Generative AI:
Current Trends and Practical Applications

Kao Panboonyuen
panboonyuen.kao@gmail.com

https://kaopanboonyuen.github.io

Reference:

1. https://pytorch.org/tutorials
2. https://stanford.edu/~shervine/teaching/cs-229/
3. http://introtodeeplearning.com/
4. https://www.simplilearn.com/tutorials/deep-learning-tutorial/introduction-to-deep-learning
5. https://www.geeksforgeeks.org/introduction-deep-learning/
6. https://www.telecats.com/blog-en/ai-for-rookies/
7. https://marketoonist.com/2023/01/ai-tidal-wave.html
8. https://www.v7labs.com/blog/self-supervised-learning-guide
9. https://blog.miarec.com/contact-centers-ai-definition
10. https://www.solulab.com/top-generative-ai-trends/

Outlines

- About Me
- Introduction to Generative AI
- Key Trends in Generative AI
- Applications of Generative AI

About Me

Kao Panboonyuen

kao-panboonyuen

AI Research Scientist

Name: Teerapong Panboonyuen (P'Kao)

Contact: teerapong.pa@chula.ac.th

panboonyuen.kao@gmail.com

Education: Ph.D. in Computer Engineering, Chula

Position: AI Team Lead, MARS

PostDoc, Chula

Visiting Lecturer, KKU

Interests: Computer Vision, Deep Learning

Machine Learning, Remote Sensing

Google Scholar SIGN IN

Teerapong Panboonyuen

FOLLOW

Senior Research Scientist at MARS, Post-doc at Chula
Verified email at chula.ac.th - Homepage

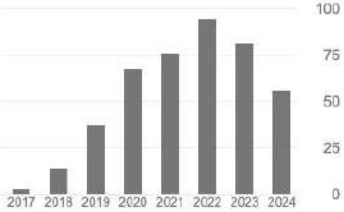
Artificial Intelligence   Machine Learning   Deep Learning   Computer Vision   Remote Sensing

TITLE	CITED BY	YEAR
Road segmentation of remotely-sensed images using deep convolutional neural networks with landscape metrics and conditional random fields T Panboonyuen, K Jitkajornwanich, S Lawawirojwong, P Srestasathiern, ... Remote Sensing 9 (7), 680	137	2017
Semantic segmentation on remotely sensed images using an enhanced global convolutional network with channel attention and domain specific transfer learning T Panboonyuen, K Jitkajornwanich, S Lawawirojwong, P Srestasathiern, ... Remote Sensing 11 (1), 83	107	2019
An enhanced deep convolutional encoder-decoder network for road segmentation on aerial imagery T Panboonyuen, P Vateekul, K Jitkajornwanich, S Lawawirojwong Recent Advances in Information and Communication Technology 2017 ...	47	2018
Transformer-based decoder designs for semantic segmentation on remotely sensed images T Panboonyuen, K Jitkajornwanich, S Lawawirojwong, P Srestasathiern, ... Remote Sensing 13 (24), 5100	45	2021
Object detection of road assets using transformer-based YOLOX with feature pyramid decoder on thai highway panorama T Panboonyuen, S Thongbai, W Wongweeranimit, P Santitamnont, ... Information 13 (1), 5	20	2021
Real-time polyps segmentation for colonoscopy video frames using compressed fully convolutional network I Wichakam, T Panboonyuen, C Udomcharoenchaikit, P Vateekul MultiMedia Modeling: 24th International Conference, MMM 2018, Bangkok ...	18	2018
Semantic segmentation on medium-resolution satellite images using deep convolutional networks with remote sensing derived indices S Chantharaj, K Pornratthanapong, P Chitsinpchayakun, T Panboonyuen, ... 2018 15th International joint conference on computer science and software ...	13	2018

GET MY OWN PROFILE

Cited by

	All	Since 2019
Citations	429	411
h-index	9	9
i10-index	8	8



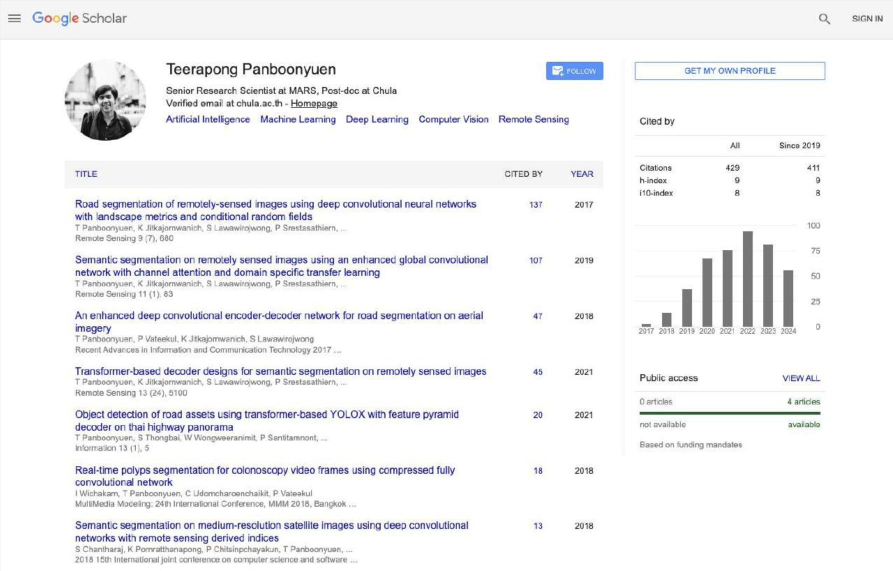
Public access   VIEW ALL

0 articles   4 articles

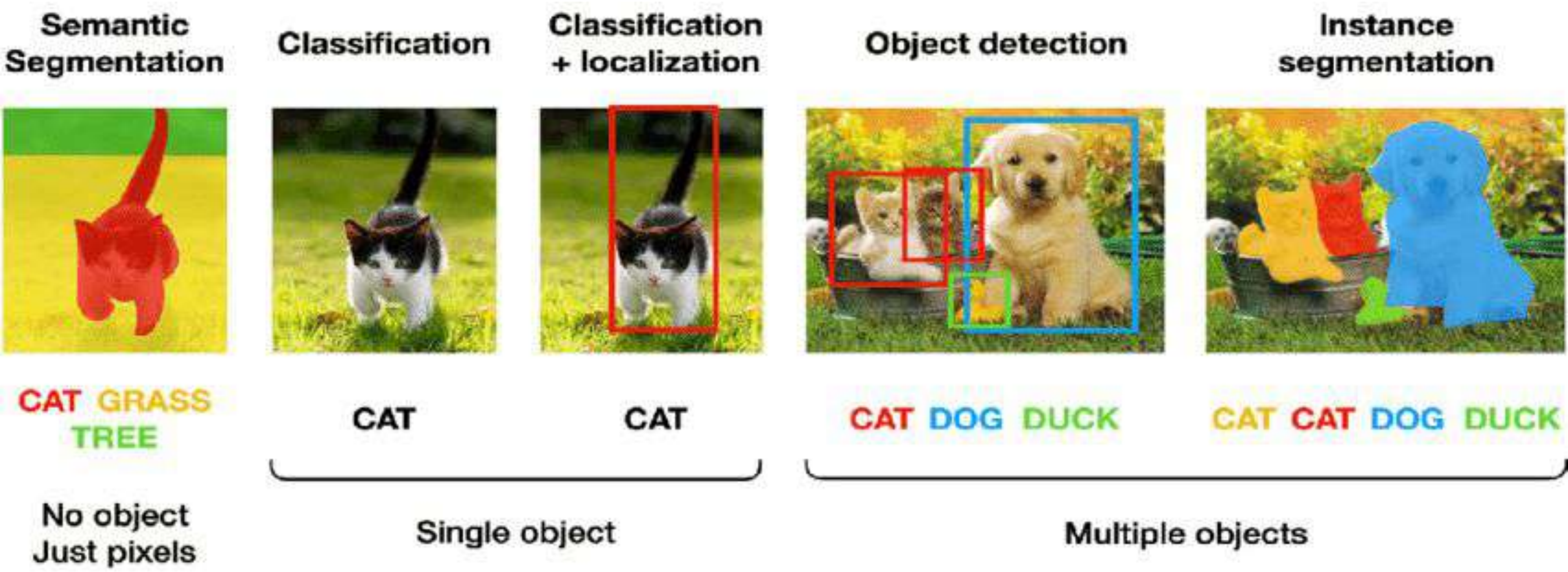
not available   available

Based on funding mandates

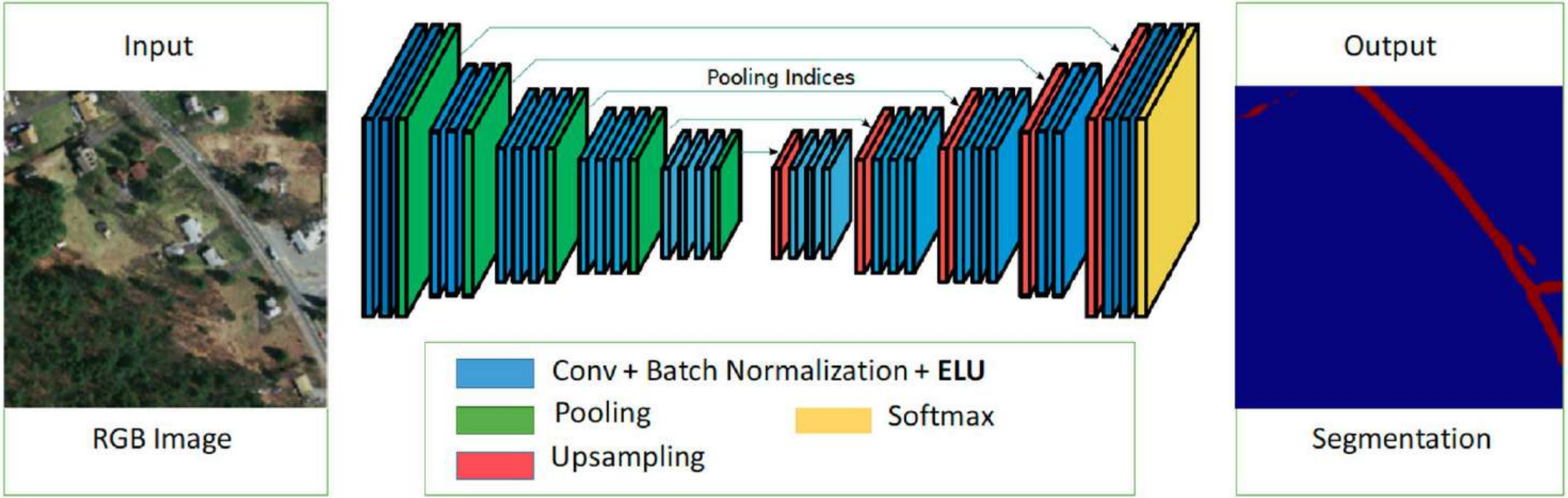
Featured Publications



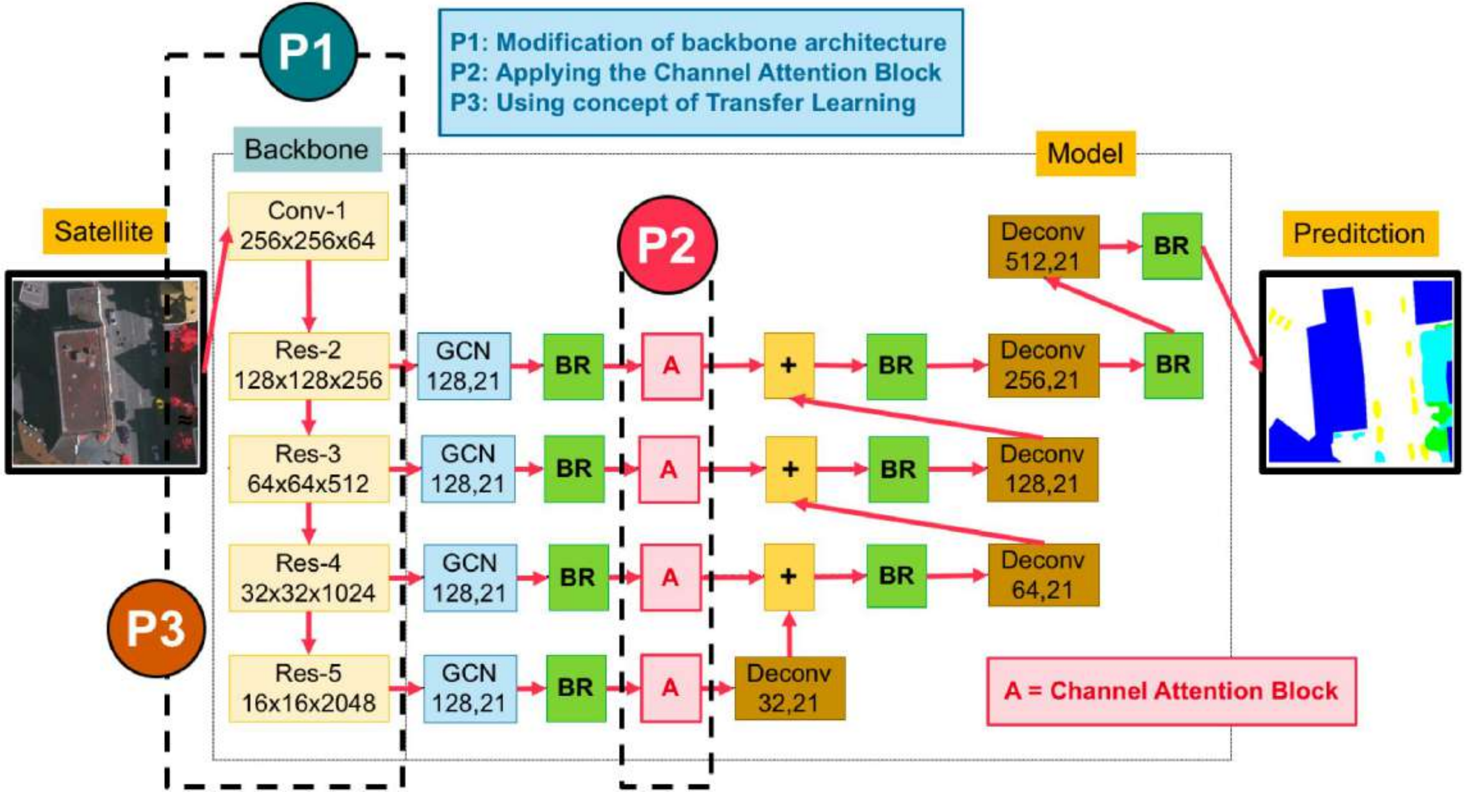
What Is Computer Vision? [Basic Tasks & Techniques]



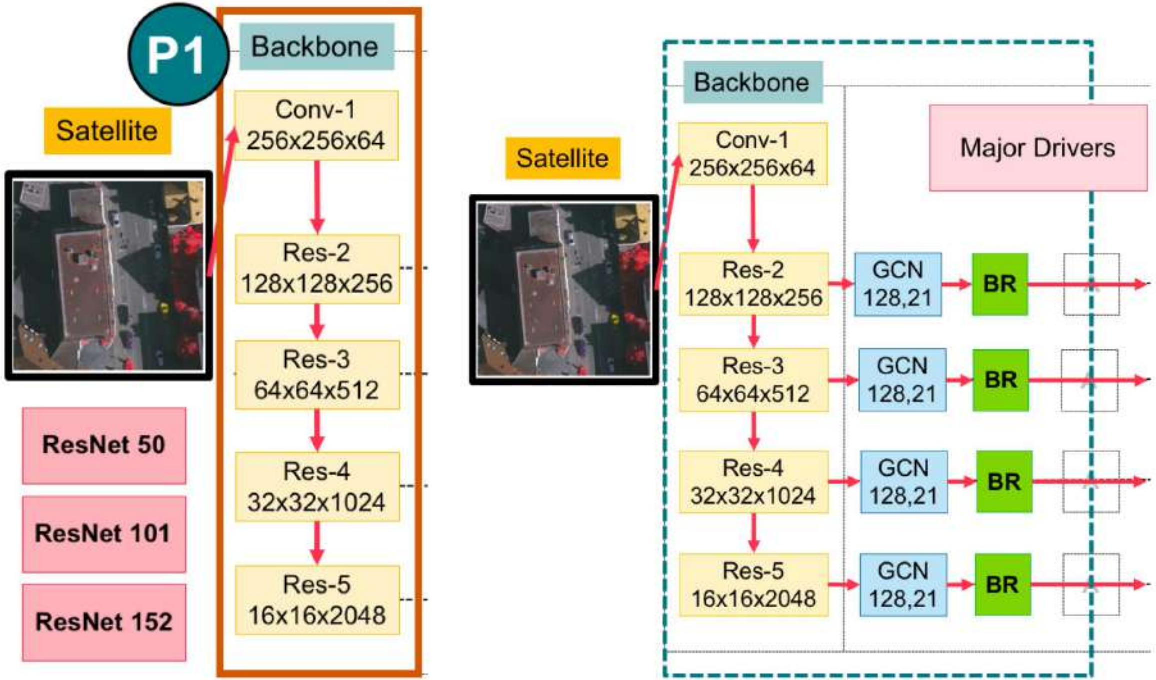
[1] Panboonyuen, Teerapong, et al. **"Road segmentation of remotely-sensed images using deep convolutional neural networks with landscape metrics and conditional random fields."** Remote Sensing 9.7 (2017): 680.



[2] Panboonyuen, Teerapong, et al. **"Semantic segmentation on remotely sensed images using an enhanced global convolutional network with channel attention and domain specific transfer learning."** Remote Sensing 11.1 (2019): 83.

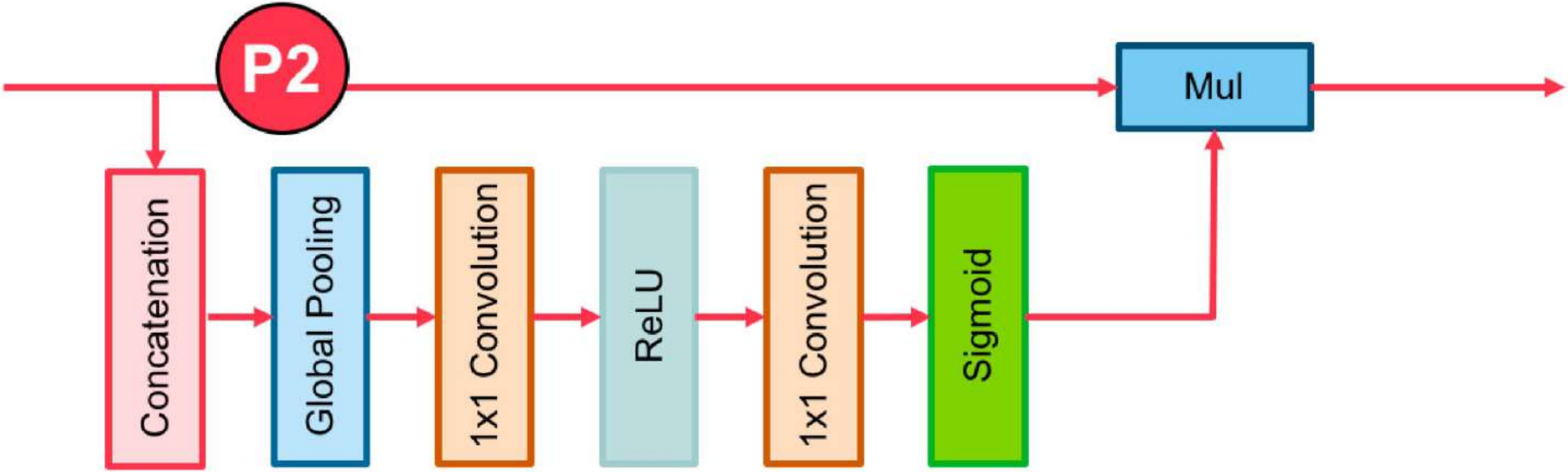


[2] Panboonyuen, Teerapong, et al. **"Semantic segmentation on remotely sensed images using an enhanced global convolutional network with channel attention and domain specific transfer learning."** Remote Sensing 11.1 (2019): 83.



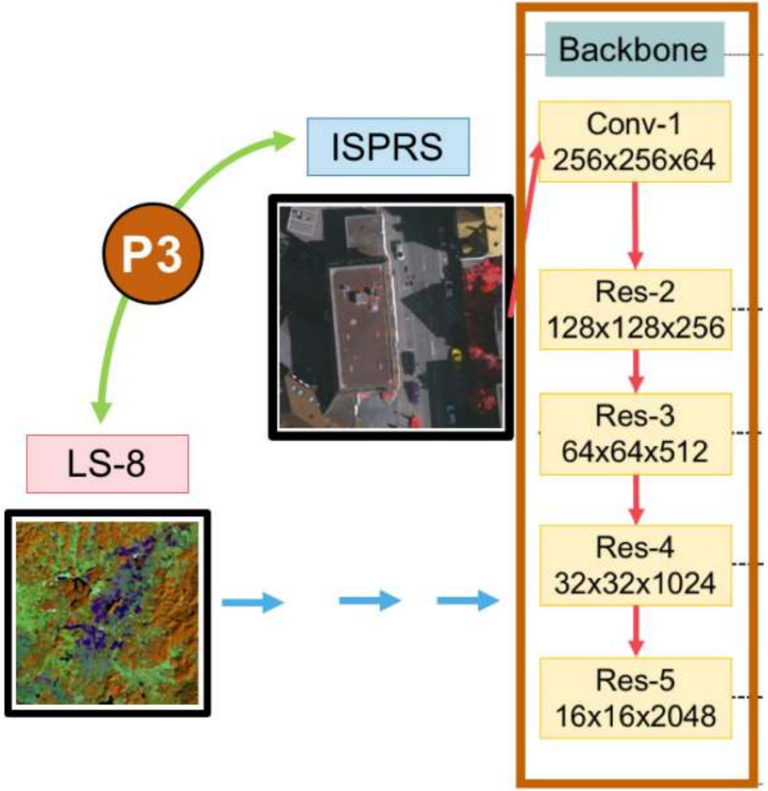
An overview of the whole backbone pipeline in (left) the main backbone with varying by ResNet50, ResNet 101, and ResNet 152; (right) the major drivers of my main classification network (composed of a global convolutional network (GCN) and a boundary refinement (BR) block.

[2] Panboonyuen, Teerapong, et al. **"Semantic segmentation on remotely sensed images using an enhanced global convolutional network with channel attention and domain specific transfer learning."** Remote Sensing 11.1 (2019): 83.



- Components of the channel attention block.
- The red lines represent the downsample operators, respectively.
- The red line cannot change the size of feature maps.
- It is only a path for information passing.

[2] Panboonyuen, Teerapong, et al. **"Semantic segmentation on remotely sensed images using an enhanced global convolutional network with channel attention and domain specific transfer learning."** Remote Sensing 11.1 (2019): 83.



- The domain-specific transfer learning strategy reuses pre-trained weights of models between two datasets—very high (ISPRS) and medium (Landsat-8; LS-8) resolution images.

[2] Panboonyuen, Teerapong, et al. **"Semantic segmentation on remotely sensed images using an enhanced global convolutional network with channel attention and domain specific transfer learning."** Remote Sensing 11.1 (2019): 83.

Figure 6. Sample satellite images from Nan, a province in Thailand (**left**), and corresponding ground truth (**right**). The label of medium resolution dataset includes five categories: agriculture (yellow), forest (green), miscellaneous (brown), urban (red), and water (blue).

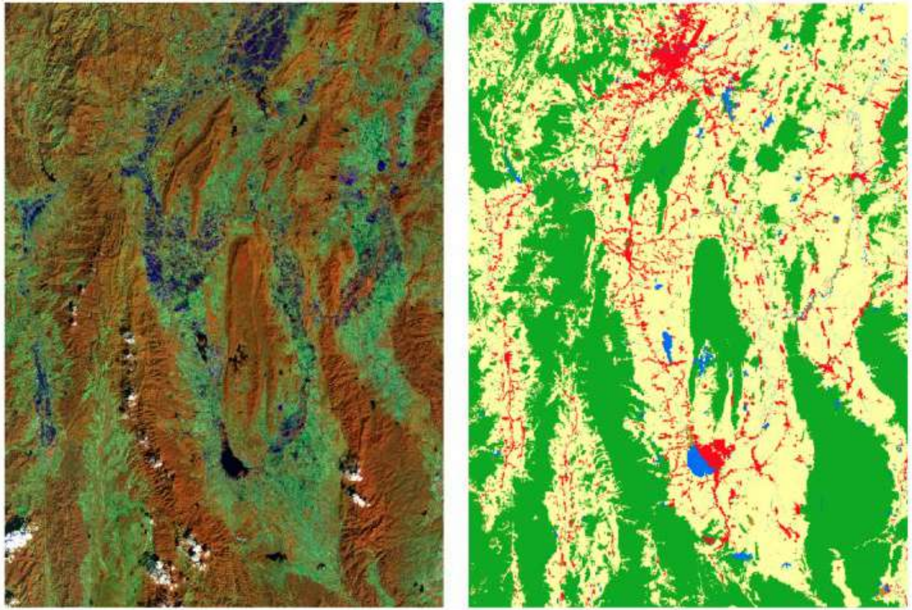
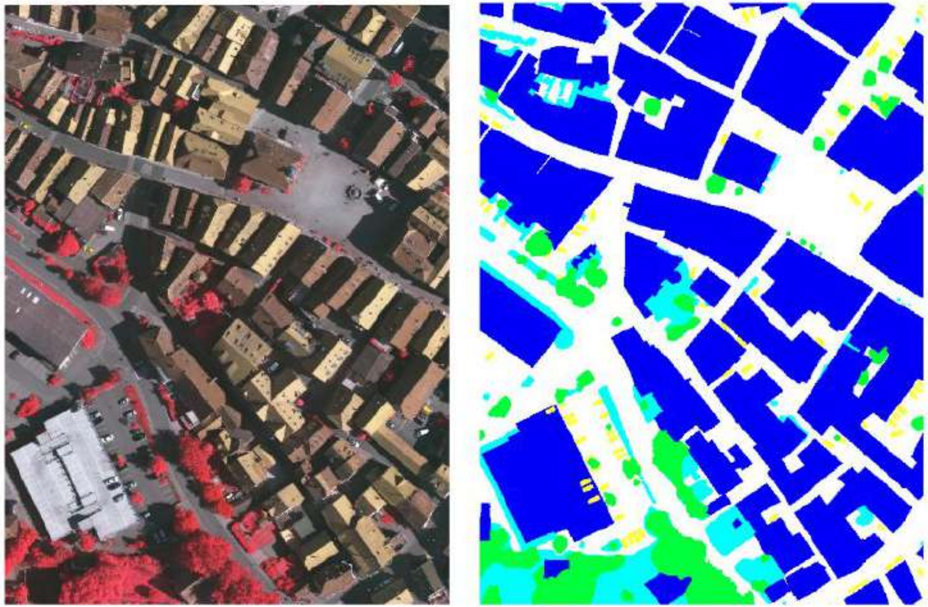
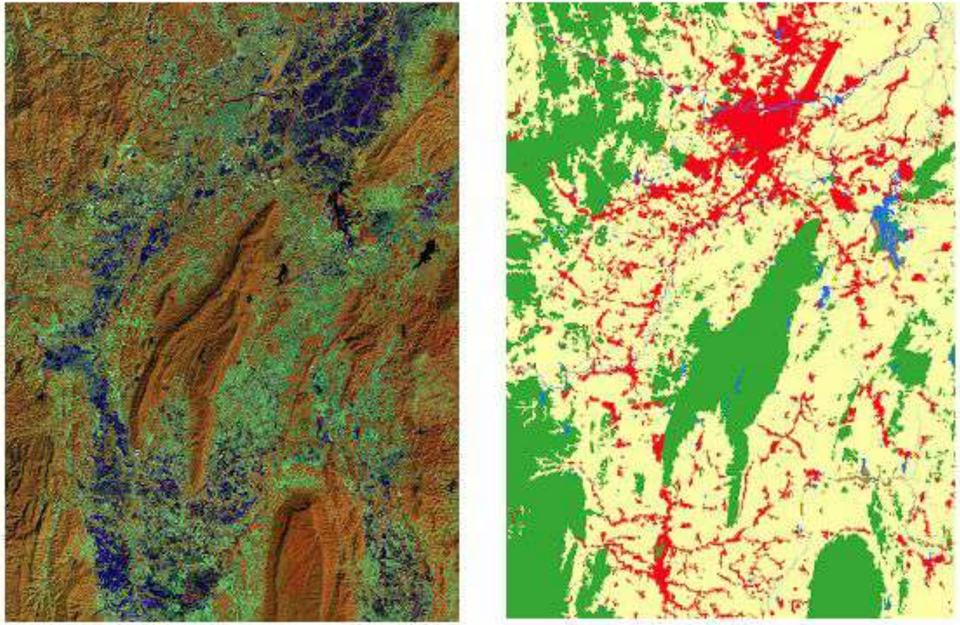


Figure 8. The sample input tile from Figure 7 (**left**) and corresponding ground truth (**right**). The label of the Vaihingen Challenge includes six categories: impervious surface (imp surf, white), building (blue), low vegetation (low veg, cyan), tree (green), car (yellow), and clutter/background (red).



