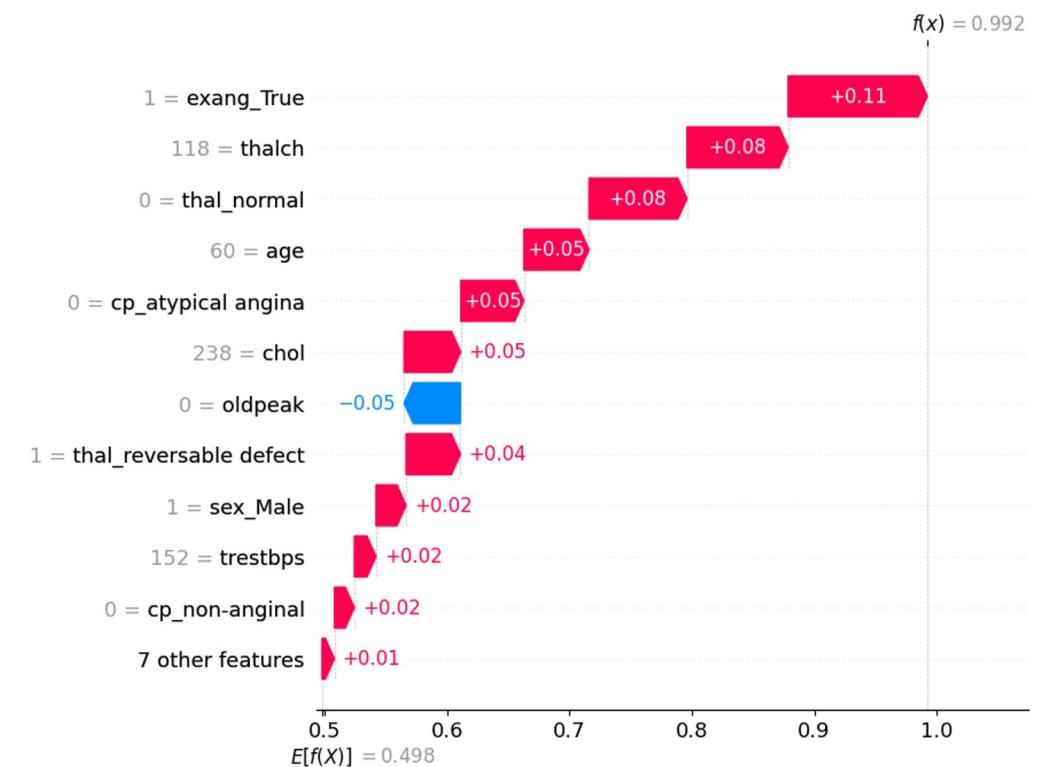
**Local SHAP waterfall plot:** starting from the **baseline prediction**, each feature pushes the patient's **risk up (red)** or down (**blue**) until reaching the final predicted risk.