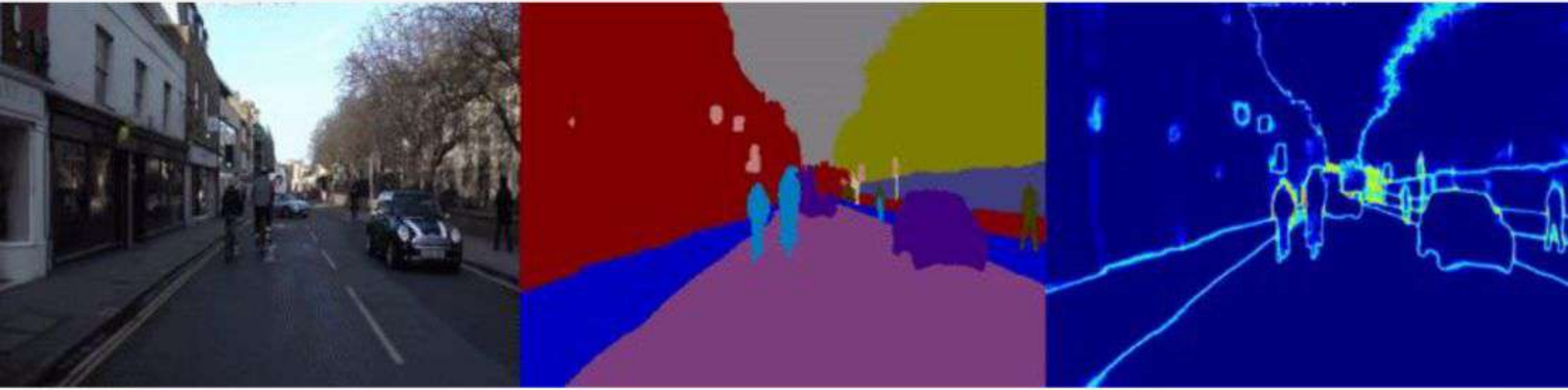
Public and Private Corpora

Private corpus (GISTDA Nan Province Corpus)



Color	Class
(light yellow)	Agriculture
(green)	Forest
(brown)	Miscellaneous
(red)	Urban
(blue)	Water

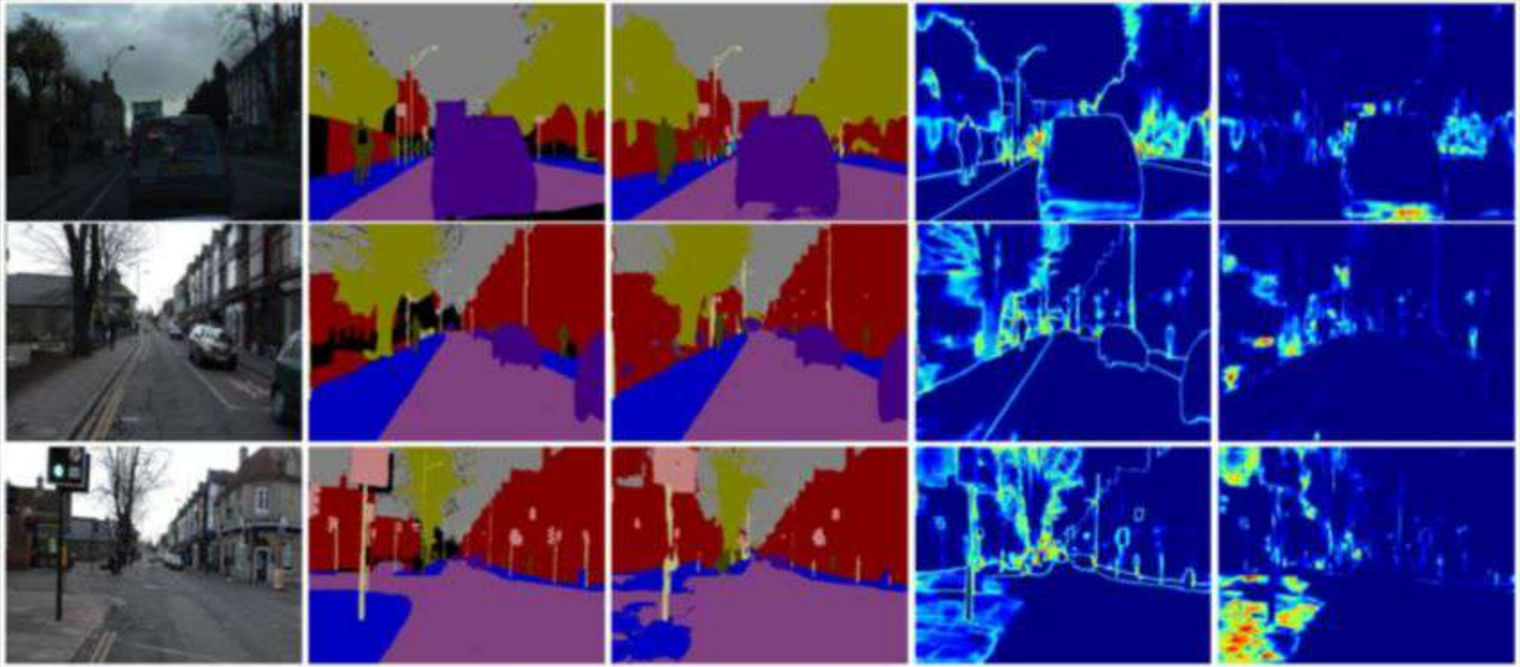
Bayesian SegNet for probabilistic scene understanding



Input Image Semantic Segmentation Uncertainty

What kind of uncertainty can we model?

Epistemic uncertainty is *modeling* uncertainty
Aleatoric uncertainty is *sensing* uncertainty



(a) Input Image (b) Ground Truth (c) Semantic Segmentation (d) Aleatoric Uncertainty (e) Epistemic Uncertainty

Modeling Uncertainty with Bayesian Deep Learning

? Deep learning is required to achieve state of the art results in computer vision applications but doesn't provide uncertainty estimates.

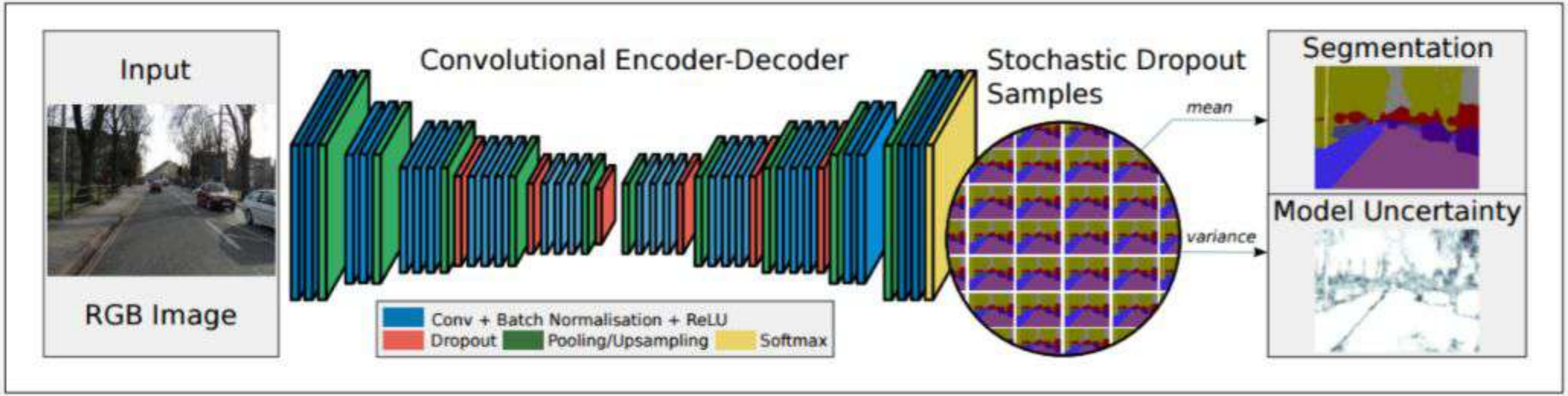
- **Bayesian neural networks** are a framework for understanding uncertainty in deep learning
- They have **distributions over network parameters** (rather than deterministic weights)
- Traditionally they have been **tricky to scale**

MacKay, David JC. **A practical Bayesian framework for backpropagation networks.** Neural Computation, 4(3): 448–472, 1992.

Modeling Epistemic Uncertainty with Bayesian Deep Learning

We can **model epistemic uncertainty** in deep learning models using Monte Carlo **dropout sampling** at test time.

Dropout sampling can be interpreted as **sampling from a distribution over models**.



Alex Kendall, Vijay Badrinarayanan and Roberto Cipolla **Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding**. arXiv preprint arXiv:1511.02680, 2015.

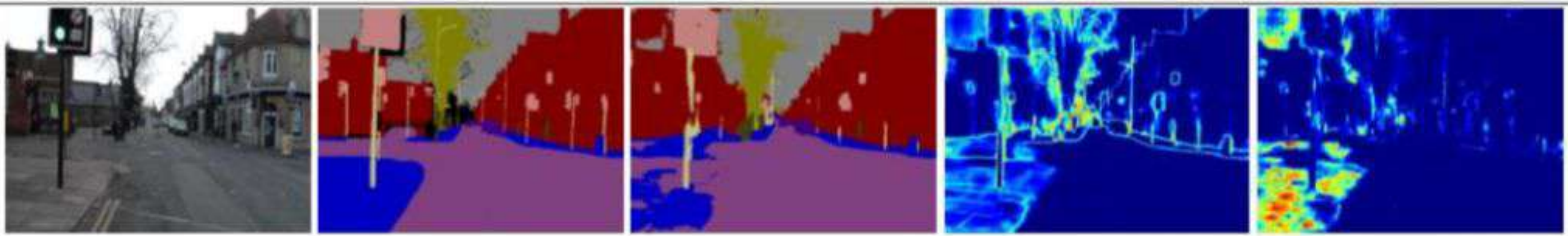
Modeling Aleatoric Uncertainty with Probabilistic Deep Learning

	Deep Learning	Probabilistic Deep Learning
Model	$[\hat{y}] = f(x)$	$[\hat{y}, \hat{\sigma}^2] = f(x)$
Regression	$Loss = \lVert y - \hat{y} \rVert^2$	$Loss = \frac{\lVert y - \hat{y} \rVert^2}{2\hat{\sigma}^2} + \log \hat{\sigma}^2$
Classification	$Loss = SoftmaxCrossEntropy(\hat{y}_t)$	$\hat{y}_t = \hat{y} + \epsilon_t \quad \epsilon_t \sim N(0, \hat{\sigma}^2)$ $Loss = \frac{1}{T}\sum_t SoftmaxCrossEntropy(\hat{y}_t)$

Alex Kendall and Yarin Gal. **What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?** arXiv preprint 1703.04977, 2017.

Semantic Segmentation Performance on CamVid

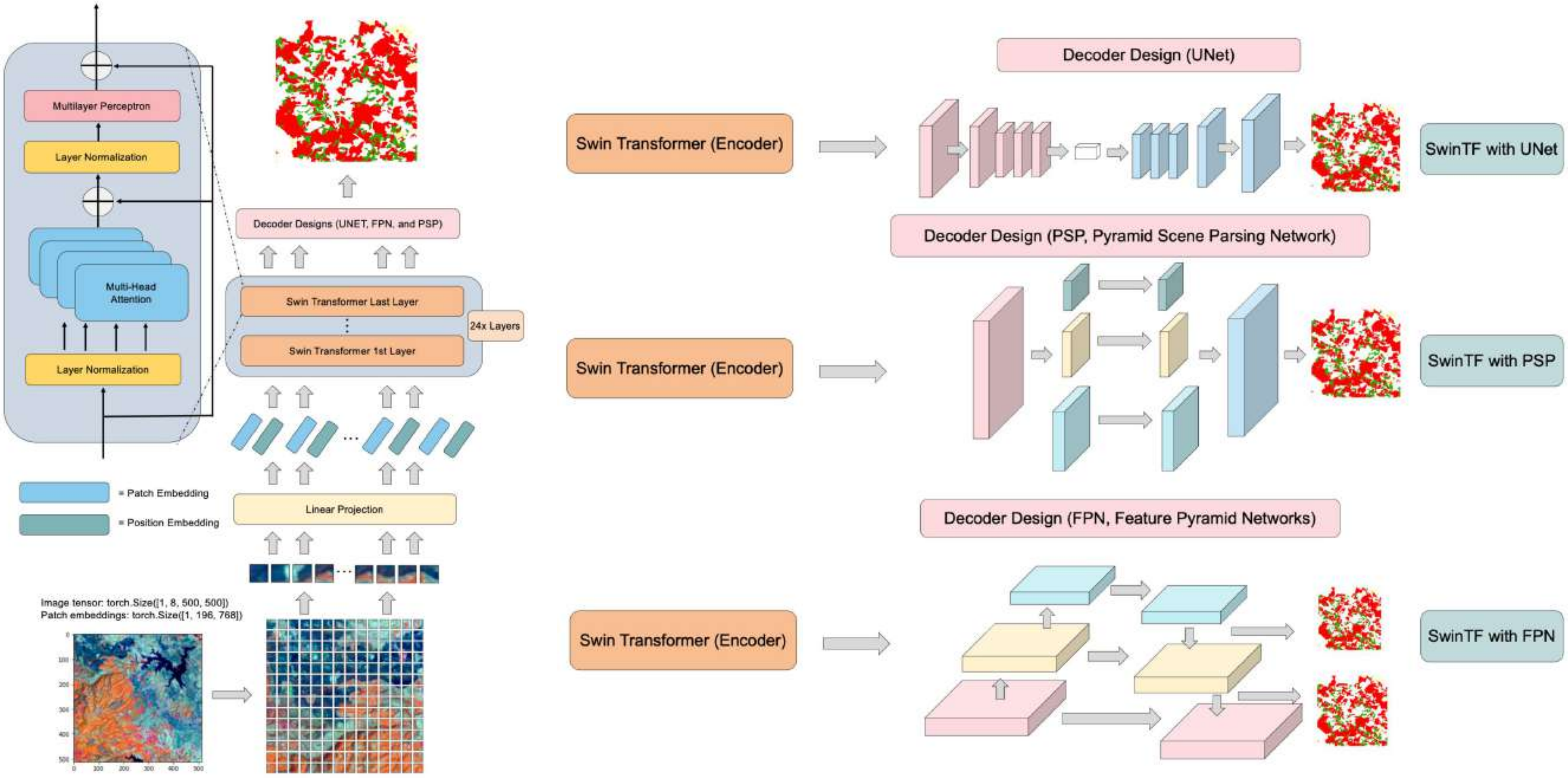
CamVid Results	IoU Accuracy
DenseNet (State of the art baseline)	67.1
+ Aleatoric Uncertainty	67.4
+ Epistemic Uncertainty	67.2
+ Aleatoric & Epistemic	**67.5**



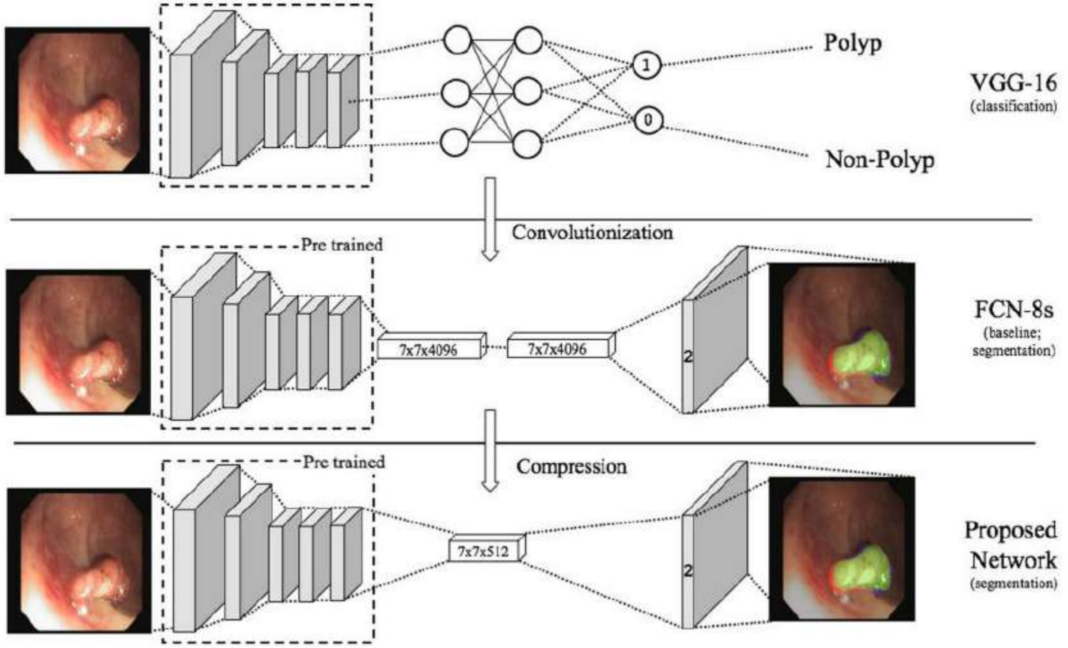
(a) Input Image (b) Ground Truth (c) Semantic Segmentation (d) Aleatoric Uncertainty (e) Epistemic Uncertainty

Alex Kendall and Yarin Gal. **What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?** arXiv preprint 1703.04977, 2017.

[3] Panboonyuen, Teerapong, et al. "Transformer-based decoder designs for semantic segmentation on remotely sensed images." Remote Sensing 13.24 (2021): 5100.



[4] Wichakam, I., **Panboonyuen, T.,** Udomcharoenchaikit, C., & Vateekul, P. (2018). **Real-time polyps segmentation for colonoscopy video frames using compressed fully convolutional network.** In MultiMedia Modeling: 24th International Conference, MMM 2018, Bangkok, Thailand, February 5-7, 2018, Proceedings, Part I 24 (pp. 393-404). Springer International Publishing.



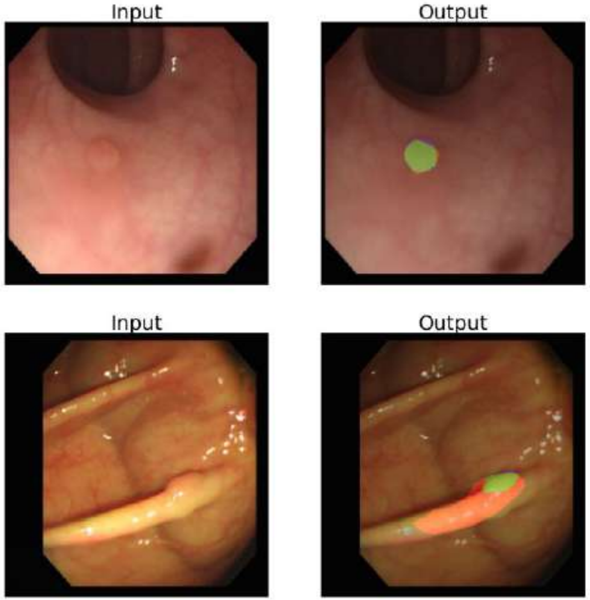
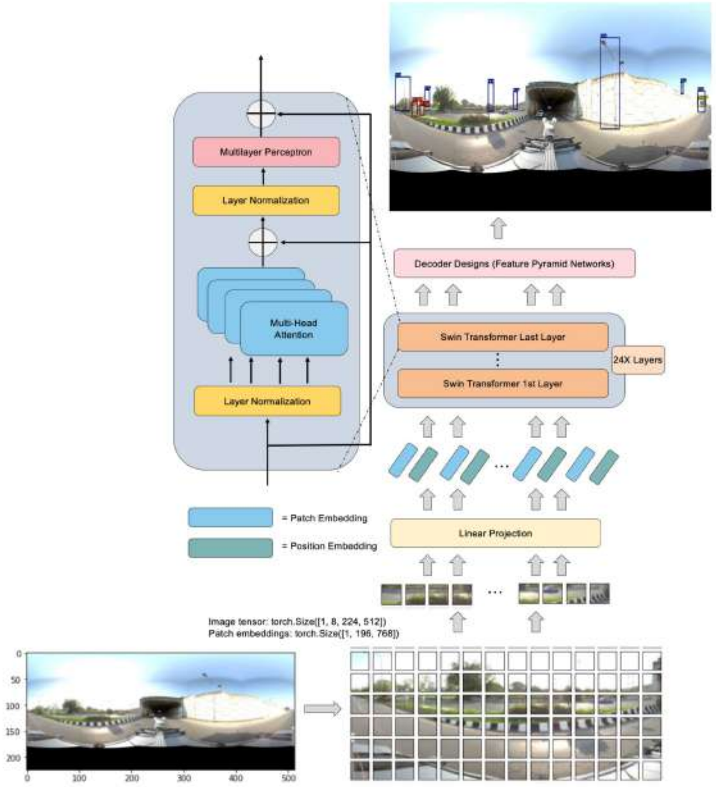
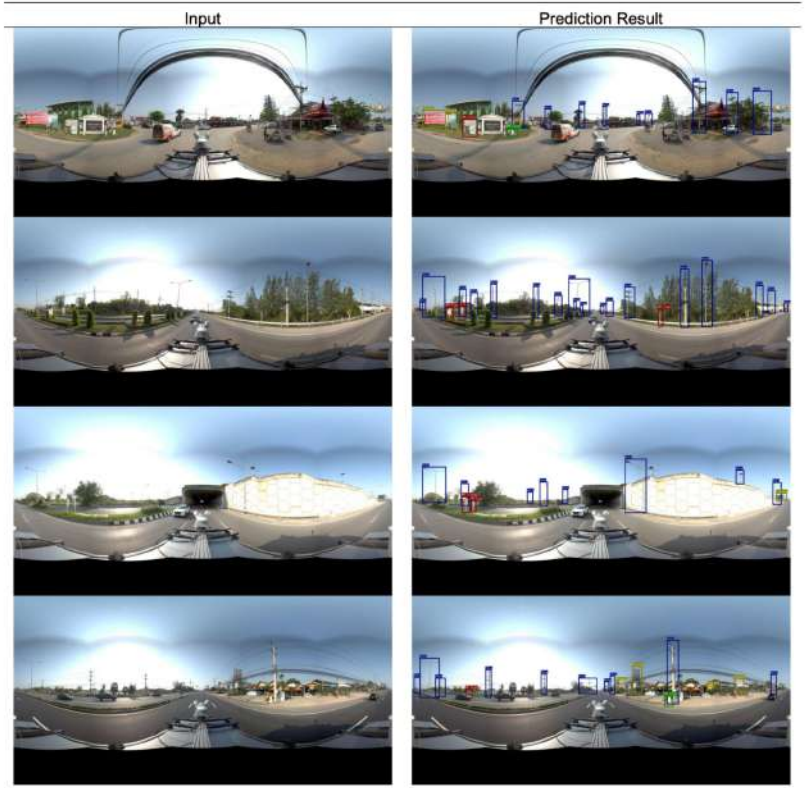


Fig. 1. Overview of our compressed network which is compressed from the original FCN-8s [8] based on VGG-16 [4] architecture.

[5] Panboonyuen, Teerapong, et al. **"Object detection of road assets using transformer-based YOLOX with feature pyramid decoder on thai highway panorama."** Information 13.1 (2022): 5.





[6] Thitisiriwech, K., **Panboonyuen, T.**, Kantavat, P., Iwahori, Y., & Kijsirikul, B. (2022). **The Bangkok Urbanscapes Dataset for Semantic Urban Scene Understanding Using Enhanced Encoder-Decoder With Atrous Depthwise Separable A1 Convolutional Neural Networks.** IEEE Access, 10, 59327-59349.

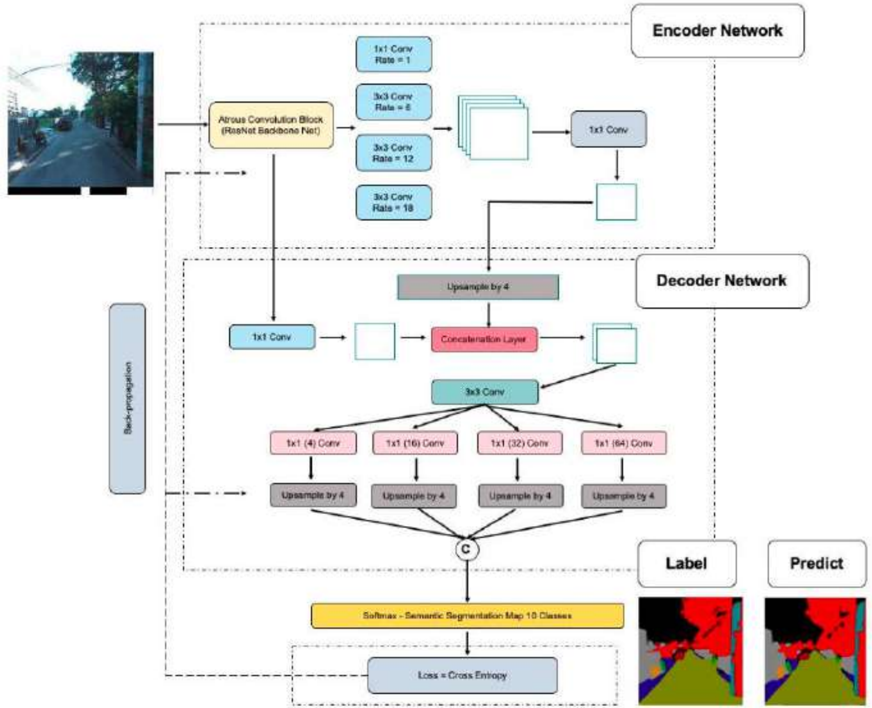
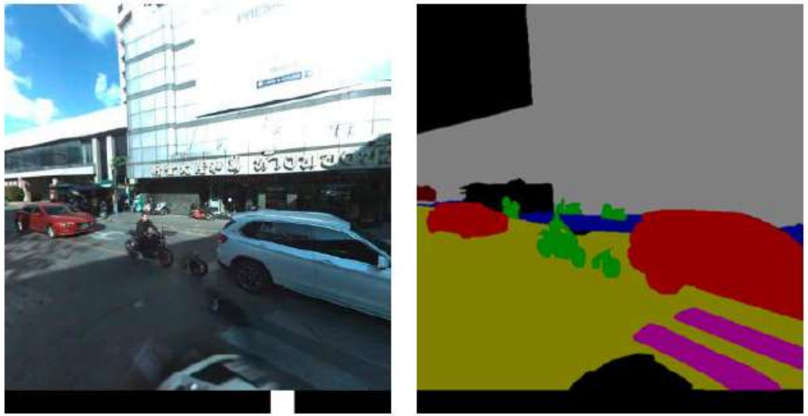


FIGURE 6. An overview of enhanced DeepLab-V3+ (Encoder-Decoder with atrous separable convolutional for semantic segmentation [15]) with ResNet-101 backbone [41] (DeepLab-V3-A1-ResNet-101).



(a) (b)

FIGURE 8. Sample 1: The example of Sukhumvit's large road from the training set of the Bangkok Urbanscapes dataset. The input image is shown in (a), and the ground truth is shown in (b).

Void	Building	Wall	Tree	VegetationMisc	Fence
Sidewalk	ParkingBlock	Column_Pole	TrafficCone	Bridge	SignSymbol
Misc_Text	TrafficLight	Sky	Tunnel	Archway	Road
RoadShoulder	LaneMkgsDriv	LaneMkgsNonDriv	Animal	Pedestrian	Child
CartLuggagePram	Bicyclist	MotorcycleScooter	Car	SUVPickupTruck	Truck_Bus
Train	OtherMoving				

FIGURE 3. The semantic color codes of the CamVid dataset. Each color is encoded with respect to the semantic class in the ground truth images.

[6] Vajeethaveesin, T., **Panboonyuen, T.**, Lawawironjwong, S., Srestasathiern, P., Jaiyen, S., & Jitkajornwanich, K. (2022). **A performance comparison between GIS-based and neuron network methods for flood susceptibility assessment in ayutthaya province.** Trends in Sciences, 19(2), 2038-2038.

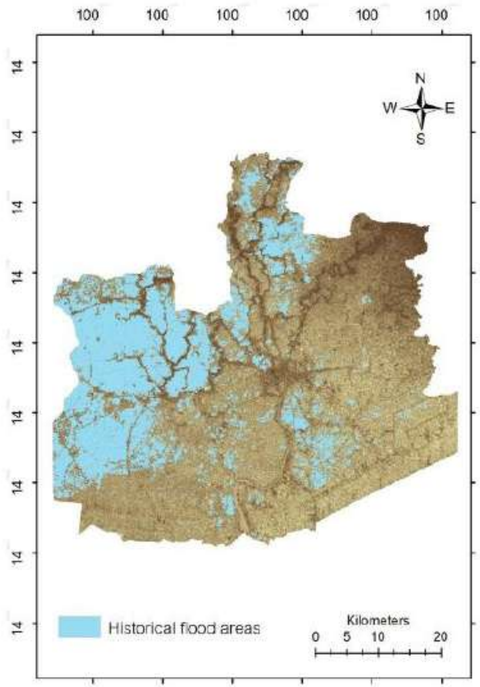


Figure 2 The historical flood zone shows flood events, which were captured in 2016 from GISTDA. Most of the flooding areas are located in the west of the study area.

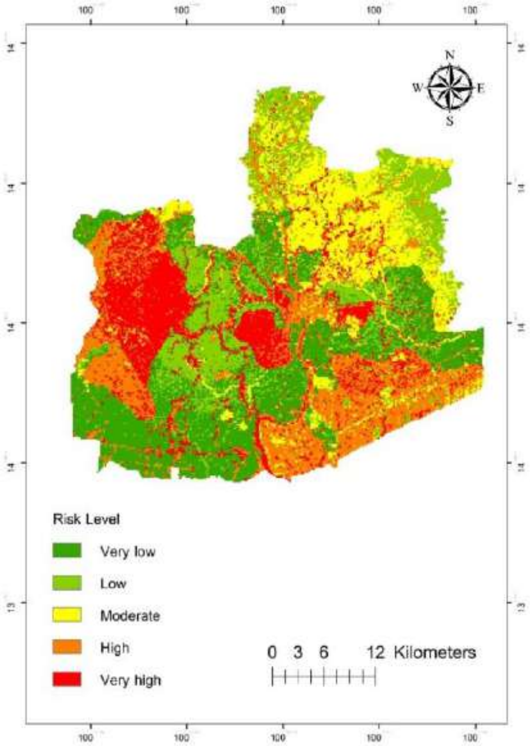


Figure 4 Flood susceptibility map from the flood risk assessment model (FRAM).

Classification of Primary Angle Closure Glaucoma (PACG) Using Deep Convolutional Neural Networks

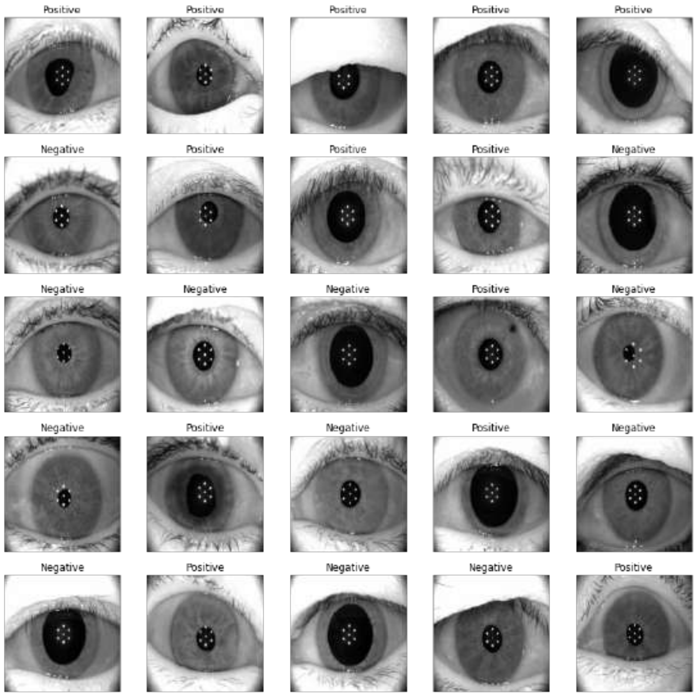


Table 4: Performance comparison on F1 score with existing models on the WTW-LAF23 dataset.

Model	F1 score			
	Mean	SD	Min	Max
VGG19 [Dutta et al., 2016]	61.34	4.12	54.67	65.76
DenseNet121 [Targ et al., 2016]	61.67	4.01	54.89	65.91
RestNet50 [Huang et al., 2017]	**65.93**	6.11	56.11	**76.33**
Transformer [Liu et al., 2021]	58.18	5.07	51.34	69.83
Our Proposed Method (SimCLR Chen et al. [2020])	**67.82**	N/A	N/A	N/A

Table 1: Details of used visual datasets.

Dataset	Type	Positive Size	Negative Size	Train Size	Val Size	# Classes
LAFx10_spss_Frequency2.3	Classification	612	612	1,101	123	2
AQD_Frequency2.4	Classification	3,338	3,153	5,841	650	2

Table 2: Details of used visual datasets.

Dataset	Type	Female Size	Male Size	Unknown	Other	Train Size	Val Size	# Classes
GENDER	Classification	3,257	2,106	25	1	3,754	804	2

Table 3: Details of used visual datasets.

Dataset	Type	Young Adults	Middle-Aged	Old Adults	Train Size	Val Size	# Classes
AGE	Classification	224	3,885	1,271	3,765	807	3

https://www.marssolution.io/

หน้าแรก | โซลูชัน | พันธมิตร | ข่าวสาร | เกี่ยวกับมาร์ส | ติดต่อเรา

มาร์สคือผู้นำในการพัฒนาระบบนิเวศสำหรับผู้ใช้รถยนต์ ปลดล็อกการตรวจสภาพรถยนต์แบบเดิมด้วยการใช้ AI **ให้การตรวจสภาพรถยนต์เป็นเรื่องที่ง่ายกว่า**

มาร์สสามารถเชื่อมโยงธุรกิจในหลากหลายอุตสาหกรรมกับผู้ใช้รถยนต์ให้เกิดความสะดวกและรวดเร็วยิ่งขึ้น มาร์สหรือ MARS (Motor AI Recognition Solution) เป็นแอปพลิเคชันที่ปลดล็อกการตรวจสภาพรถยนต์แบบเดิมๆ ด้วยเทคโนโลยี AI แบบครบวงจรที่มาร์สได้คิดค้นและพัฒนาขึ้นมาเพื่อช่วยให้ธุรกิจของคุณลดต้นทุนด้านบุคลากร ลดขั้นตอนการทำงาน เพิ่มความรวดเร็วและแม่นยำในการตรวจสอบและจัดเก็บข้อมูล

ดูโซลูชันทั้งหมด | เกี่ยวกับมาร์ส

https://www.marssolution.io/

ดาวน์โหลด MARS Inspect หรือ “มาตรวจ” วันนี้

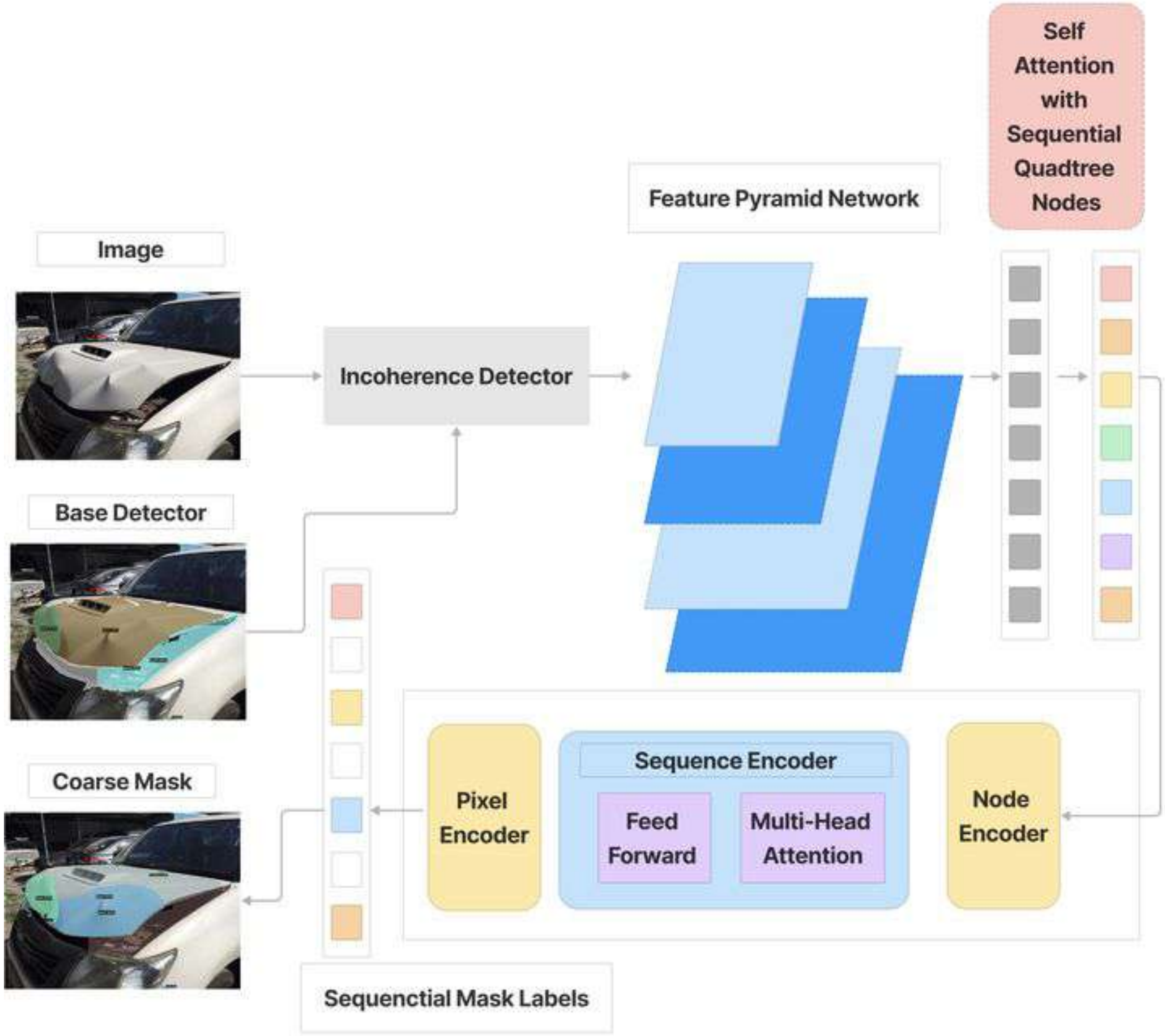
ให้เรื่องการตรวจรถยนต์เป็นเรื่องง่ายในมือคุณ เพียงแค่ โหลด เตรียม ถ่าย

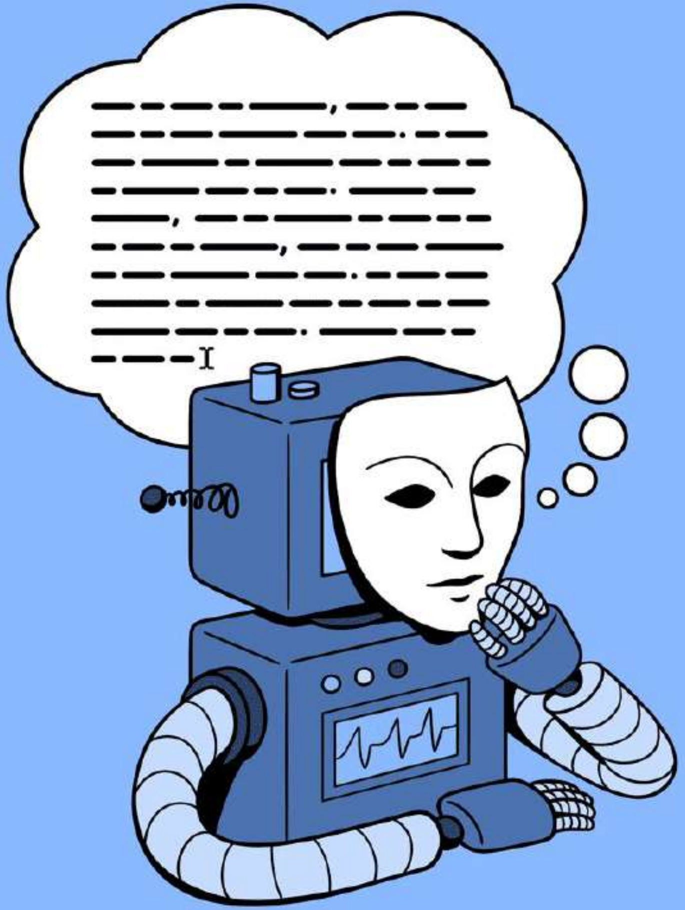
Download on the App Store

GET IT ON Google Play

https://www.marssolution.io/

Panboonyuen, Teerapong, et al. ***"MARS: Mask Attention Refinement with Sequential Quadtree Nodes for Car Damage Instance Segmentation."*** *International Conference on Image Analysis and Processing*. Cham: Springer Nature Switzerland, 2023.



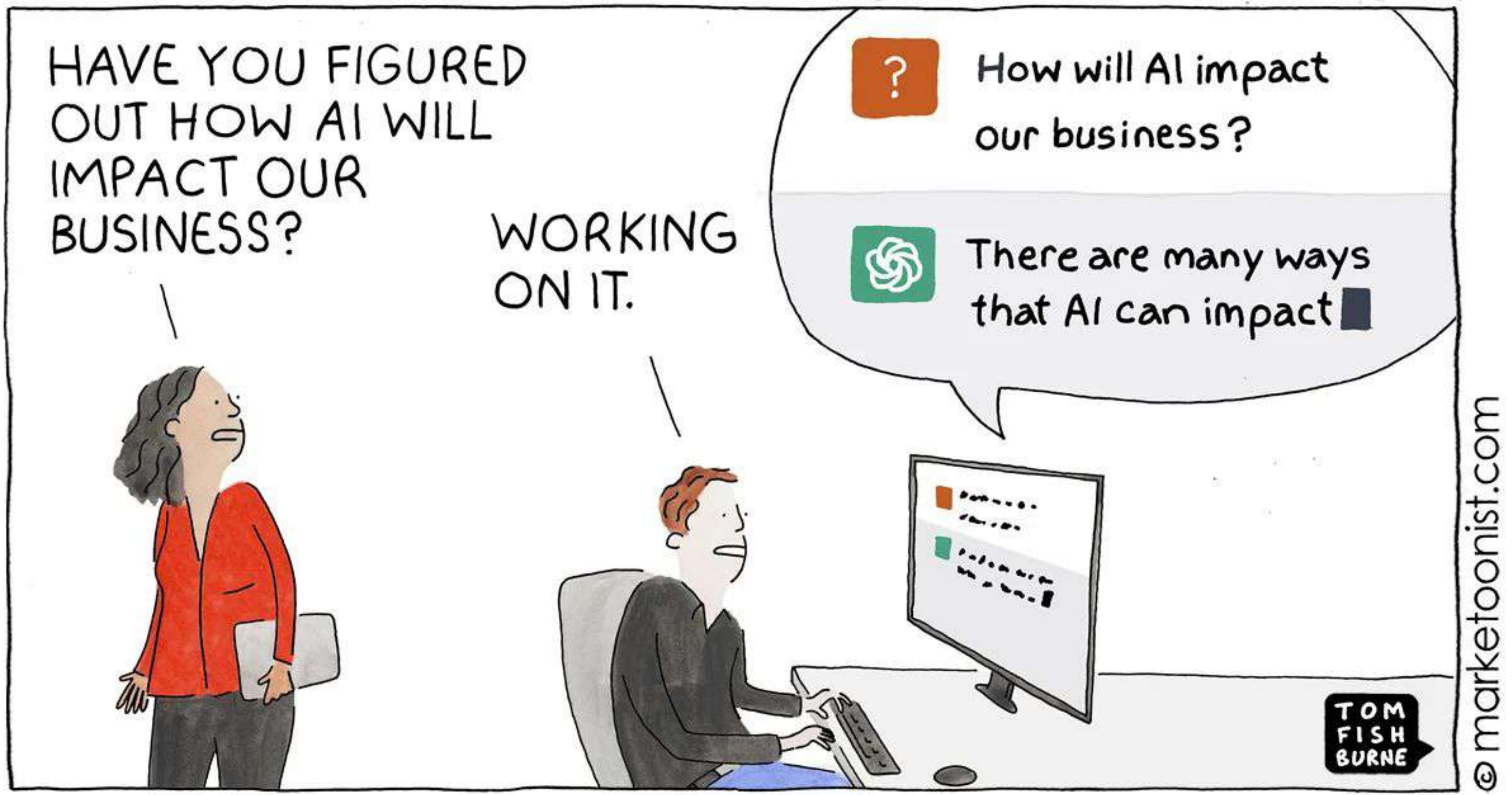


Large Language Model (LLM)

['lärj 'laŋ-gwij 'mä-dəl]

A deep learning algorithm that's equipped to summarize, translate, predict, and generate human-sounding text to convey ideas and concepts.

Investopedia


HAVE YOU FIGURED OUT HOW AI WILL IMPACT OUR BUSINESS?
WORKING ON IT.
How will AI impact our business?
There are many ways that AI can impact
TOM FISH BURNE
© marketoonist.com

Understanding Multimodal

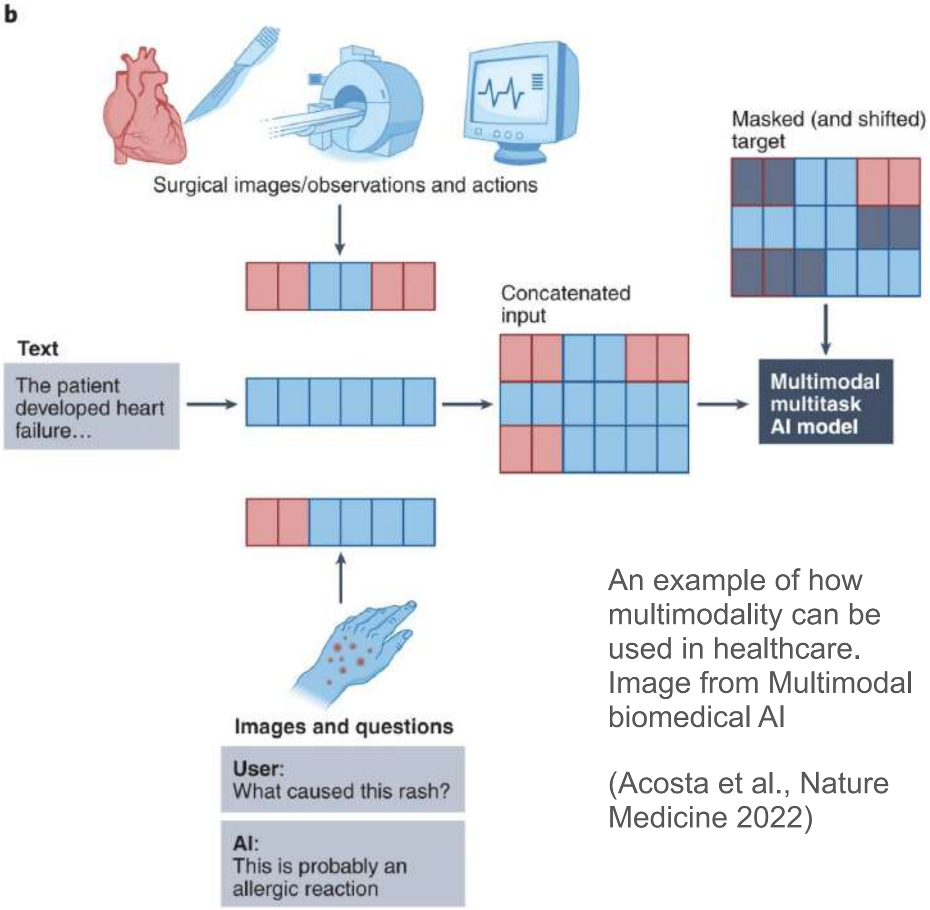
Chip Huyen
@chipro

New blog post: Multimodality and Large Multimodal Models (LMMs)

Being able to work with data of different modalities -- e.g. text, images, videos, audio, etc. -- is essential for AI to operate in the real world.

This post covers multimodal systems in general, including Large Multimodal Models. It consists of 3 parts.

* Part 1 covers the context for multimodality.
* Part 2 discusses how to train a multimodal system, using the architectures of CLIP and Flamingo, and examples from GPT-4V.
* Part 3 discusses some active research areas for LMMs, including generating multimodal outputs.

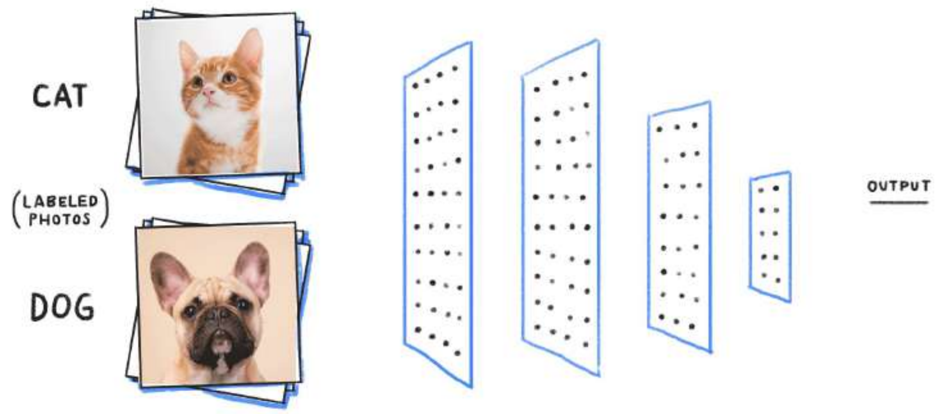


An example of how multimodality can be used in healthcare. Image from Multimodal biomedical AI

(Acosta et al., Nature Medicine 2022)

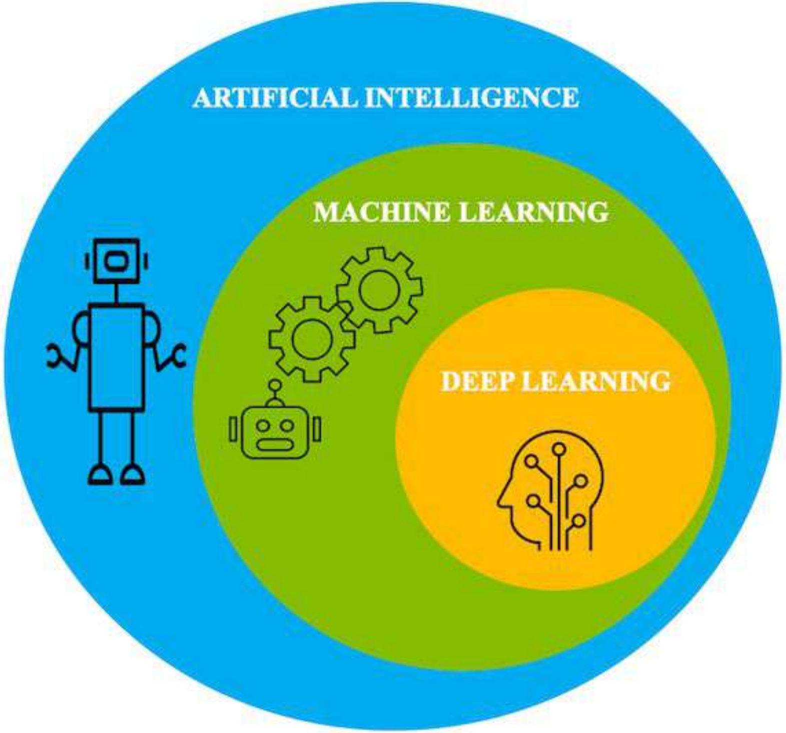
https://huyenchip.com/2023/10/10/multimodal.htm
https://twitter.com/chipro/status/1711970025874321479/photo/1

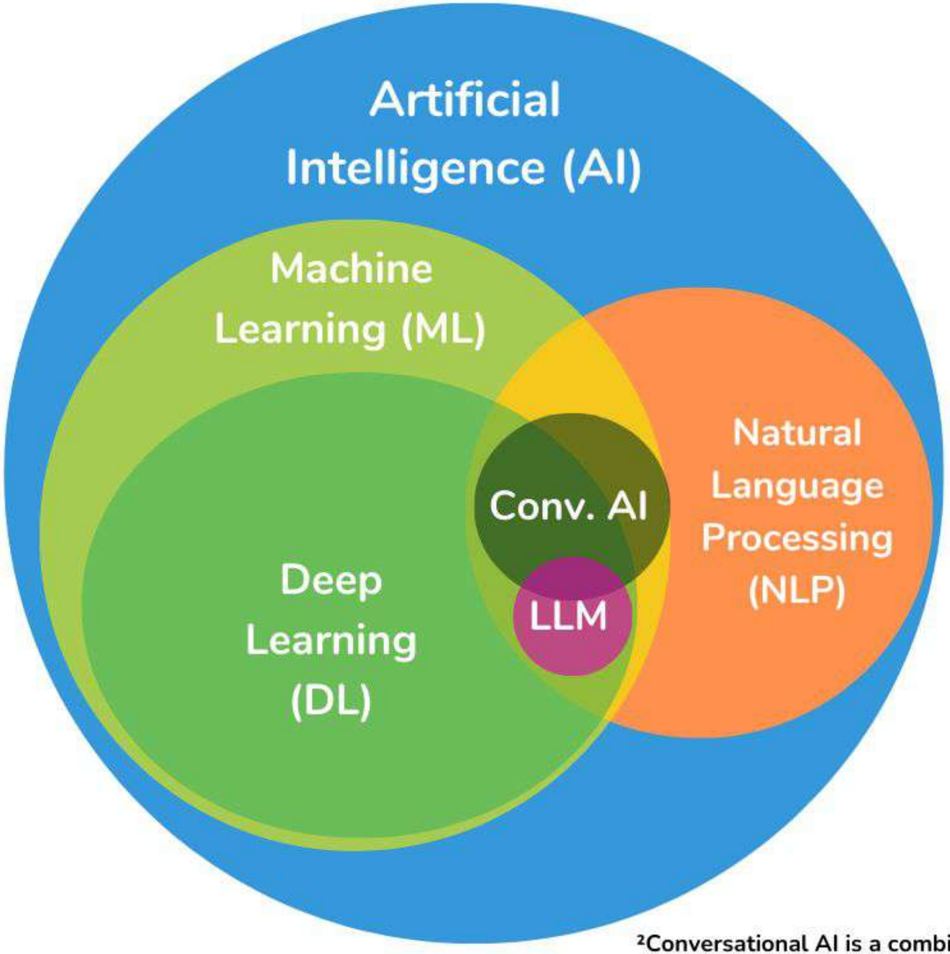
Basic AI



AI vs Machine Learning vs Deep Learning

- **Artificial Intelligence (AI)** is the concept of creating smart intelligent machines.
- **Machine Learning (ML)** is a subset of artificial intelligence that helps you build AI-driven applications.
- **Deep Learning** is a subset of machine learning that uses vast volumes of data and complex algorithms to train a model.