$$\text{Importance}_{i,j} = |\phi_{i,j}|$$

Sample	SHAP(Age)
1	+0.20
2	-0.10
3	+0.30

**Global feature importance:**

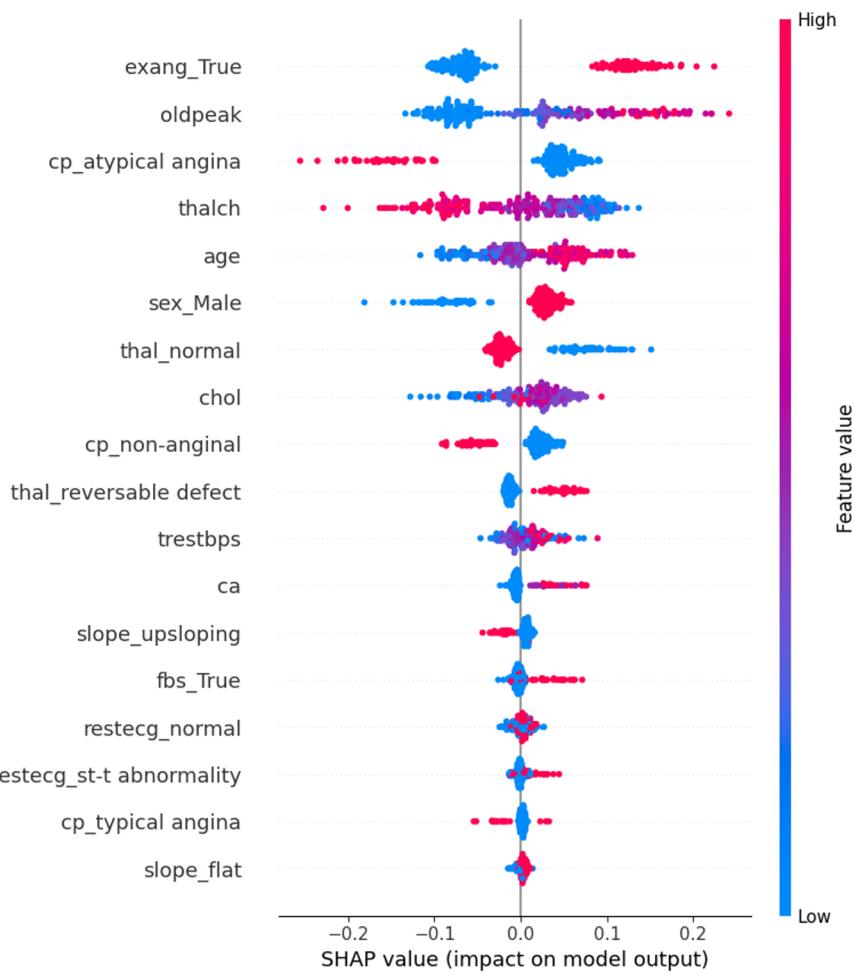
$$\frac{|0.20| + |-0.10| + |0.30|}{3} = \frac{0.60}{3} = 0.20$$



# Expected Output [Overall model]

**SHAP summary plot:** features are ranked by **importance**, and each dot shows whether a patient's feature value pushes the model toward **higher risk (right)** or **lower risk (left)**, with color indicating low vs high feature values.

**Ex. exang\_True:** When this feature is **high (red = 1, exercise-induced angina)**, dots are mostly on the **right**, meaning it **pushes the model toward higher heart-disease risk**.



+

**Grad-CAM**



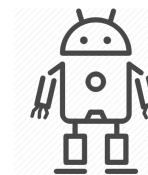
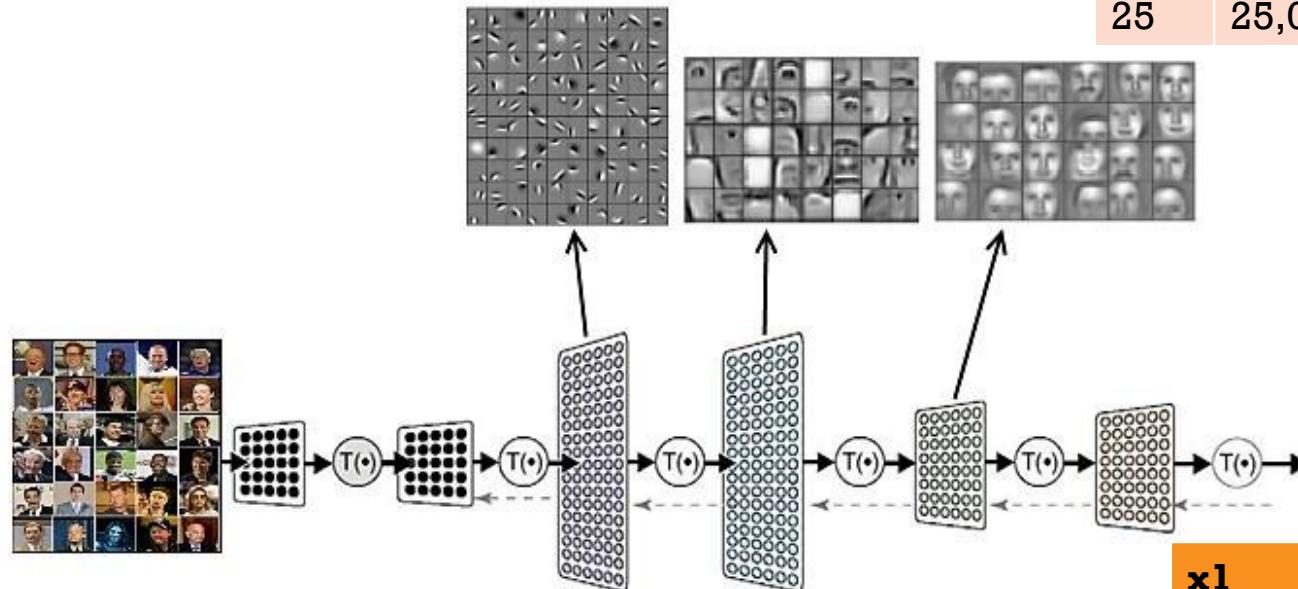
# Deep Learning – Basics (cont.)