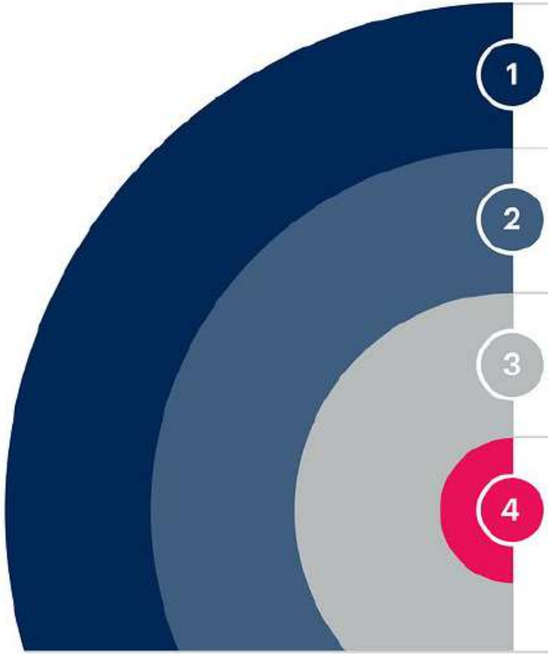


- Artificial Intelligence (AI)
- Machine Learning (ML)
- Deep Learning (DL)
- Natual Language Processing (NLP)
- Large Language Model (LLM)[1]
- Conversational AI (Conv. AI)[2]

[1]LLM is an intersection of DL and NLP

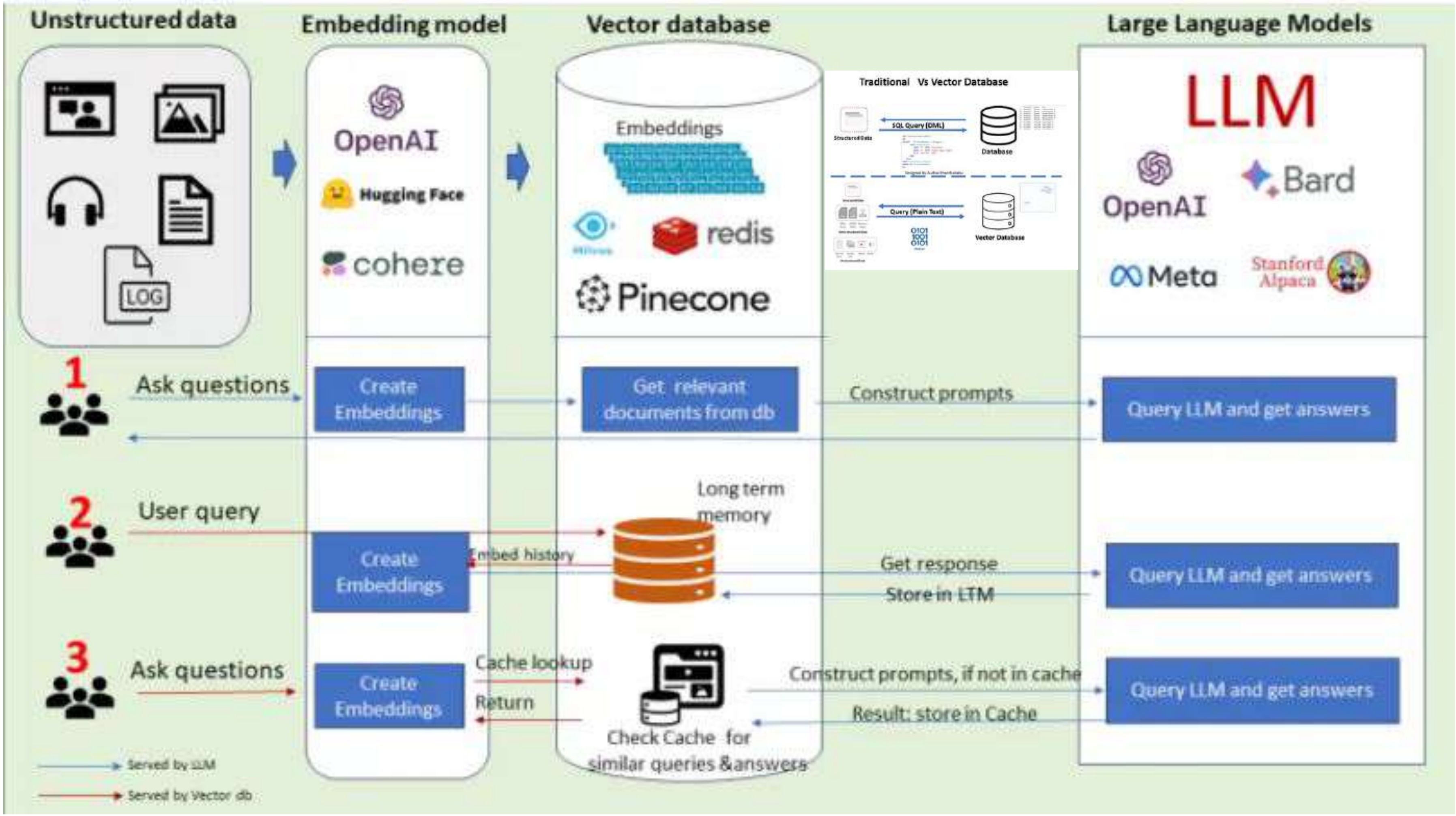
[2]Conversational AI is a combination of ML and NLP. It may include DL and LLM, but that isn't always the case.

What Is Generative AI?



1. **Generative AI** — AI techniques that learn from data about existing artifacts and use this to generate new artifacts.
2. **Foundation Models** — Large machine learning models that are trained on a broad set of unlabeled data and are fine-tuned to a wide range of applications.
3. **Large Language Models (LLM)** — AI that is trained on vast amounts of text to interpret and generate human-like textual output.
4. **ChatGPT** — An OpenAI service that incorporates a conversational chatbot with LLM to create content.

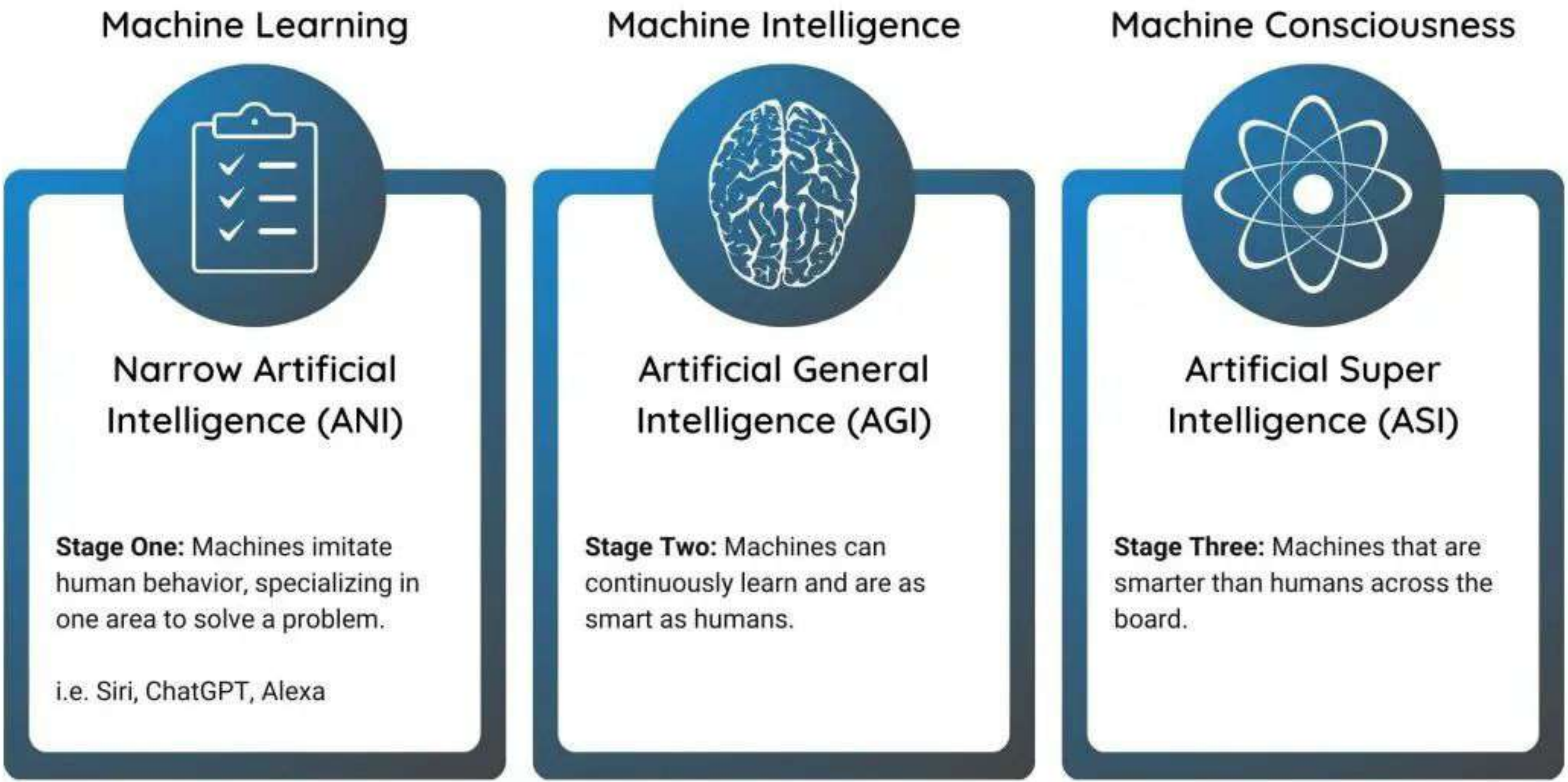
Source: Gartner



Unstructured data
Embedding model
Vector database
Large Language Models
OpenAI
Hugging Face
cohere
Embeddings
redis
Pinecone
Traditional Vs Vector Database
SQL Query (DML)
Database
Query (Plain Text)
Vector Database
LLM
OpenAI
Bard
Meta
Stanford Alpaca
1
Ask questions
Create Embeddings
Get relevant documents from db
Construct prompts
Query LLM and get answers
2
User query
Long term memory
Create Embeddings
Embed history
Get response
Store in LTM
Query LLM and get answers
3
Ask questions
Create Embeddings
Cache lookup
Return
Check Cache for similar queries &answers
Construct prompts, if not in cache
Result: store in Cache
Query LLM and get answers
Served by LLM
Served by Vector db

Difference between Artificial Intelligence and Artificial General Intelligence (AGI)

Artificial Intelligence (AI)	Artificial General Intelligence (AGI)
• AI mimics human intelligence. • AI is trained by data scientists on specific, singular, or a limited number of tasks. Examples: image recognition, speech-to-text transcription, text generation, chatbots. • Is not self-aware; has no consciousness or ability to think on its own.	• AGI equals human intelligence, theoretically. • AGI possesses common sense and creativity and expresses emotions. • Has the ability to learn, apply knowledge, generalize, and plan ahead. • Not fully realized yet. Some doubt if it ever will be.


Machine Learning
Narrow Artificial Intelligence (ANI)
Stage One: Machines imitate human behavior, specializing in one area to solve a problem.
i.e. Siri, ChatGPT, Alexa
Machine Intelligence
Artificial General Intelligence (AGI)
Stage Two: Machines can continuously learn and are as smart as humans.
Machine Consciousness
Artificial Super Intelligence (ASI)
Stage Three: Machines that are smarter than humans across the board.

Meta scientist Yann LeCun says AI won't destroy jobs forever

15 June 2023

Share

Chris Vallance
Technology reporter

Prof Yann LeCun is known as one of the three godfathers of AI and works as Facebook-owner Meta's top AI scientist

https://www.tiktok.com/@20vc_tok/video/7233812824508599557

Forward-Forward Algorithm: Will it replace Backpropagation?

The Forward-Forward Algorithm: Some Preliminary Investigations

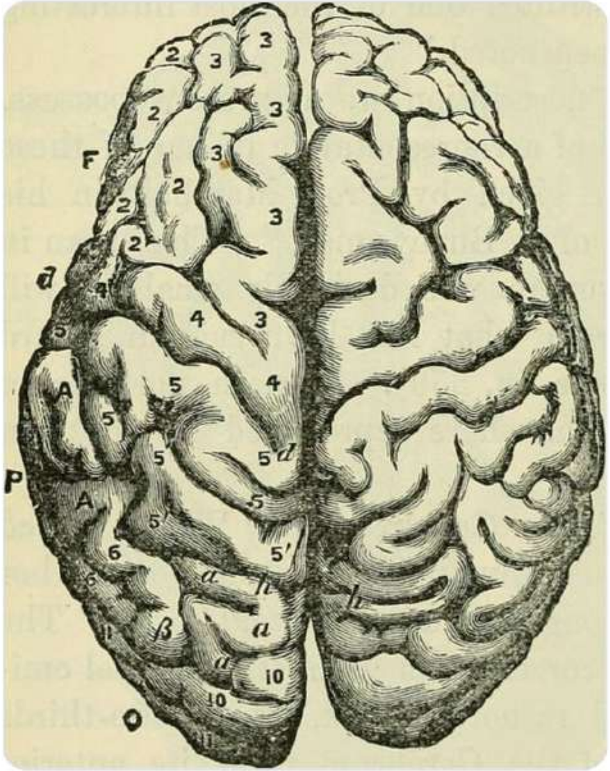
Geoffrey Hinton
Google Brain
geoffhinton@google.com

Abstract

The aim of this paper is to introduce a new learning procedure for neural networks and to demonstrate that it works well enough on a few small problems to be worth serious investigation. The Forward-Forward algorithm replaces the forward and backward passes of backpropagation by two forward passes, one with positive (*i.e.* real) data and the other with negative data which could be generated by the network itself. Each layer has its own objective function which is simply to have high goodness for positive data and low goodness for negative data. The sum of the squared activities in a layer can be used as the goodness but there are many other possibilities, including minus the sum of the squared activities. If the positive and negative passes can be separated in time, the negative passes can be done offline, which makes the learning much simpler in the positive pass and allows video to be pipelined through the network without ever storing activities or stopping to propagate derivatives.

Martin Görner @martin_gorner · Dec 5, 2022

I seems very unlikely that the human brain uses back propagation to learn. There is little evidence of backprop mechanics in biological brains (no error derivatives propagating backwards, no storage of neuron activities to use in a packprop pass, ...).



Martin Görner
@martin_gorner

Also, the brain can learn from a continuous stream of incoming data and does not need to stop to run a backprop pass. Yes, sleep is beneficial for learning somehow, but we can learn awake too.

https://jalammar.github.io/illustrated-transformer

The Illustrated Transformer

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

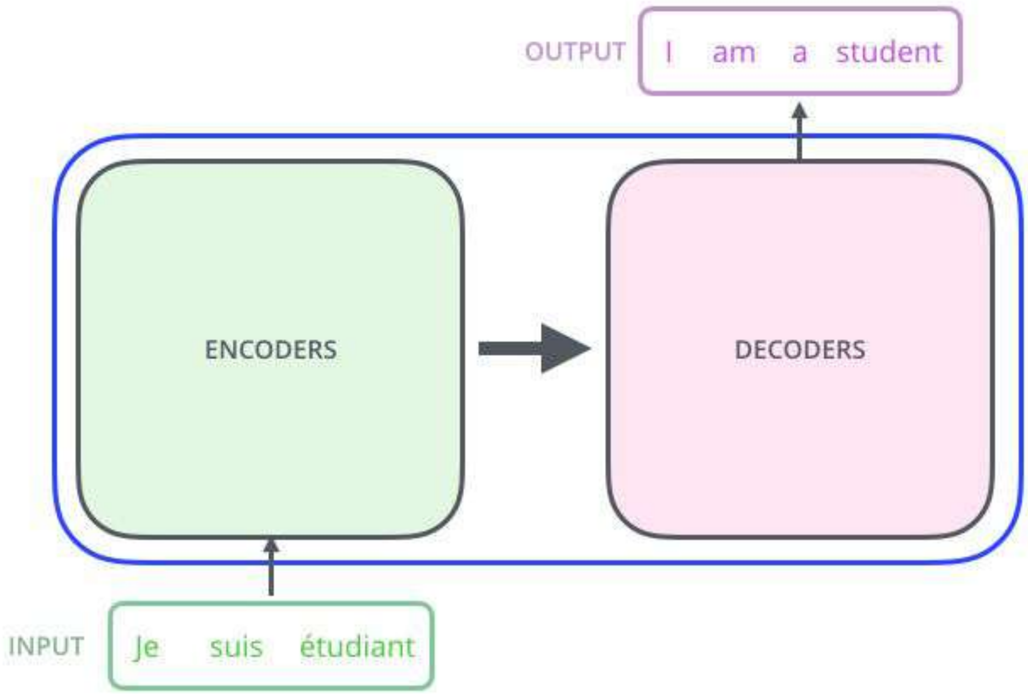
Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.0 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature.

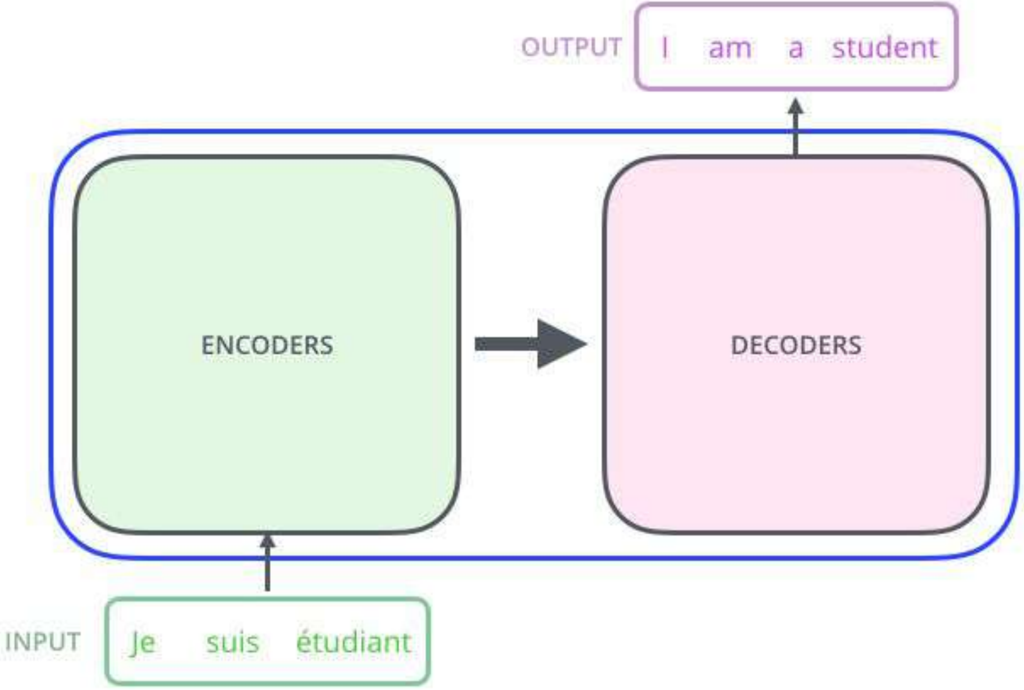
https://jalammar.github.io/illustrated-transformer

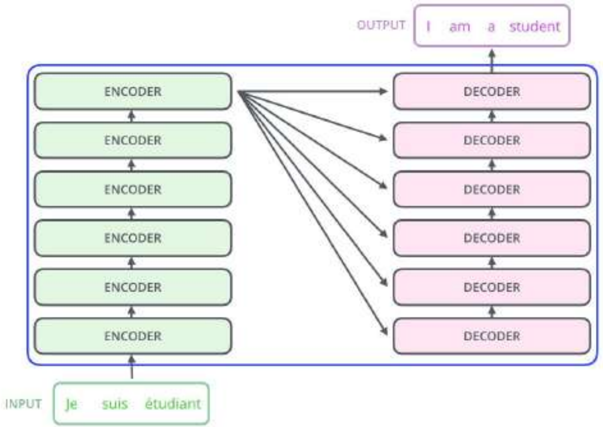
The Illustrated Transformer

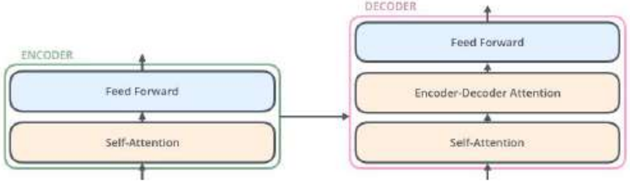


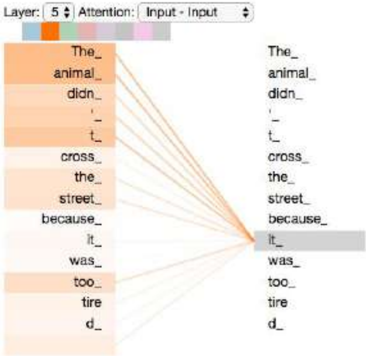
https://jalammar.github.io/illustrated-transformer

The Illustrated Transformer



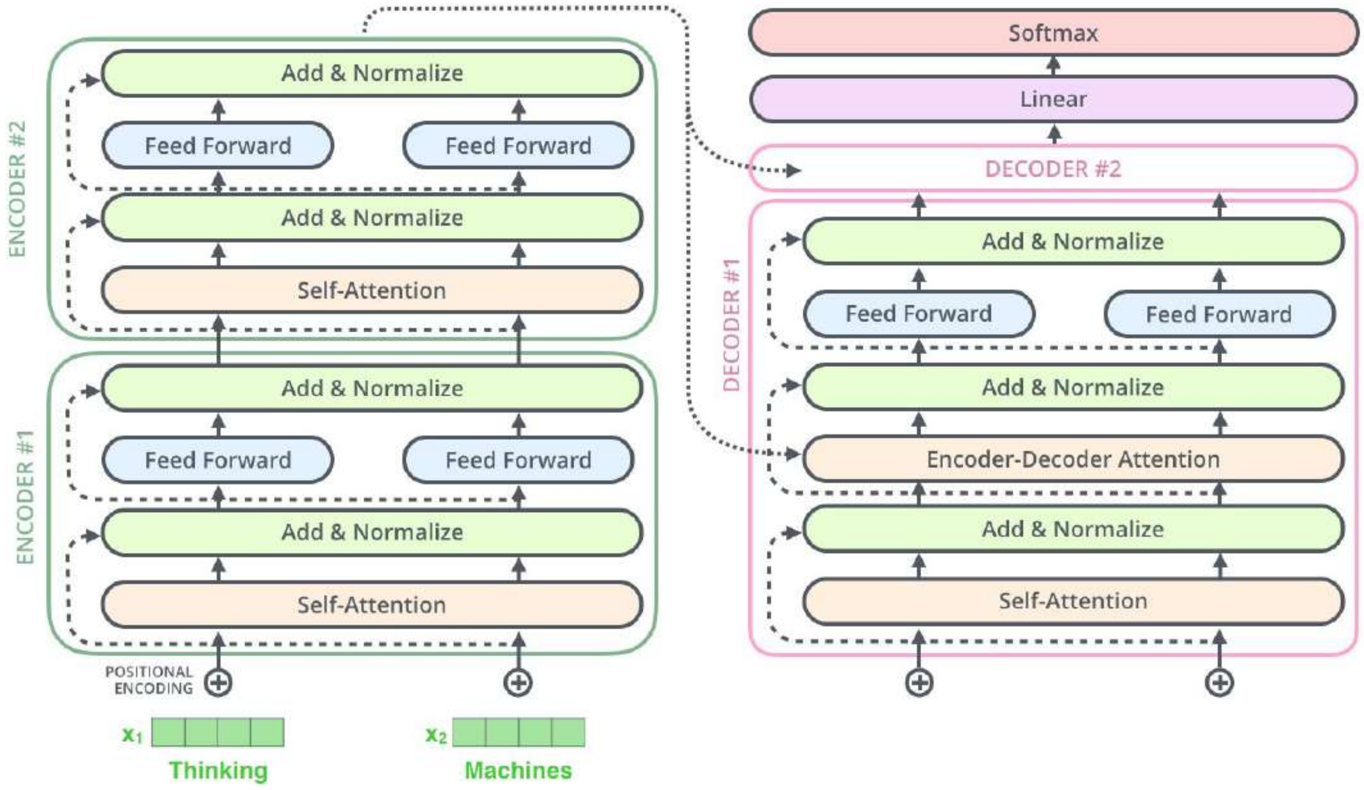




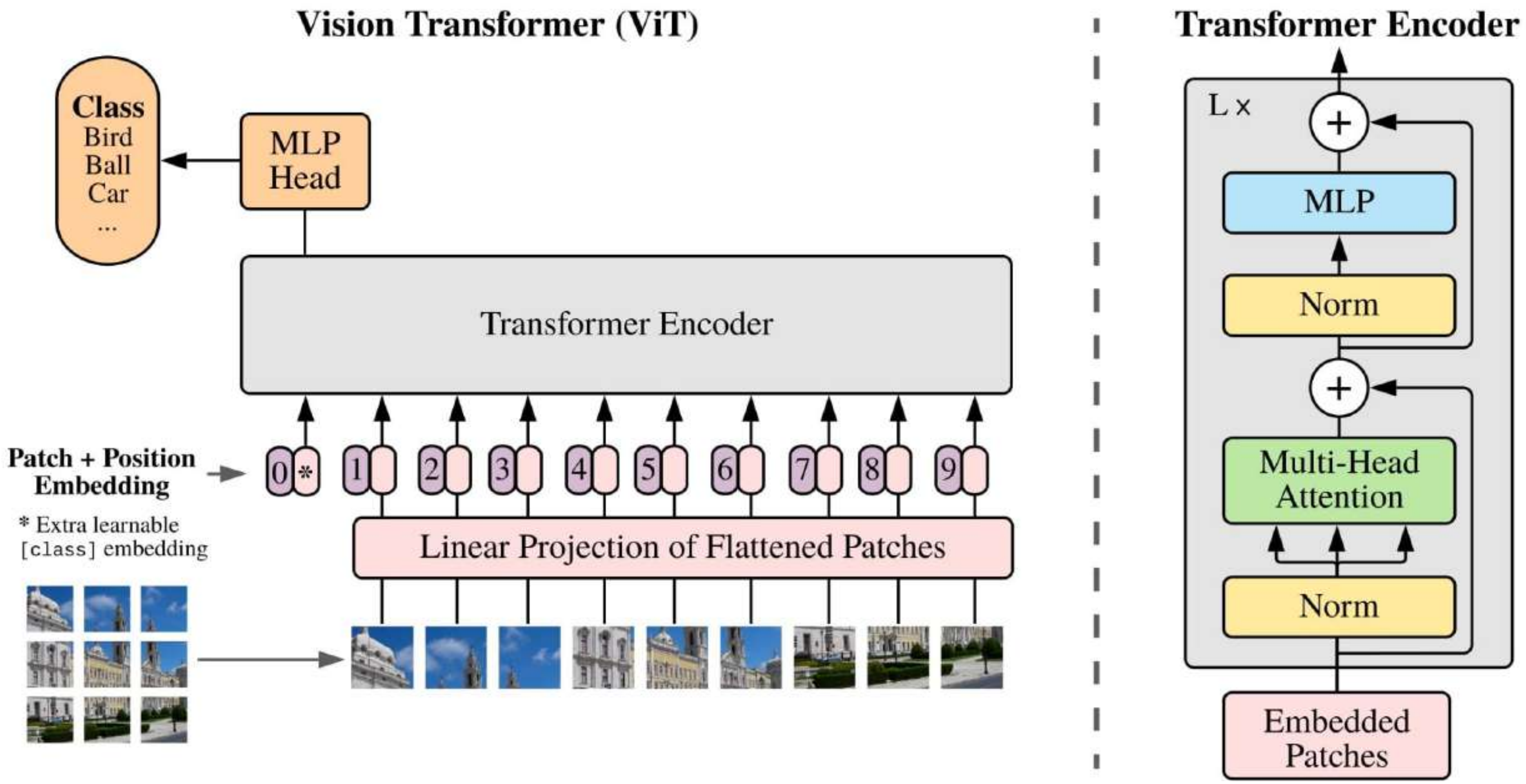


https://jalammar.github.io/illustrated-transformer

The Illustrated Transformer

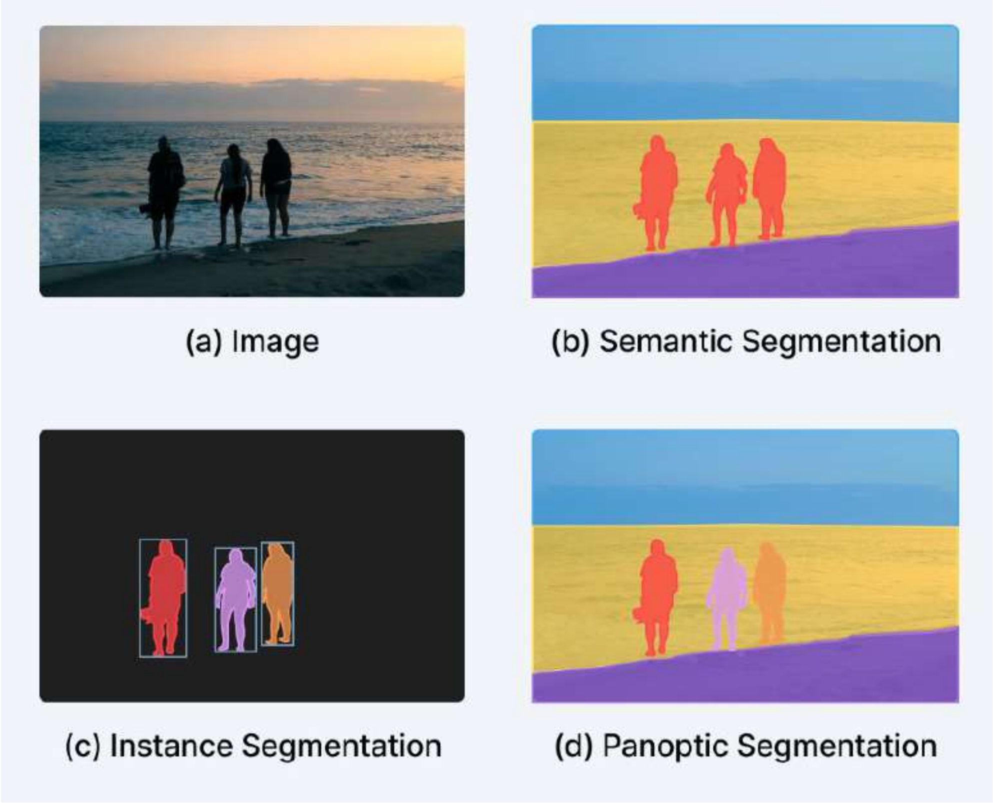


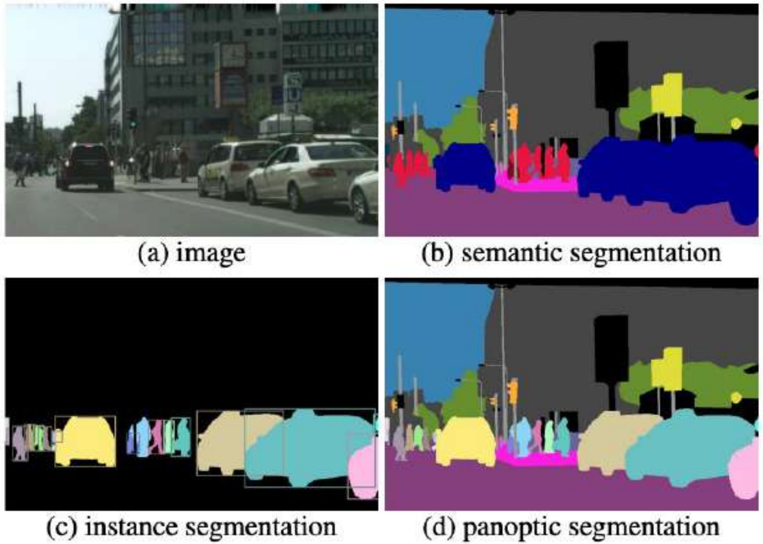
Vision Transformer (ViT)



Panoptic Segmentation

Alexander Kirillov[1,2] Kaiming He[1] Ross Girshick[1] Carsten Rother[2] Piotr Dollár[1]
[1]Facebook AI Research (FAIR) [2]HCI/IWR, Heidelberg University, Germany





Stuff & things in panoptic segmentation

Things: In the realm of computer vision, "things" typically refer to objects that

$$\mathrm{PQ} = \underbrace{\frac{\sum_{(p,g)\in TP} \mathrm{IoU}(p,g)}{|TP|}}_{\text{segmentation quality (SQ)}} \times \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{recognition quality (RQ)}}.$$

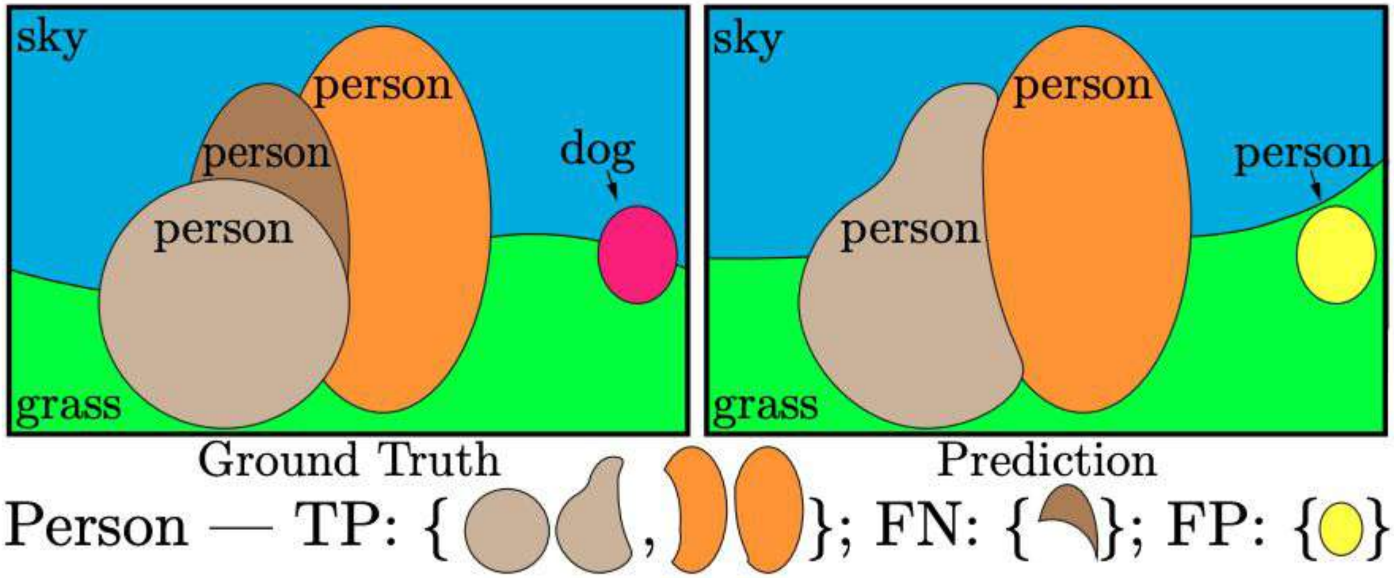


Figure 2: Toy illustration of ground truth and predicted panoptic segmentations of an image. Pairs of segments of the same color have IoU larger than 0.5 and are therefore matched. We show how the segments for the *person* class are partitioned into true positives TP, false negatives FN, and false positives FP.

$$\mathrm{PQ} = \underbrace{\frac{\sum_{(p,g)\in TP} \mathrm{IoU}(p,g)}{|TP|}}_{\text{segmentation quality (SQ)}} \times \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{recognition quality (RQ)}}.$$

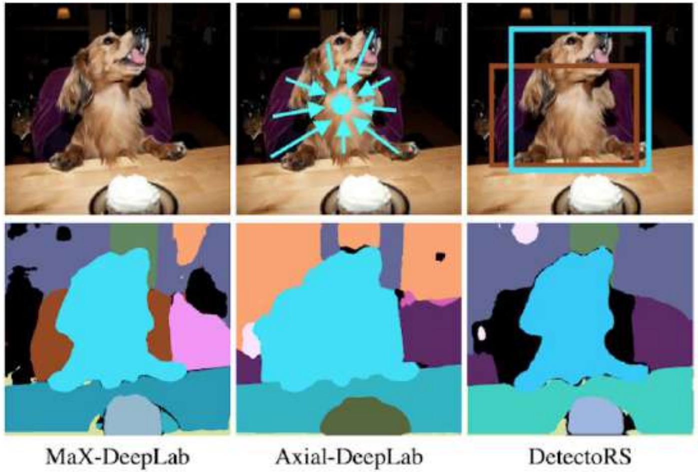
https://blog.research.google/2021/04/max-deeplab-dual-path-transformers-for.html

Google Research

MaX-DeepLab: Dual-Path Transformers for End-to-End Panoptic Segmentation

Method		PQ
Box-based	DETR	46.0 (-5.3)
	DetectoRS	49.6 (-1.7)
Box-free	Panoptic-DeepLab	41.4 (-9.9)
	Axial-DeepLab	44.2 (-7.1)
	MaX-DeepLab	**51.3**

Comparison on COCO test-dev set.

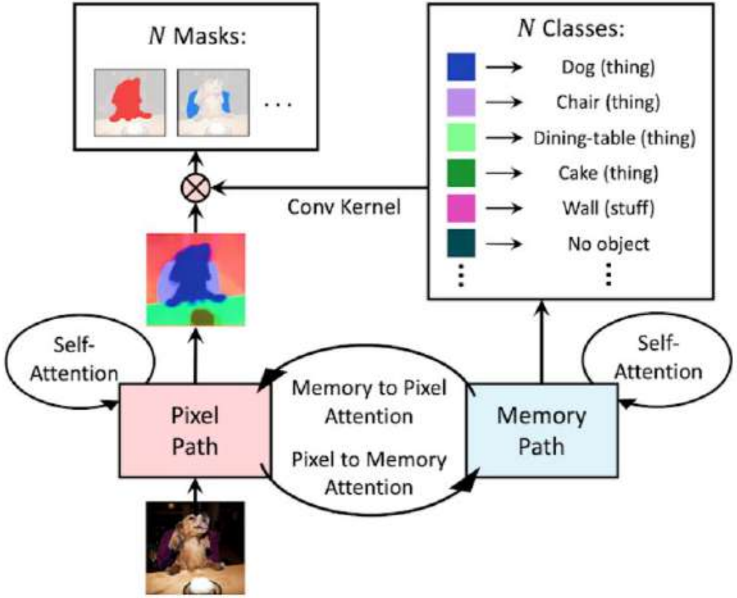


A case study for MaX-DeepLab and state-of-the-art box-free and box-based methods.

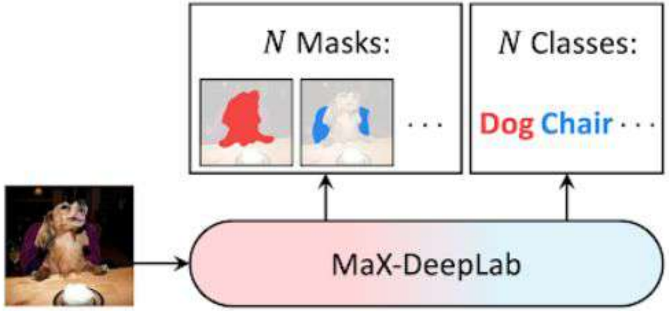
https://blog.research.google/2021/04/max-deeplab-dual-path-transformers-for.html

Google Research

MaX-DeepLab: Dual-Path Transformers for End-to-End Panoptic Segmentation



An overview of the dual-path transformer architecture



MaX-DeepLab directly predicts N masks and N classes with a CNN and a mask transformer.

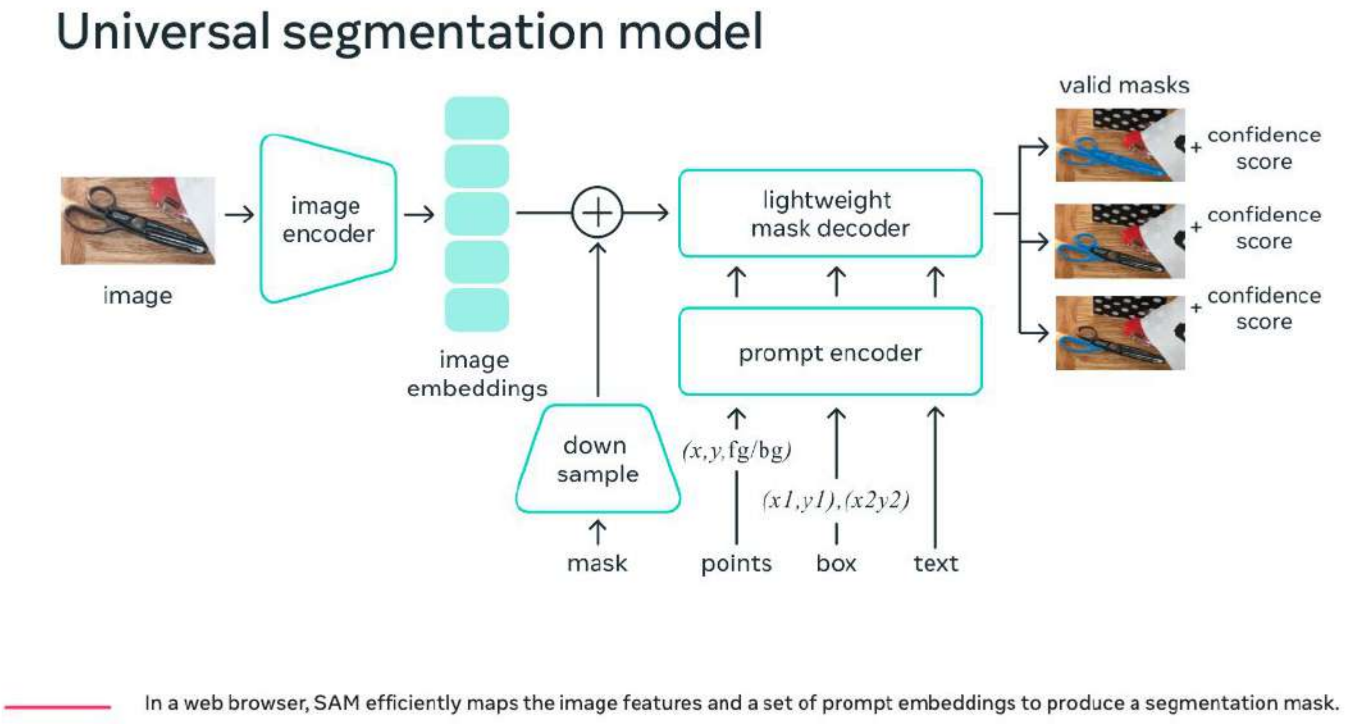
When the centers of the dog and the chair are close to each other, Axial-DeepLab merges them into one object.

Segment Anything Model (SAM)

- a new AI model from Meta AI that can "cut out" any object, in any image, with a single click

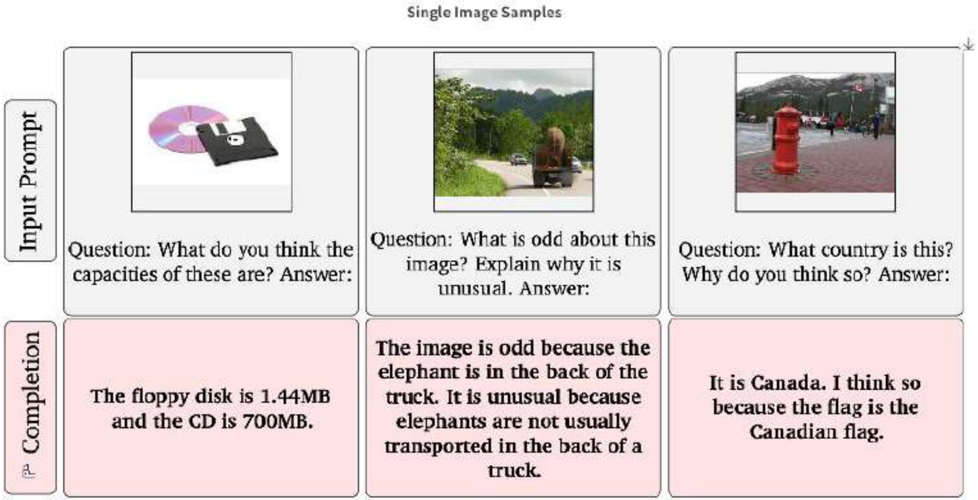
https://ai.meta.com/blog/segment-anything-foundation-model-image-segmentation

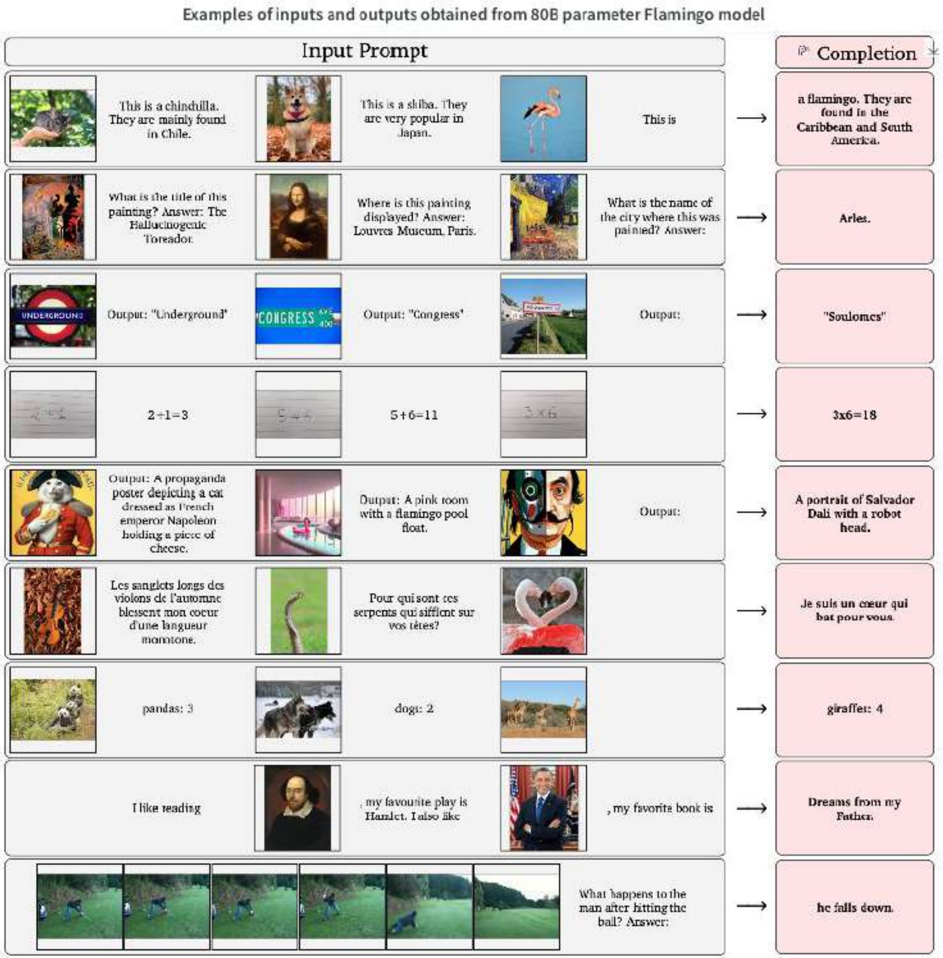
Segment Anything Model (SAM)

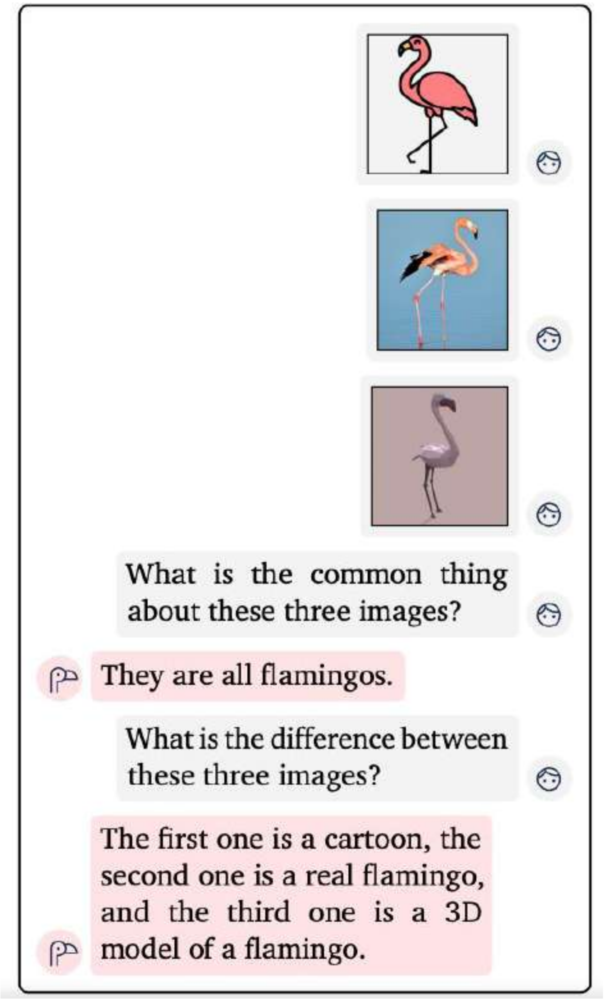


DeepMind Flamingo: A Visual Language Model for Few-Shot Learning

Flamingo is a new visual language model (VLM) capable of multimodal tasks like captioning, visual dialogue, classification, and visual question answering. As you can see, it works rather well:

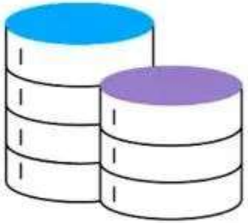



Examples of inputs and outputs obtained from 80B parameter Flamingo model
Input Prompt
Completion
This is a chinchilla. They are mainly found in Chile.
This is a shiba. They are very popular in Japan.
This is
a flamingo. They are found in the Caribbean and South America.
What is the title of this painting? Answer: The Hallucinogenic Toreador.
Where is this painting displayed? Answer: Louvres Museum, Paris.
What is the name of the city where this was painted? Answer:
Arles.
Output: "Underground"
Output: "Congress"
Output:
"Soulomes"
2+1=3
5+6=11
3x6=18
Output: A propaganda poster depicting a cat dressed as French emperor Napoleon holding a piece of cheese.
Output: A pink room with a flamingo pool float.
Output:
A portrait of Salvador Dali with a robot head.
Les sanglots longs des violons de l'automne blessent mon coeur d'une langueur monotone.
Pour qui sont ces serpents qui sifflent sur vos têtes?
Je suis un cœur qui bat pour vous.
pandas: 3
dogs: 2
giraffes: 4
I like reading
, my favourite play is Hamlet. I also like
, my favorite book is
Dreams from my Father.
What happens to the man after hitting the ball? Answer:
he falls down.

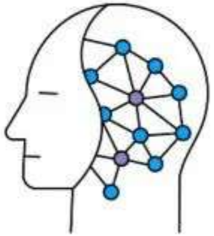

What is the common thing about these three images?
They are all flamingos.
What is the difference between these three images?
The first one is a cartoon, the second one is a real flamingo, and the third one is a 3D model of a flamingo.