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  $\rightarrow$  nose  $\rightarrow$  face). The output layer combines those features to make predictions.



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

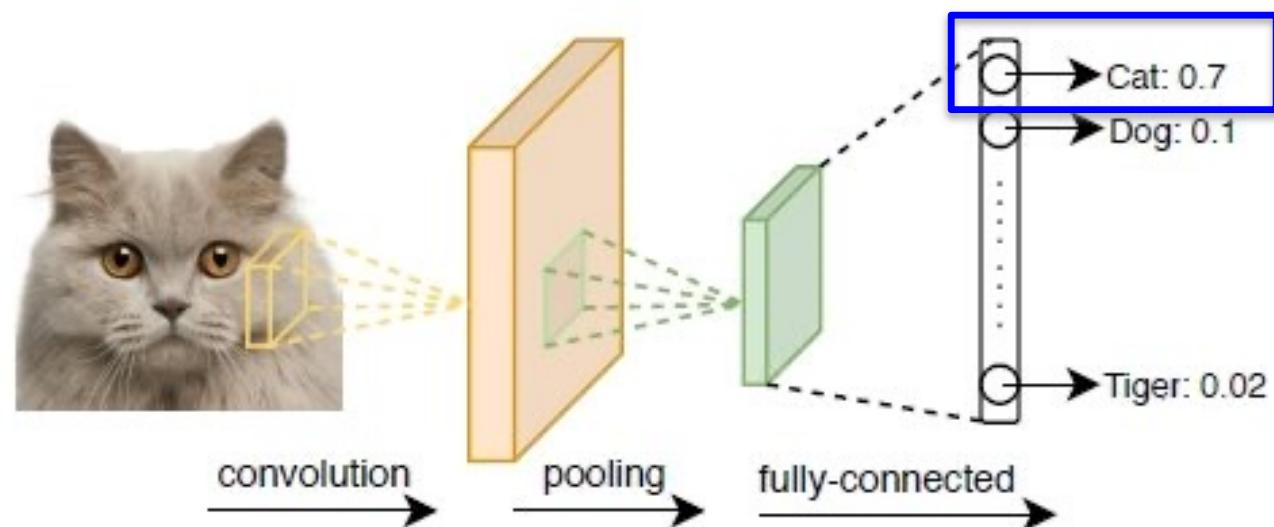


x1	x2	x3	x4	Corona
0.7	0.2	-0.5	-0.1	Yes



# Image Classification: Why does model predict “Cat”?

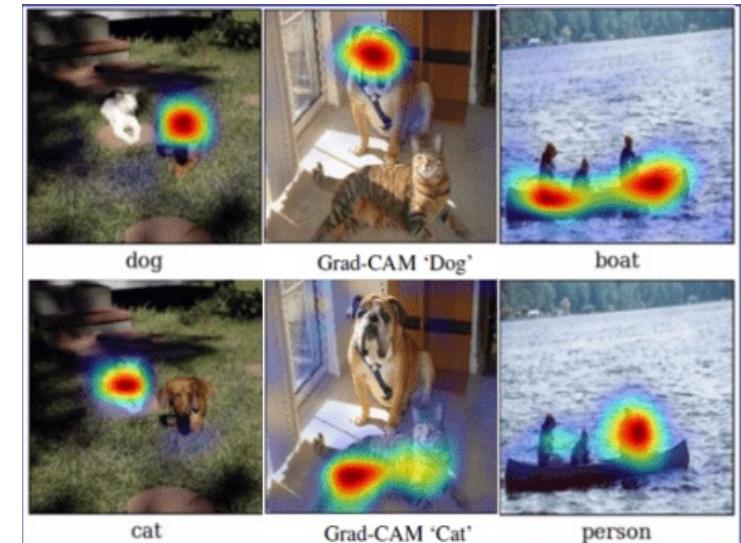
## Convolutional Neural Network



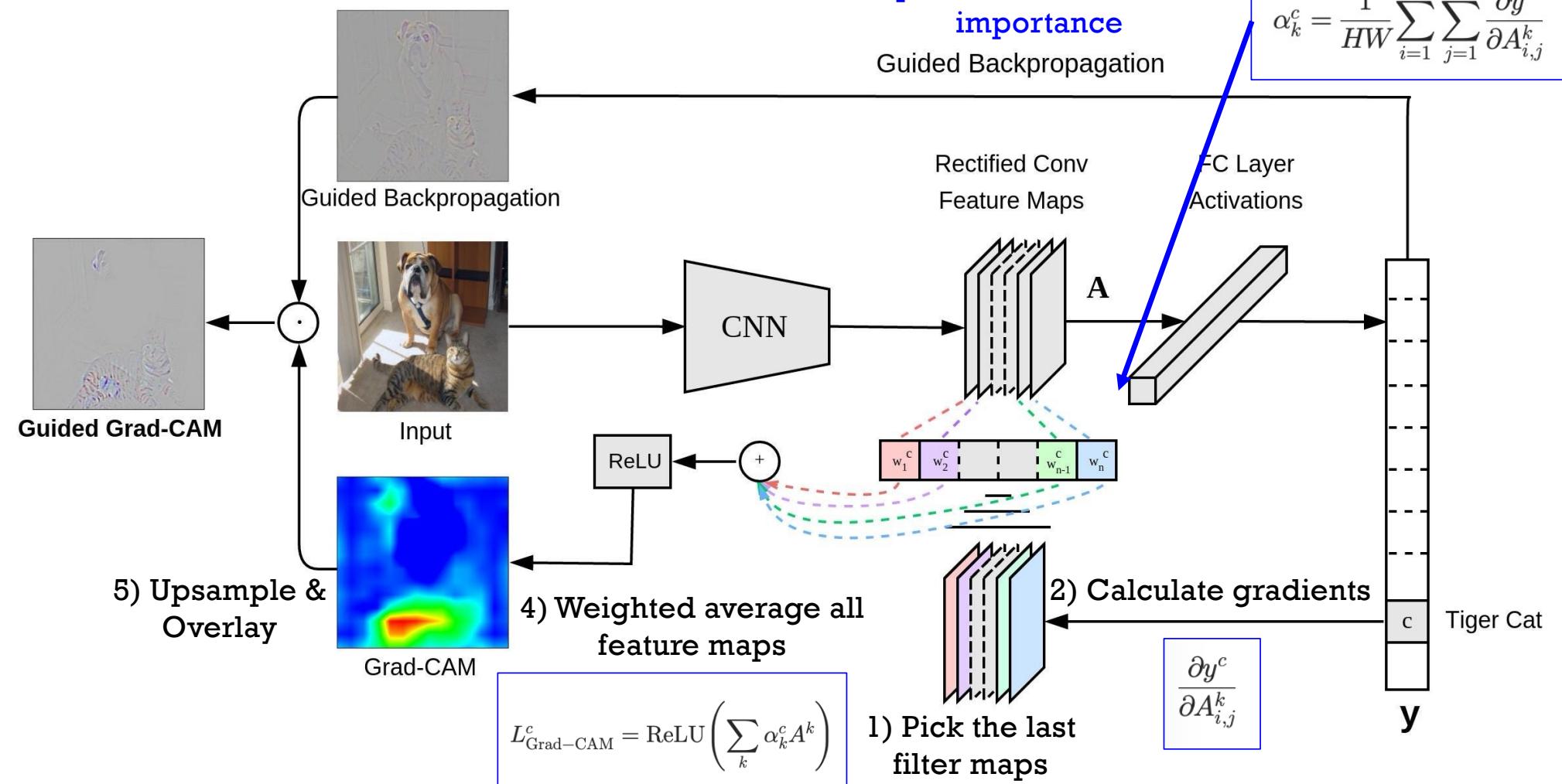


# Gradient-weighted Class Activation Mapping (Grad-CAM; ICCV-2017)

- A technique to explain **CNN** predictions on images
- Shows which **regions of an image** influence a model's decision
- Produces **class-specific heatmaps**
- Widely used in **medical imaging**, vision QA, and model debugging
- How's about Transformer?
  - ViT (medical images) → Transformer Grad-CAM / Layer-wise Relevance Propagation (LRP)
  - Text Transformers (BERT) → Token attribution / SHAP



# How to calculate Grad-CAM?



+ Thank you  
& any questions