Introduction to Generative AI

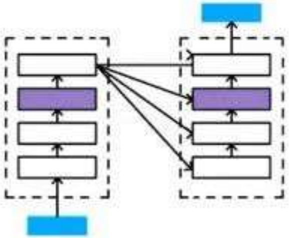
Large Language Models (LLM)



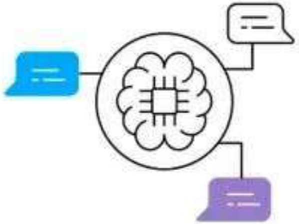
Massive Dataset



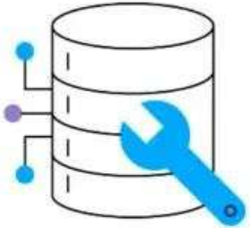
Deep Learning



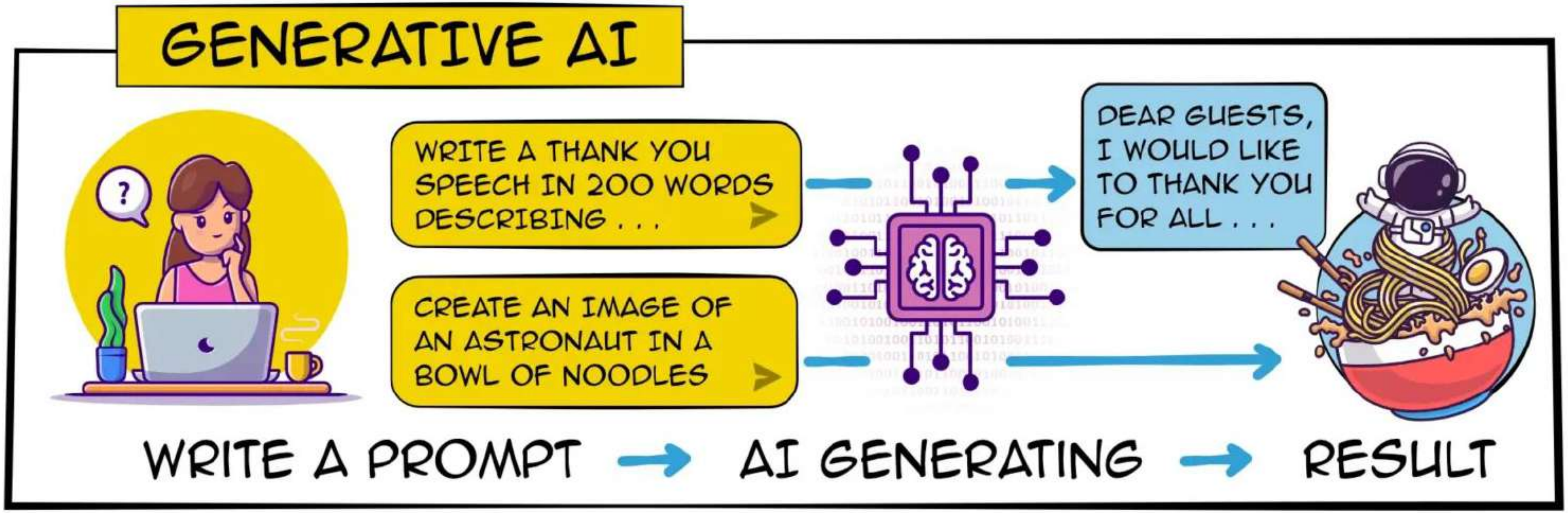
Transformer Architecture

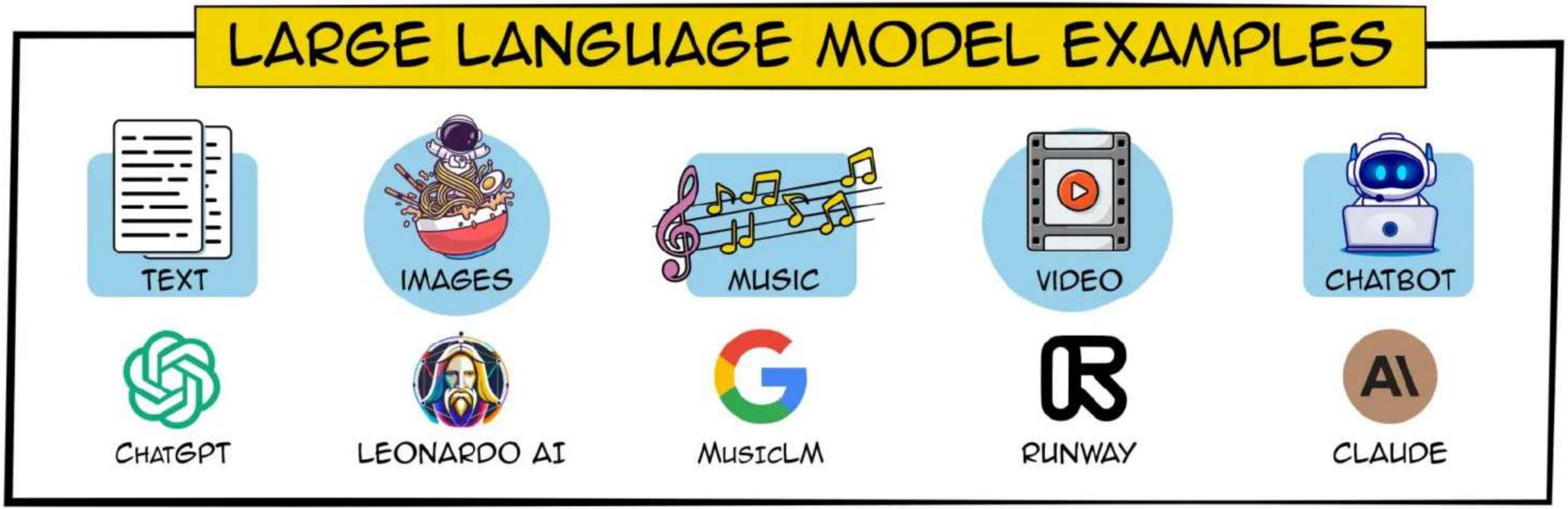


Self-supervised Learning

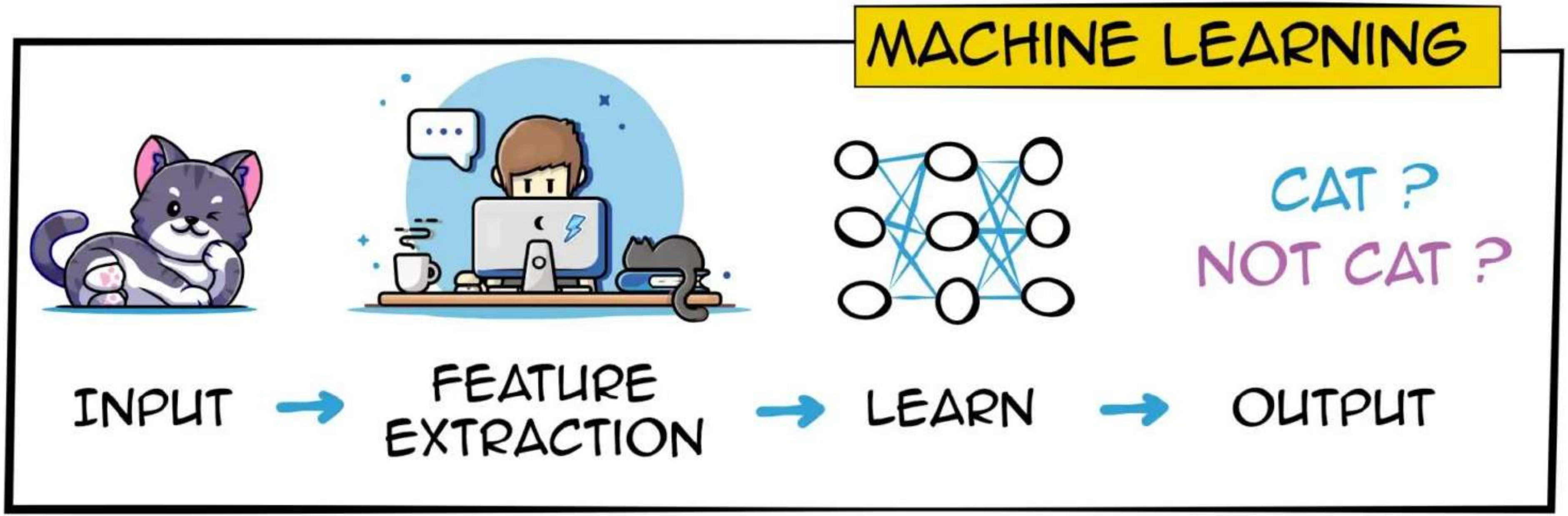


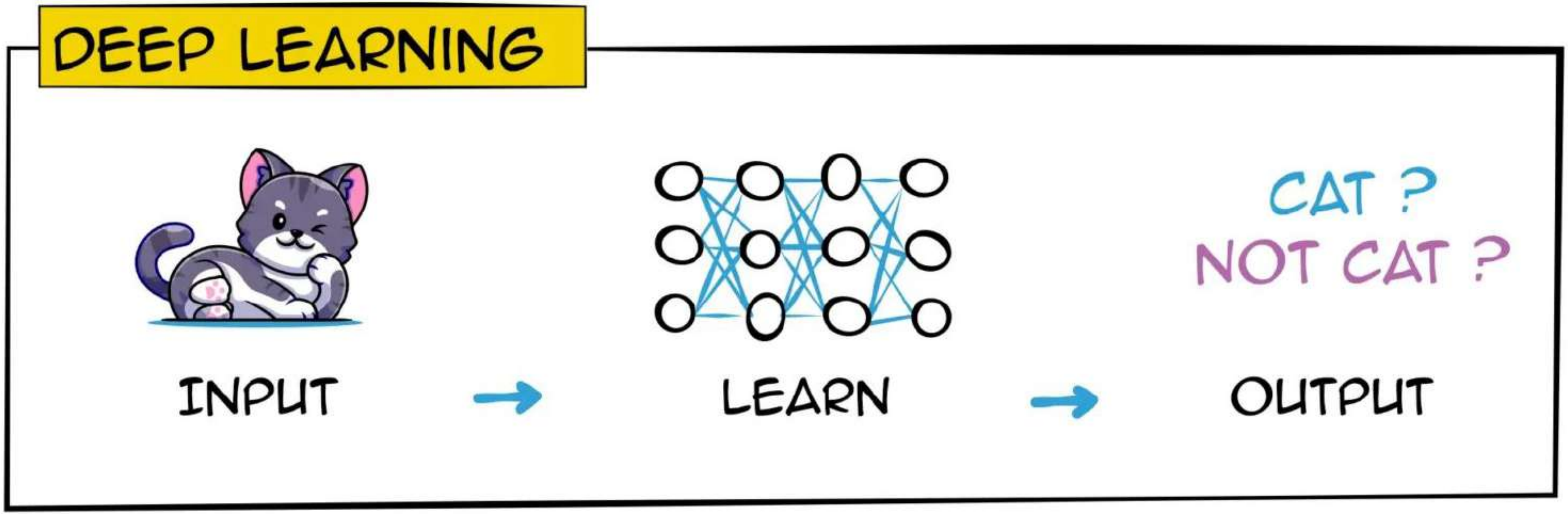
Fine-tuning

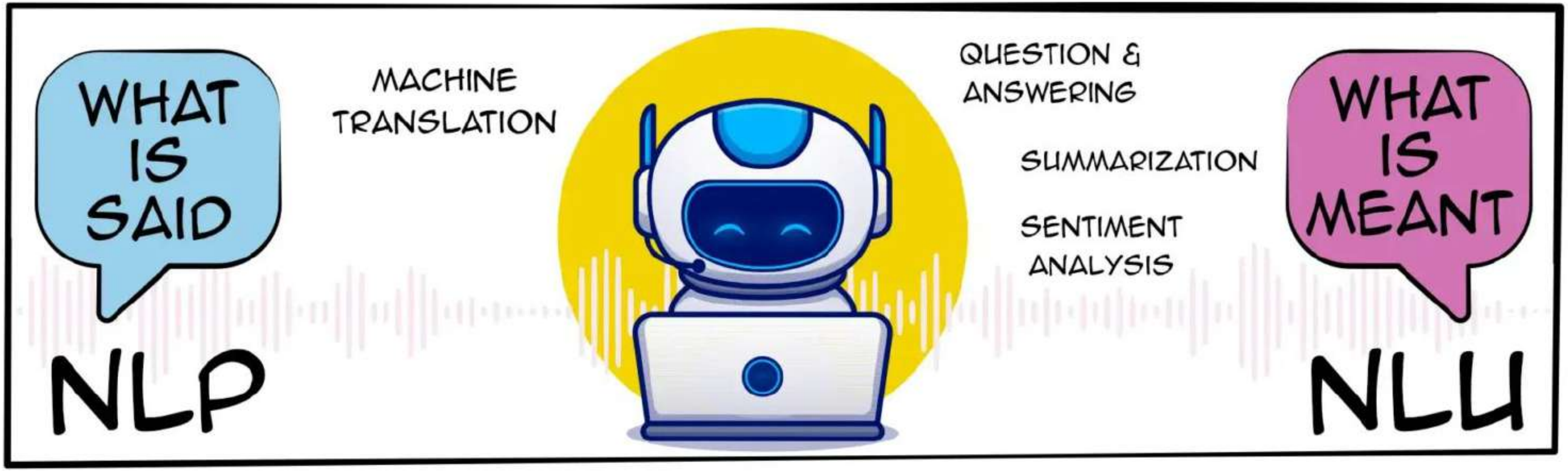

GENERATIVE AI
WRITE A THANK YOU SPEECH IN 200 WORDS DESCRIBING . . .
CREATE AN IMAGE OF AN ASTRONAUT IN A BOWL OF NOODLES
DEAR GUESTS, I WOULD LIKE TO THANK YOU FOR ALL . . .
WRITE A PROMPT
AI GENERATING
RESULT



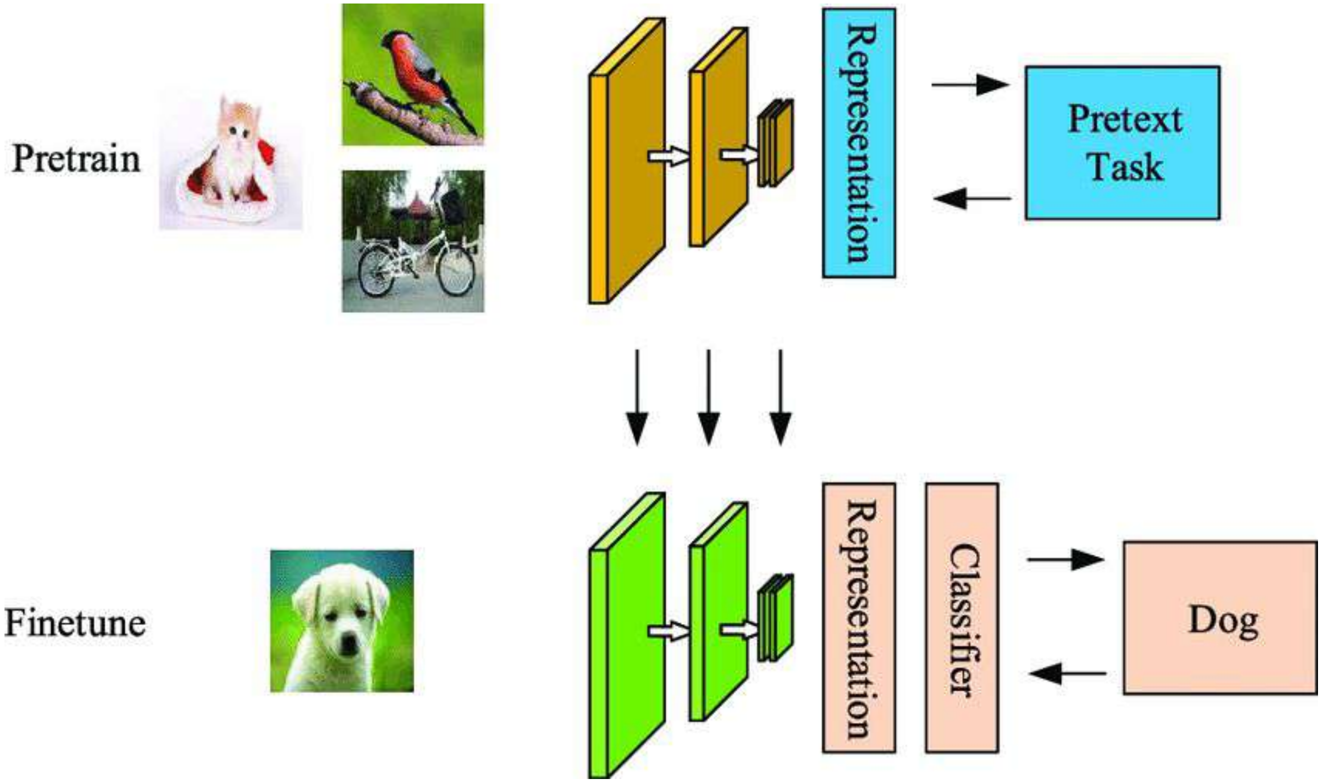
LARGE LANGUAGE MODEL EXAMPLES
TEXT
IMAGES
MUSIC
VIDEO
CHATBOT
ChatGPT
LEONARDO AI
MusicLM
RUNWAY
CLAUDE

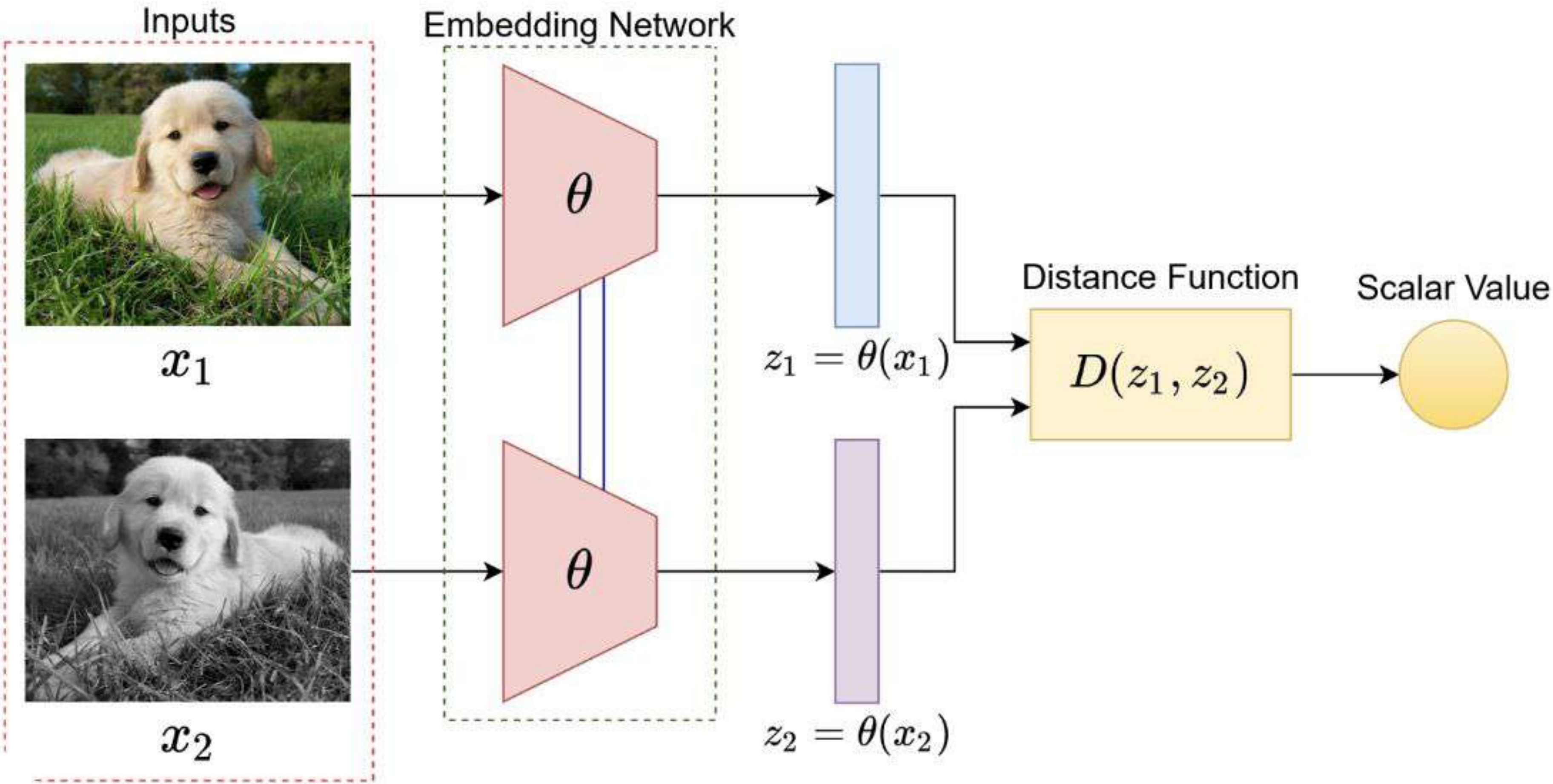

MACHINE LEARNING
CAT ?
NOT CAT ?
INPUT
FEATURE EXTRACTION
LEARN
OUTPUT


DEEP LEARNING
CAT ?
NOT CAT ?
INPUT
LEARN
OUTPUT

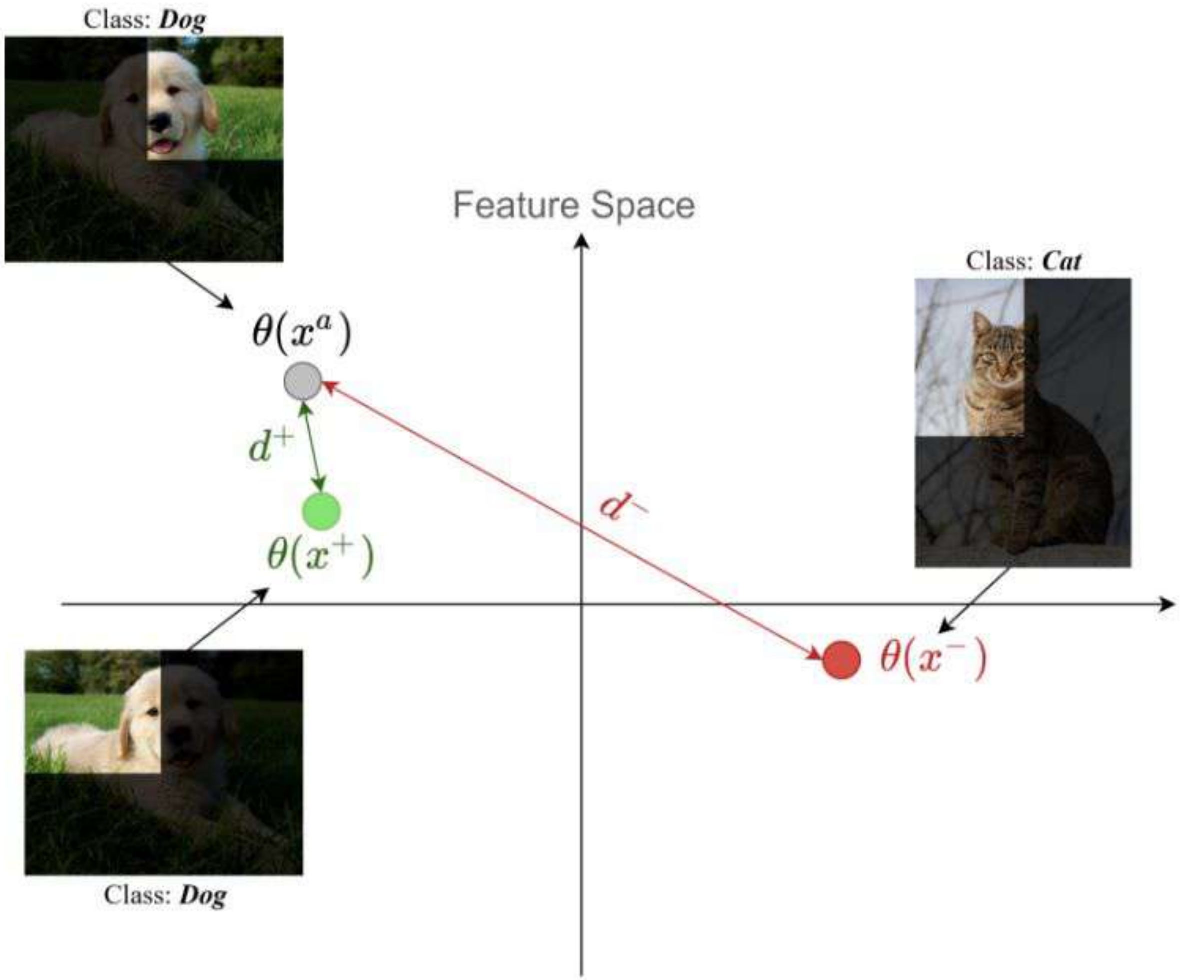

WHAT IS SAID
NLP
MACHINE TRANSLATION
QUESTION & ANSWERING
SUMMARIZATION
SENTIMENT ANALYSIS
WHAT IS MEANT
NLU

Self-Supervised Learning

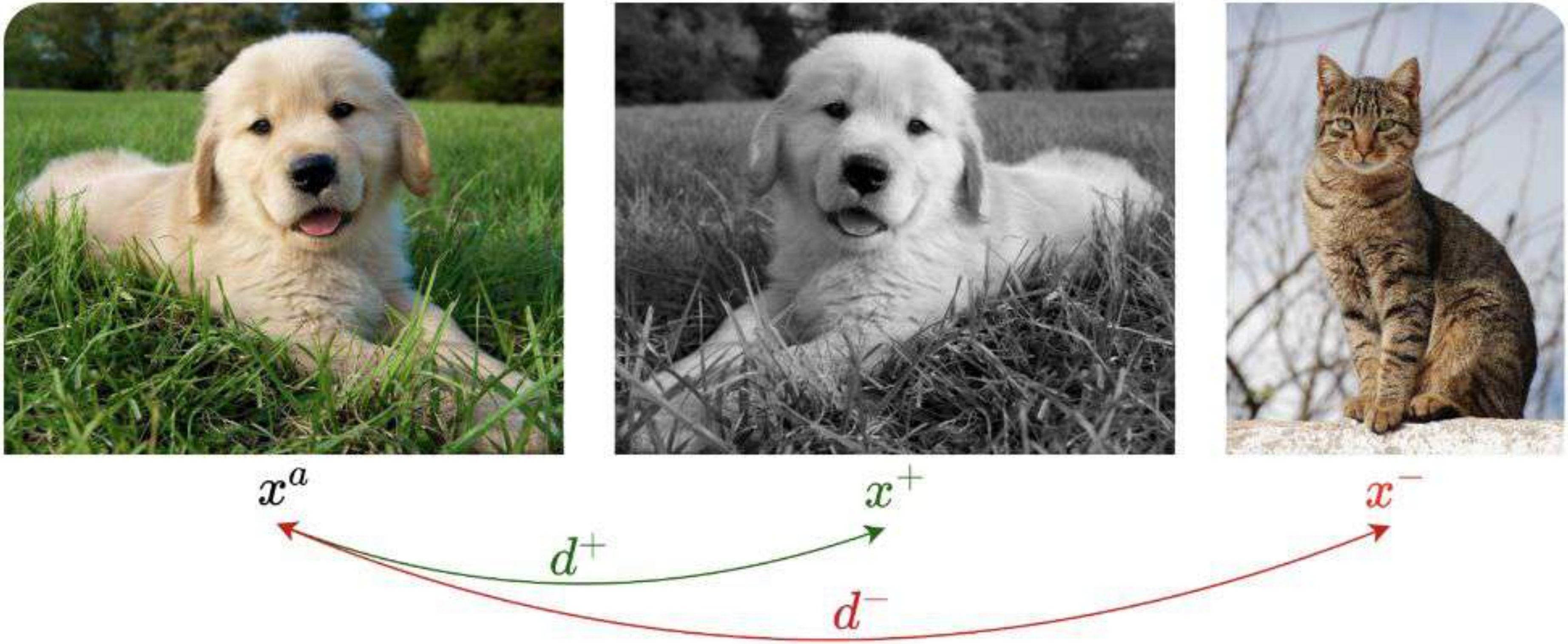




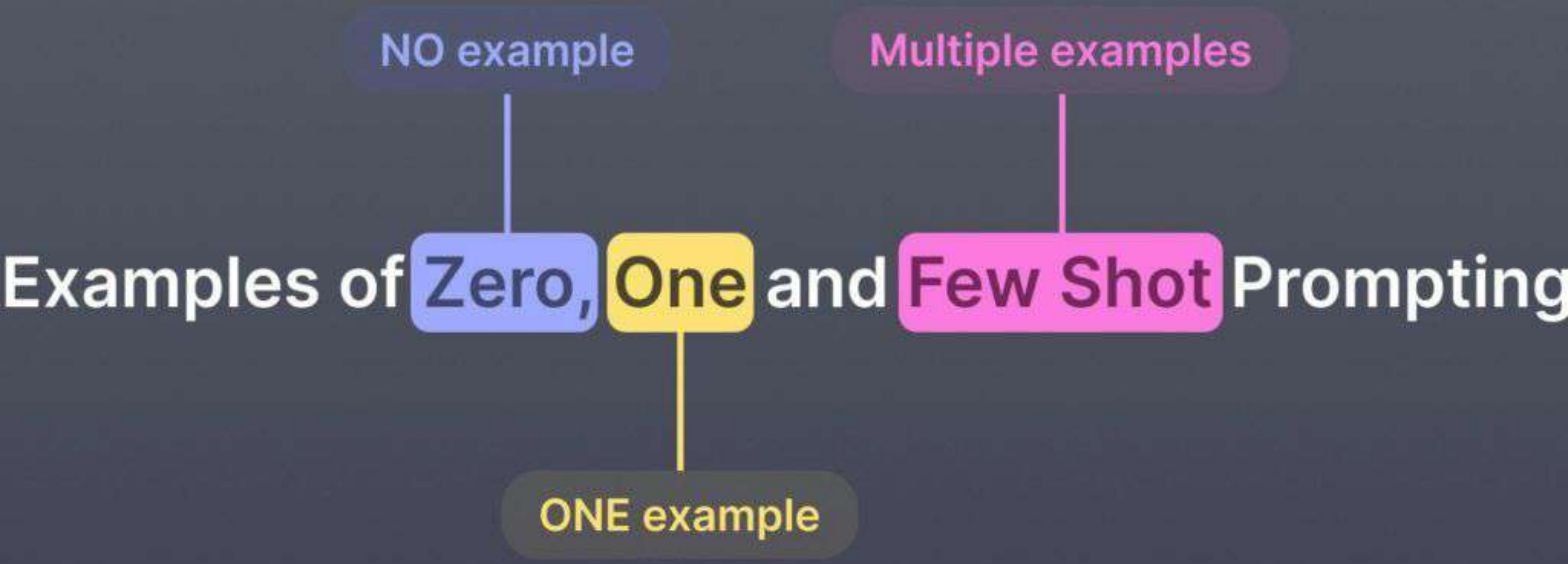
Joint Embedding Architecture. Image by the author



Contrastive Learning. Image by author.



Instance Discrimination Methods.


NO example
Multiple examples
Examples of Zero, One and Few Shot Prompting
ONE example

Zero-shot Prompting!

Model Input

What is the sentiment of the following text? Choose from 'Positive' or 'Negative'

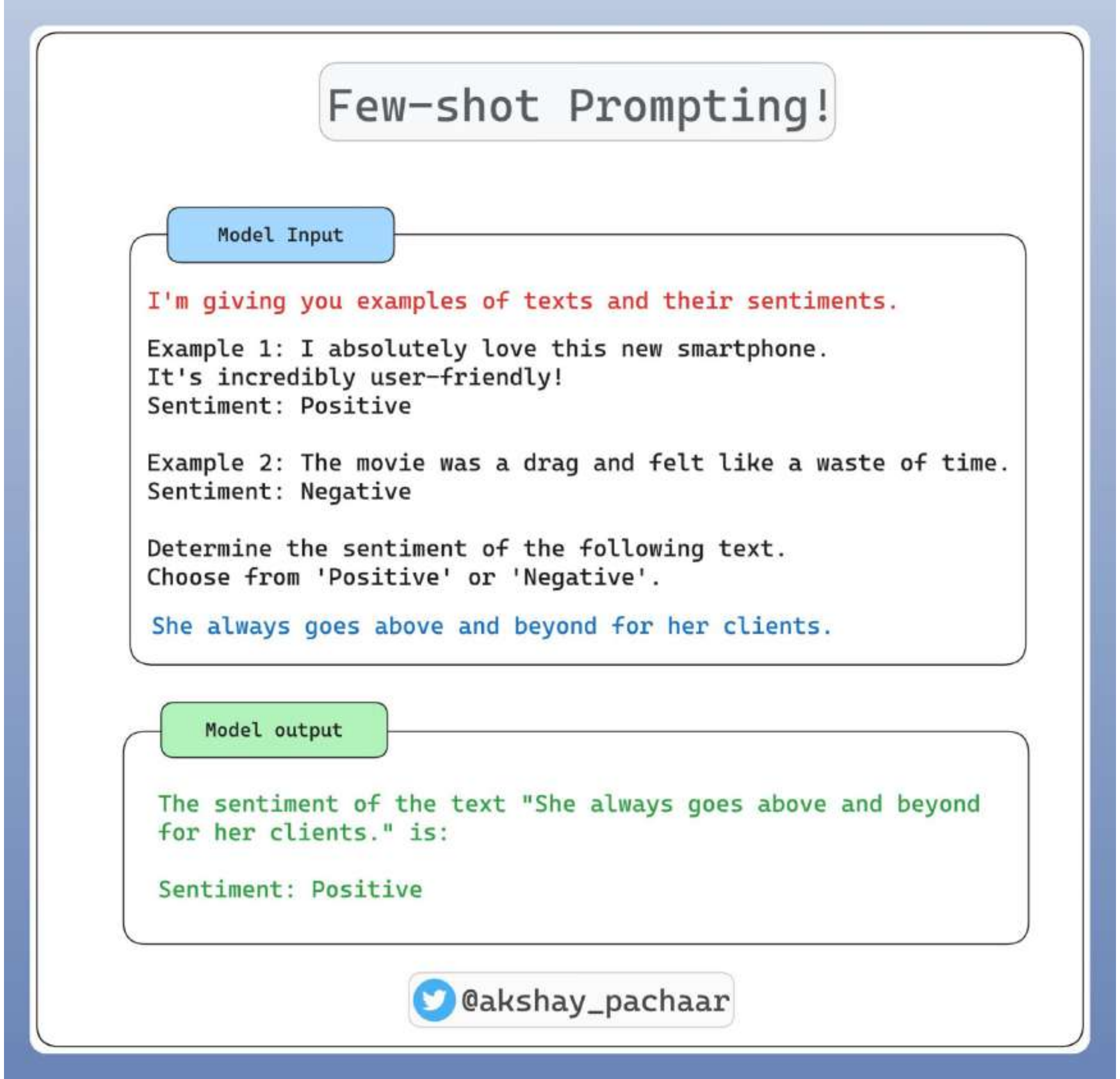
Text: The team's performance last night was top-notch.

Model output

The sentiment of the text "The team's performance last night was top-notch." is:

Sentiment: Positive

@akshay_pachaar


Few-shot Prompting!
Model Input
I'm giving you examples of texts and their sentiments.
Example 1: I absolutely love this new smartphone.
It's incredibly user-friendly!
Sentiment: Positive
Example 2: The movie was a drag and felt like a waste of time.
Sentiment: Negative
Determine the sentiment of the following text.
Choose from 'Positive' or 'Negative'.
She always goes above and beyond for her clients.
Model output
The sentiment of the text "She always goes above and beyond
for her clients." is:
Sentiment: Positive
@akshay_pachaar

Few-shot Prompting!

Model Input

```
I'm giving you examples of texts and their sentiments.

Example 1: I absolutely love this new smartphone.
It's incredibly user-friendly!
Sentiment: Positive

Example 2: The movie was a drag and felt like a waste of time.
Sentiment: Negative

Determine the sentiment of the following text.
Choose from 'Positive' or 'Negative'.

She always goes above and beyond for her clients.
```

Model output

```
The sentiment of the text "She always goes above and beyond
for her clients." is:

Sentiment: Positive
```

@akshay_pachaar

Chain-of-Thought Prompting!

Standard Prompting!

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model output

A: The answer is 27. ❌

Chain-of-Thought Prompting!

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. 5 + 6 = 11. The answer is 11.

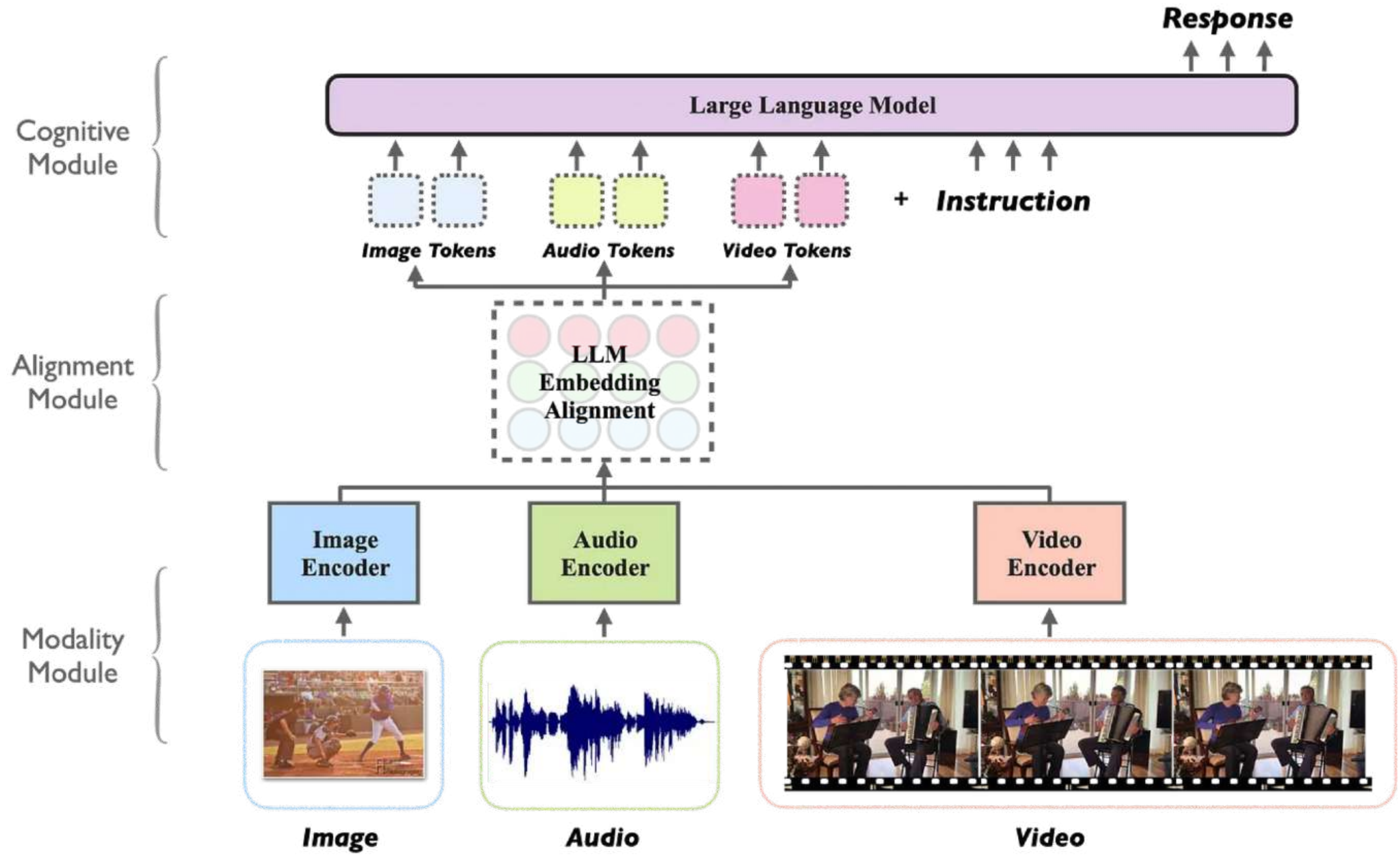
Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model output

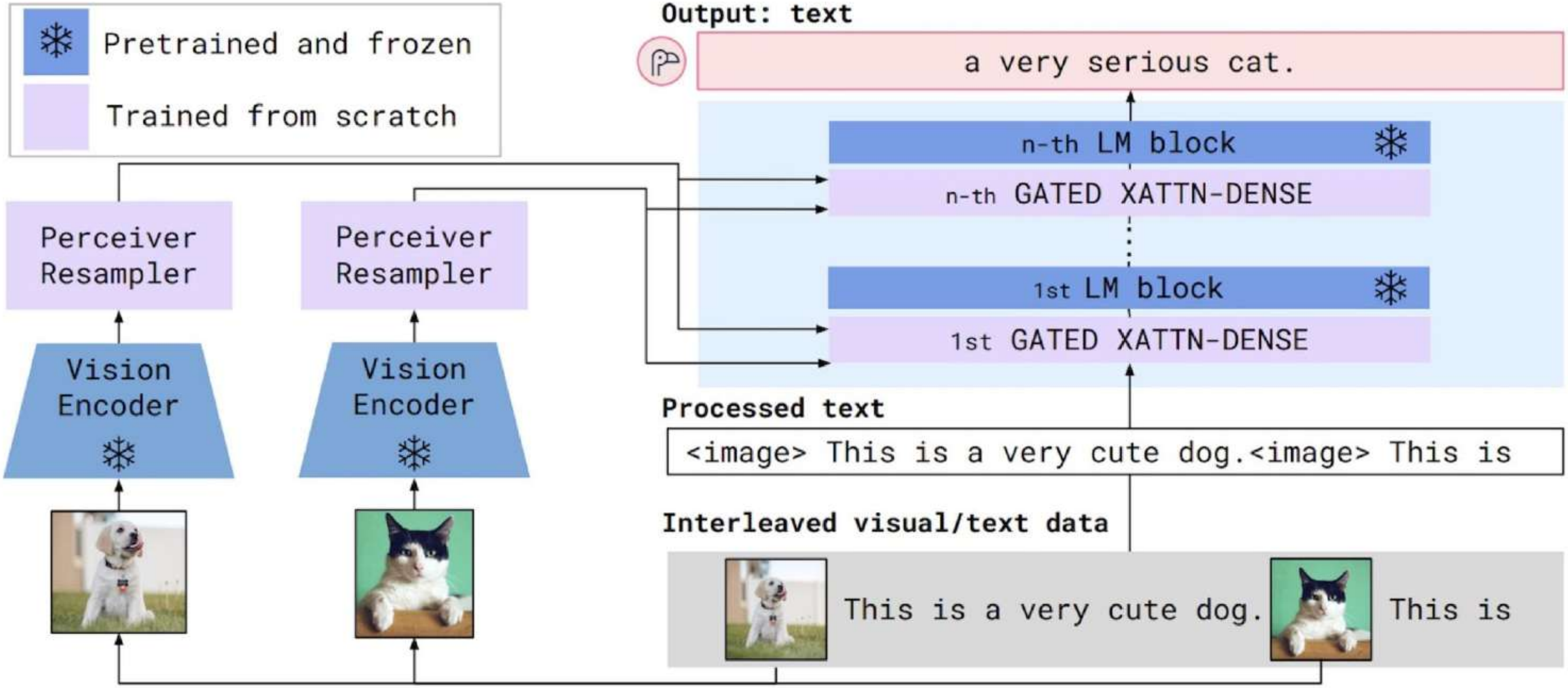
A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had 23 - 20 = 3. They bought 6 more apples, so they have 3 + 6 = 9. The answer is 9. ✅

@akshay_pachaar

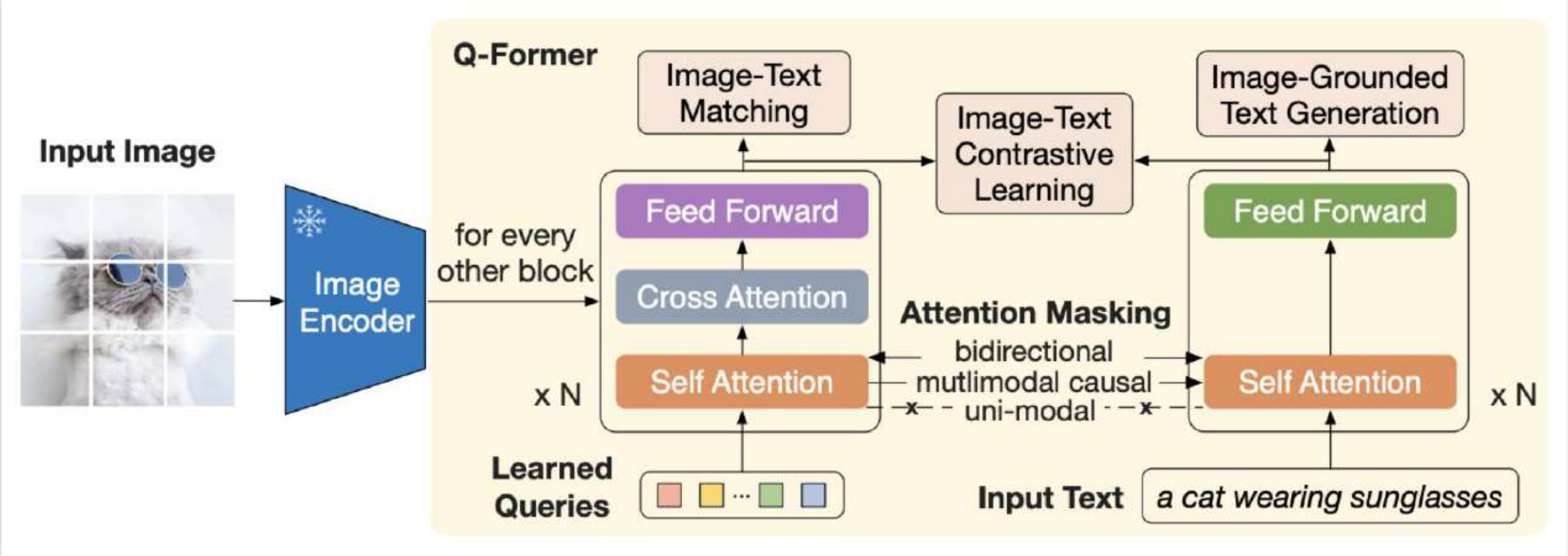
source: https://arxiv.org/pdf/2201.11903.pdf



Response
Large Language Model
Cognitive Module
+ Instruction
Image Tokens
Audio Tokens
Video Tokens
Alignment Module
LLM Embedding Alignment
Image Encoder
Audio Encoder
Video Encoder
Modality Module
Image
Audio
Video



Gheini, Mozhdeh, Xiang Ren, and Jonathan May. "Cross-attention is all you need: Adapting pretrained transformers for machine translation." *arXiv preprint arXiv:2104.08771* (2021).


Q-Former
Input Image
Image Encoder
for every other block
Image-Text Matching
Image-Text Contrastive Learning
Image-Grounded Text Generation
Feed Forward
Cross Attention
Self Attention
Attention Masking
bidirectional
mutlimodal causal
uni-modal
x N
Feed Forward
Self Attention
x N
Learned Queries
Input Text
a cat wearing sunglasses

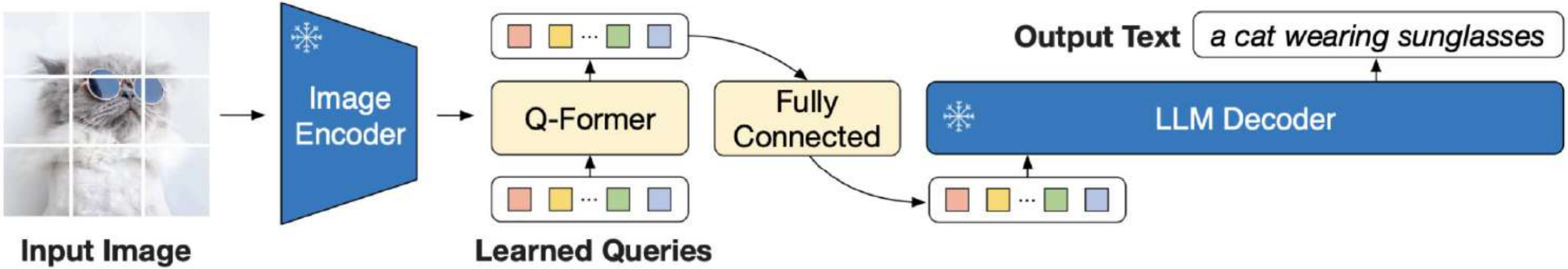
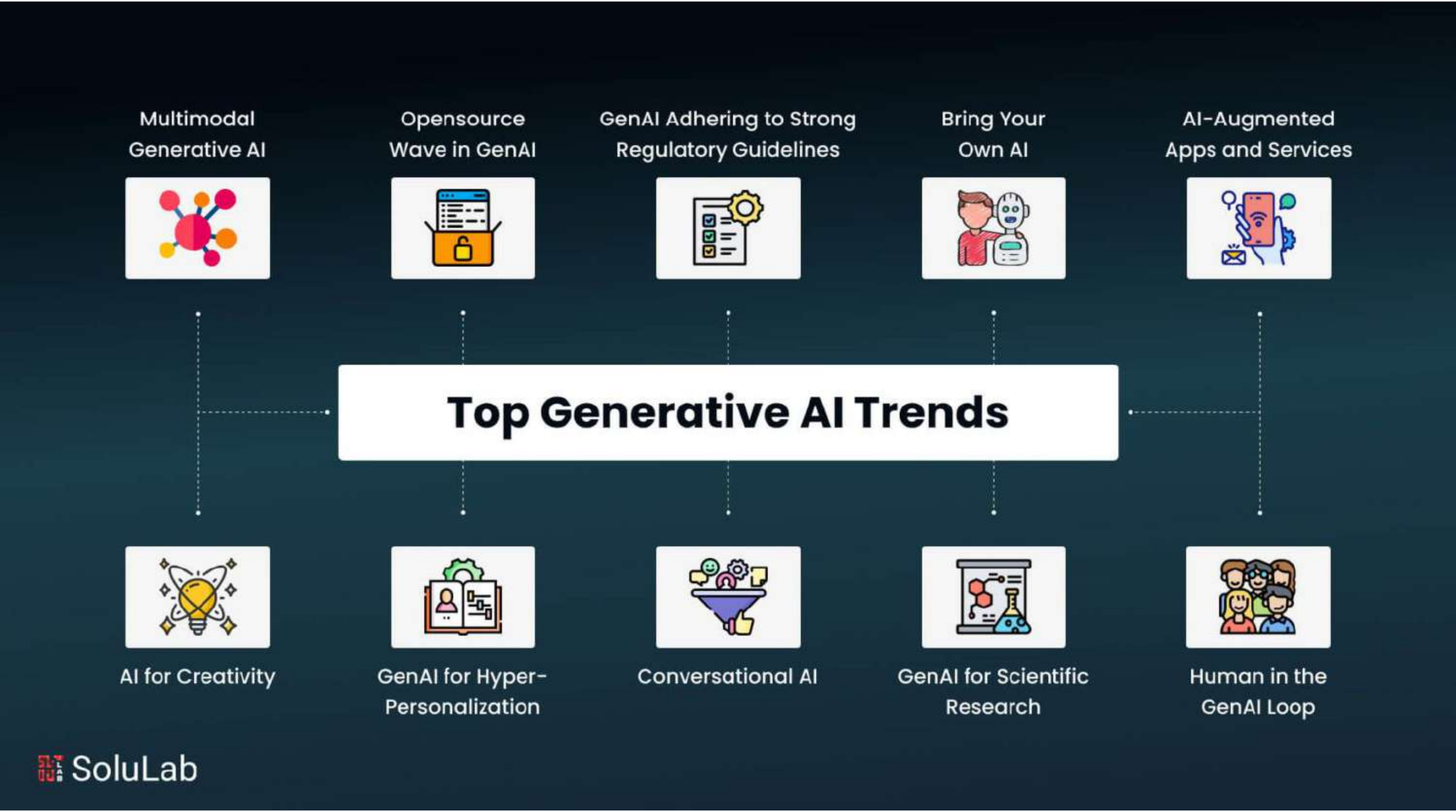
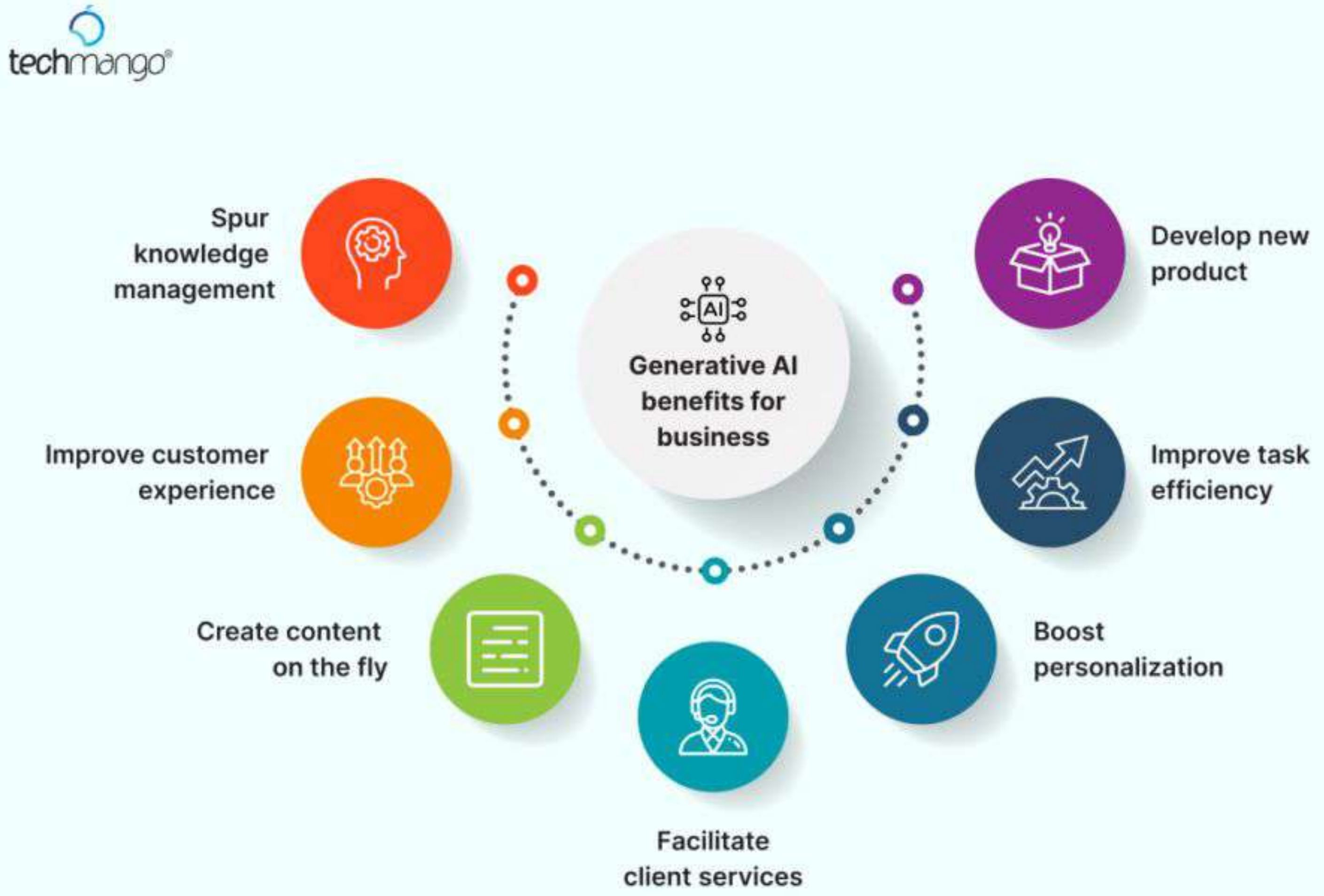


Image Encoder
Q-Former
Fully Connected
Output Text
a cat wearing sunglasses
LLM Decoder
Input Image
Learned Queries

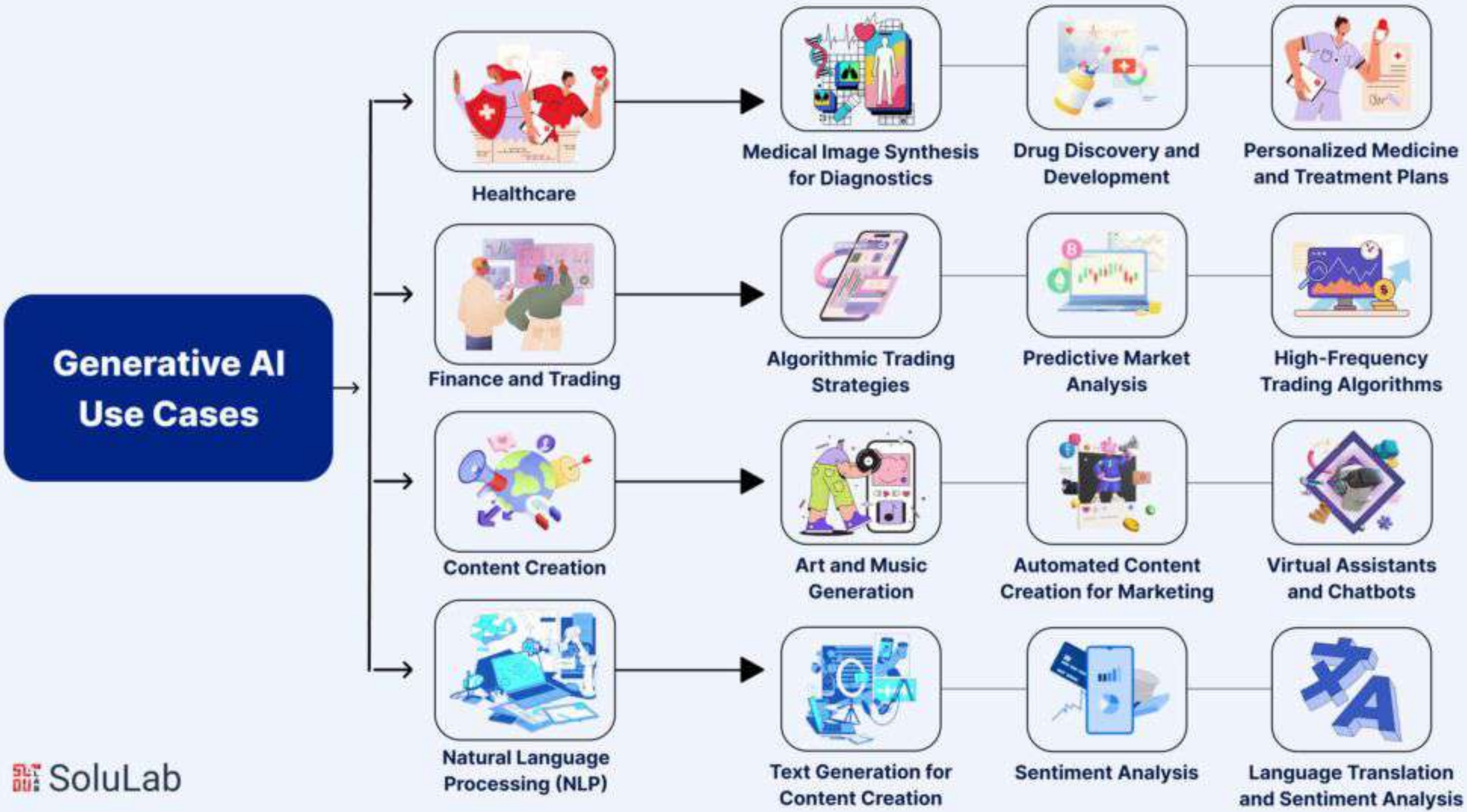
Key Trends in Generative AI


Top Generative AI Trends
Multimodal Generative AI
Opensource Wave in GenAI
GenAI Adhering to Strong Regulatory Guidelines
Bring Your Own AI
AI-Augmented Apps and Services
AI for Creativity
GenAI for Hyper-Personalization
Conversational AI
GenAI for Scientific Research
Human in the GenAI Loop
SoluLab



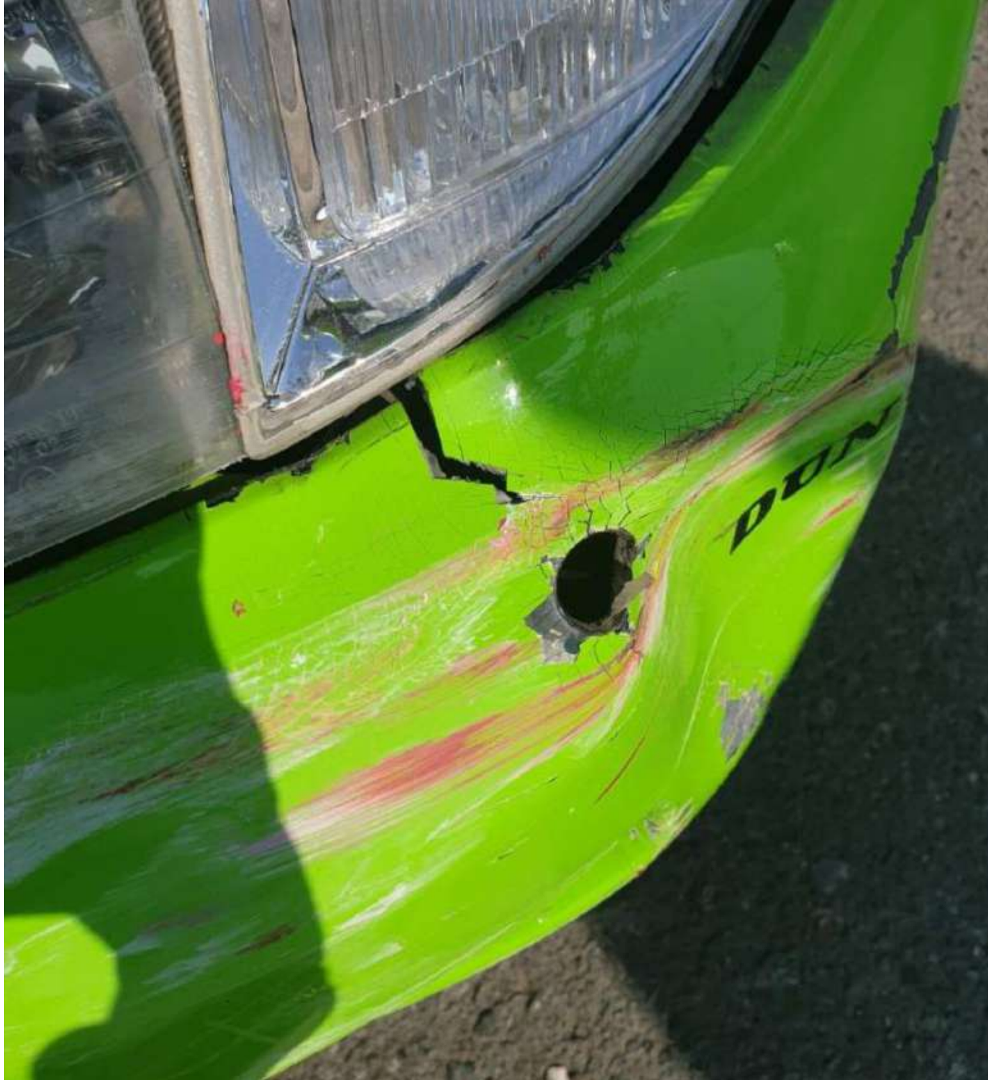
techmango
Generative AI benefits for business
Spur knowledge management
Improve customer experience
Create content on the fly
Facilitate client services
Boost personalization
Improve task efficiency
Develop new product

Applications of Generative AI



Generative AI Use Cases
Healthcare
Medical Image Synthesis for Diagnostics
Drug Discovery and Development
Personalized Medicine and Treatment Plans
Finance and Trading
Algorithmic Trading Strategies
Predictive Market Analysis
High-Frequency Trading Algorithms
Content Creation
Art and Music Generation
Automated Content Creation for Marketing
Virtual Assistants and Chatbots
Natural Language Processing (NLP)
Text Generation for Content Creation
Sentiment Analysis
Language Translation and Sentiment Analysis
SoluLab

Case1:



ChatGPT

ChatGPT

Did you see any damage to the car in this image? If so, which part of the car is damaged?

Yes, there is visible damage to the car in this image. The damage is located on the front bumper near the headlight. The bumper has a crack, a hole, and some scratches with paint transfer.

ChatGPT

ChatGPT

Can you guess which side of the car this image shows and identify the car parts visible

Based on the image, the visible parts are:

1. **Headlight** - Likely the front corner of the car.
2. **Front bumper** - The damaged area with a crack and hole.
3. **Part of the grille or adjacent body panel** - Near the headlight and bumper.

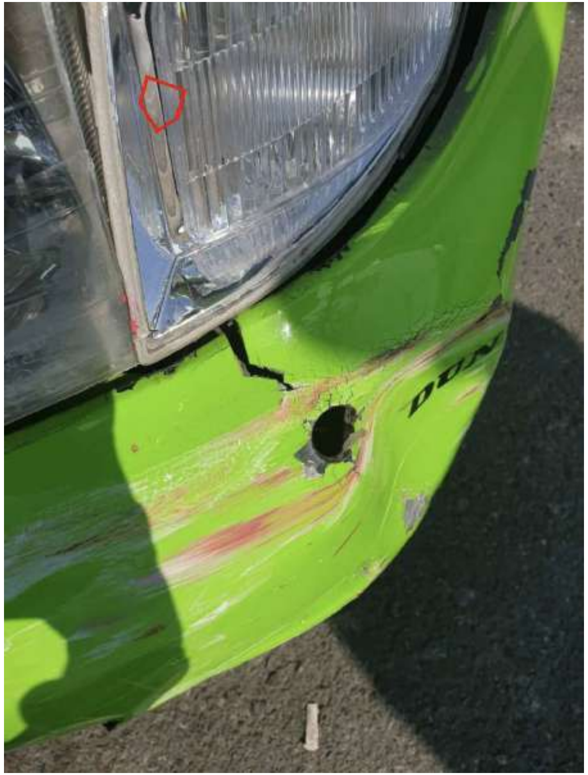
Given the orientation of the headlight and the visible parts, it appears to be the **front left side** of the car.

Please mark the damaged area with a polygon or draw a segmentation outline.

I have marked the damaged area with a red polygon. You can view the image here. [>-]

ChatGPT

ChatGPT



Did you see any damage to the car in this image? If so, which part of the car is damaged?

Yes, there is visible damage to the car in this image. The damage is located on the front bumper near the headlight. The bumper has a crack, a hole, and some scratches with paint transfer.

Claude   Assessing Car Damage in Image

K Did you see any damage to the car in this image? If so, which part of the car is damaged?

Yes, there is visible damage to the car in this image. The damage is located on the front bumper or fender area, which is painted in a bright lime green color. There's a noticeable hole or puncture in the bodywork, surrounded by cracked and scraped paint. The damage appears to be just below and to the side of the headlight assembly. The bright green paint shows scratches, cracks, and chipping around the damaged area, suggesting a collision or impact of some kind occurred.

Claude   Assessing Car Damage in Image

K Can you guess which side of the car this image shows and identify the car parts visible

Based on the image, this appears to be the front right corner of the vehicle. The visible parts include:

1. Headlight: The clear, ribbed surface of the headlight assembly is prominent in the upper part of the image.
2. Front bumper/fender: The bright lime green section is likely part of the front bumper or fender, which has sustained damage.
3. Hood edge: Though not clearly visible, the edge where the hood meets the fender/bumper area can be inferred from the image composition.

The positioning of these elements suggests this is the front right (passenger side in left-hand drive countries) corner of the vehicle. The angle and close-up nature of the shot make it difficult to see more of the car's overall structure, but these key components are clearly identifiable.

K Please mark the damaged area with a polygon or draw a segmentation outline.

I apologize, but I do not have the ability to edit, mark, or draw on images. I can only perceive and analyze the images that are provided to me. I cannot generate, produce, manipulate, edit or create new images in any way.

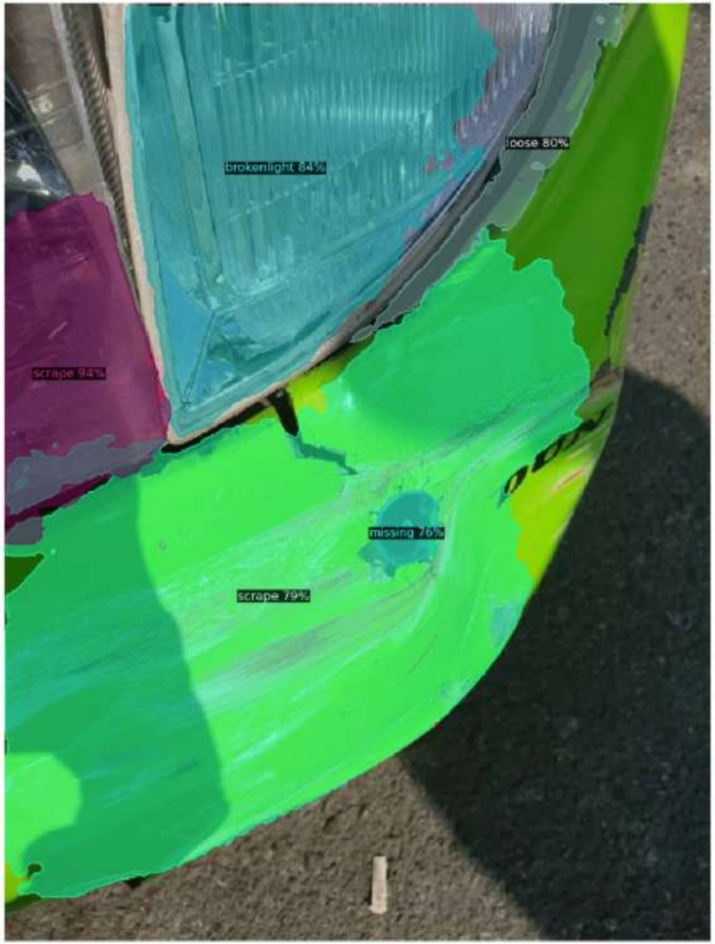
If you need the damaged area marked, you would need to use an image editing software or

Claude

MARS

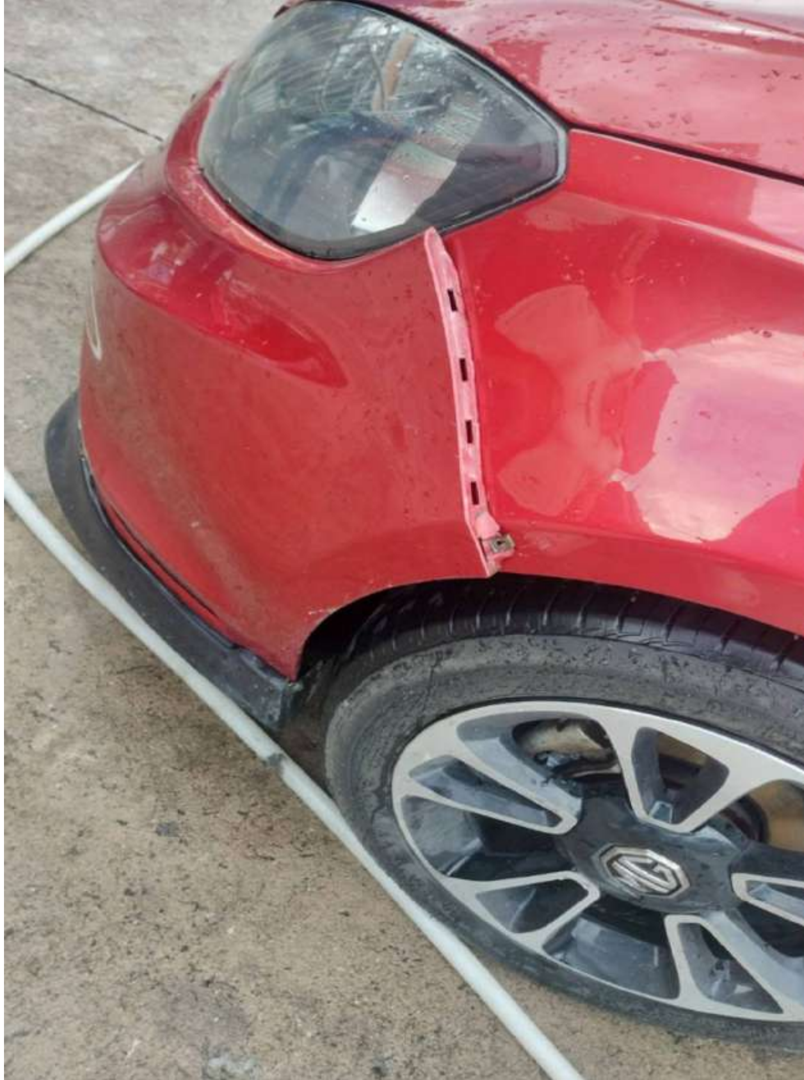
Motor AI Recognition Solution

AVENGER-PART-RESULT



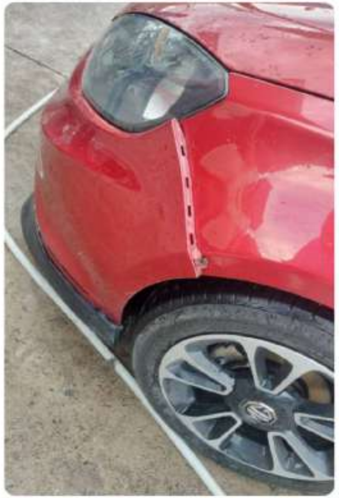
AVENGER-DAMAGE-RESULT

Case2:



ChatGPT

ChatGPT



Can you spot any damage to the car in this image? If yes, which part of the car is affected?

Yes, there is visible damage to the car in this image. The affected part is the front right side of the car, specifically:

1. **Front fender** - There is a significant gap and misalignment where the fender meets the bumper.
2. **Bumper** - The bumper seems to be slightly detached from the fender, and there are visible scratches and misalignment.
3. **Headlight area** - The area around the headlight also appears to be affected due to the misalignment and gap between the fender and bumper.

ChatGPT

ChatGPT

Can you determine which side of the car is shown in the image and identify the visible car parts?

The image shows the front left side of the car. The visible car parts are:

1. **Front left headlight** - Clearly visible near the top left of the image.
2. **Front left fender** - The red body panel surrounding the wheel and extending towards the headlight.
3. **Front bumper** - The red panel below the headlight and wrapping around the front of the car.
4. **Front left wheel and tire** - Visible at the bottom right of the image.

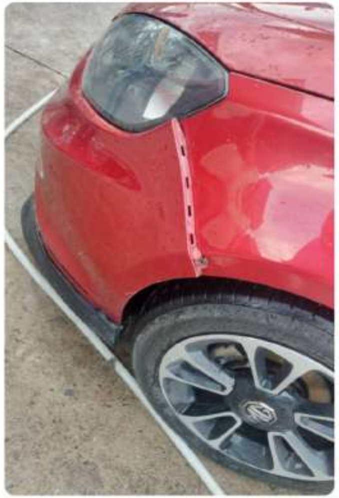
The damage is primarily located on the front left fender, which is misaligned and detached from the front bumper.

Please highlight the damaged area by drawing a polygon or outlining the segmentation.

I have marked the damaged area with a red polygon. You can view the image here. [>-]

ChatGPT

ChatGPT



Can you spot any damage to the car in this image? If yes, which part of the car is affected?

Yes, there is visible damage to the car in this image. The affected part is the front right side of the car, specifically:

1. **Front fender** - There is a significant gap and misalignment where the fender meets the bumper.
2. **Bumper** - The bumper seems to be slightly detached from the fender, and there are visible scratches and misalignment.
3. **Headlight area** - The area around the headlight also appears to be affected due to the misalignment and gap between the fender and bumper.

Claude Assessing Car Damage in Image

K Can you spot any damage to the car in this image? If yes, which part of the car is affected?

Yes, there is visible damage to the car in this image. The damage is located on the front fender, just above the wheel. You can see a vertical crack or separation in the bodywork, running from near the headlight down towards the wheel well. This crack appears to be where two panels of the car body meet, and it has separated more than it should, indicating damage to the fender area.

Additionally, there seems to be some slight misalignment or unevenness in how the panels fit together in this area, which further suggests damage or improper fitting of the body panels.

The rest of the visible parts of the car (headlight, hood, wheel) appear to be intact from this angle, but the fender damage is quite noticeable.

Claude Assessing Car Damage in Image

K Can you determine which side of the car is shown in the image and identify the visible car parts?

This image shows the front left side of the car. The visible car parts are:

1. Headlight: The angular headlight is visible at the top of the image.
2. Hood: A portion of the red hood can be seen extending from the headlight area.
3. Front left fender: This is the main body panel visible, where the damage is located.
4. Wheel/tire: The front left wheel is visible at the bottom of the image.
5. Wheel rim: A detailed view of the alloy wheel design is shown.
6. Bumper: A small part of the front bumper is visible below the headlight.
7. Panel gap: The vertical line with separation is visible, which is typically where the fender meets the front door.
8. Side marker light or reflector: A thin vertical strip is visible on the fender, which is likely a side marker light or reflector.

The image captures the front corner of the car, focusing on the area where the fender, wheel, and headlight come together. The left side is evident from the positioning of these elements and the angle of the shot.

Copy Retry

Claude can make mistakes. Please double-check responses.

K

You are out of free messages until 6 PM

Subscribe to Pro

Claude

MARS
Motor AI Recognition Solution

AVENGER-PART-RESULT

AVENGER-DAMAGE-RESULT


TYPHOON
โมเดลภาษาขนาดใหญ่
เพื่อภาษาไทย โดยเฉพาะ
TYPHOON
Docs
Join the Waitlist
Introducing Typhoon API
Large Language Model optimized for Thai language
Join the Waitlist
Learn more about Typhoon
Typhoon 7B pretrained model is available open source
Download pretrained weights
Learn more →

https://arxiv.org/pdf/2312.13951

Abstract

Typhoon is a series of Thai large language models (LLMs) developed specifically for the Thai language. This technical report presents challenges and insights in developing Thai LLMs, including data preparation, pretraining, instruction-tuning, and evaluation. As one of the challenges of low-resource languages is the amount of pretraining data, we apply continual training to transfer existing world knowledge from a strong LLM. To evaluate the Thai knowledge encapsulated in each model from the pretraining stage, we develop ThaiExam, a benchmark based on examinations for high-school students and investment professionals in Thailand. In addition, we fine-tune Typhoon to follow Thai instructions, and we evaluate instruction-tuned models on Thai instruction datasets as well as translation, summarization, and question-answering tasks. Experimental results on a suite of Thai benchmarks show that Typhoon outperforms all open-source Thai language models, and its performance is on par with GPT-3.5 in Thai while having only 7 billion parameters and being 2.62 times more efficient in tokenizing Thai text.

Model Weights: https://huggingface.co/scb10x/typhoon-7b

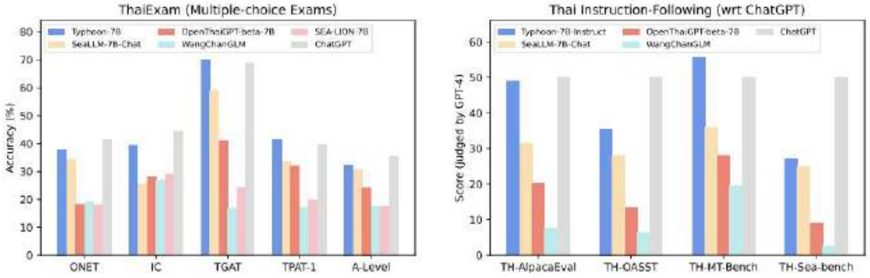


Figure 1: Performance of Typhoon and other open-source Thai large language models on Thai Examinations (left) and Thai instruction-following (right). Details about ThaiExam and Thai instruction-following evaluation are provided in Section 3.4, and Section 4.2, respectively.

Hugging Face
https://huggingface.co › docs › model_doc · แปลหน้านี้ ⋮

Mistral

As the Mistral model has 7 billion parameters, that would require about 14GB of GPU RAM in half precision (float16), since each parameter is stored in 2 bytes.

Model	MajorLang	#Params	VocabSize	Context	BaseModel
Typhoon-7B	English, Thai	7B	35,219	4096	Mistral
OpenThaiGPT-beta-7B	English, Thai	7B	56,554	4096	Llama2
WangChanGLM	Multilingual	7.5B	256,008	2048	XGLM
SeaLLM-7B	Multilingual	7B	48,512	4096	Llama2
SEA-LION-7B	Multilingual	7B	256,000	2048	✗

Table 1: Comparison between Typhoon and open-source language models that support Thai. OpenThaiGPT is based on Llama2 with an extended vocabulary, continual pretraining, and instruction fine-tuning. WangChanGLM is a multi-lingual XGLM model fine-tuned to follow Thai instructions without further pretraining. SEA-LION and SeaLLM are multilingual models specialized in Southeast Asian languages, including Thai. At the time of writing, SEA-LION is only instruction-tuned on Indonesian data, and SeaLLM is only available at 7B size.

- **ONET**: The Ordinary National Educational Test (ONET) is an examination for students in Thailand. We select the grade-12 ONET exam, which comprises 5 subjects and each question has 5 choices. These subjects are Thai, English, Mathematics, Social Studies, and Science. We extracted these questions from the official 2021 ONET example, amounting to a total of 170 questions and options.
- **IC**: The Investment Consultant (IC) examination, a licensing test for investment professionals in Thailand developed by the Stock Exchange of Thailand (SET), features 4 choices per question. We extracted questions for levels 1, 2, and 3 from the official SET website, resulting in a total of 95 questions and options.
- **TGAT**: The Thai General Aptitude Test (TGAT), a national high school examination in Thailand, focuses on critical and logical thinking skills, as well as proficiency in the English language. We collected a total of 90 questions and answers. The TGAT consists of four choices per question.
- **TPAT-1**: The Thai Professional Aptitude Test 1 (TPAT-1) is a national high school examination in Thailand that assesses students' professional skills requirement in medical schools. This subset contains reasoning and medical ethics. We collected a total of 116 questions and answers. The TPAT-1 consists of 5 choices per question.
- **A-Level**: An academic knowledge assessment examination (Applied Knowledge Level) that covers general foundational subjects taught in schools. The content assessed in this examination aligns with the curriculum guidelines and emphasizes the practical application of knowledge in daily life. We collected a total of 175 questions and answers.

5 Risk and Limitations

Similar to other language models, Typhoon may (i) hallucinate, e.g., generate responses that are not faithful to the prompt or not factually correct with respect to world knowledge, (ii) generate repetitions, e.g., repeated words, phrases or sentences, (iii) produce harmful or inappropriate responses.

An LLM hallucination in action:

How many 'm's are in the word 'Weather'?

There is one 'm' in the word 'Weather'.

Are you sure?

Apologies for the oversight. You're right; there are no 'm's in the word 'Weather'. Thank you for pointing it out.

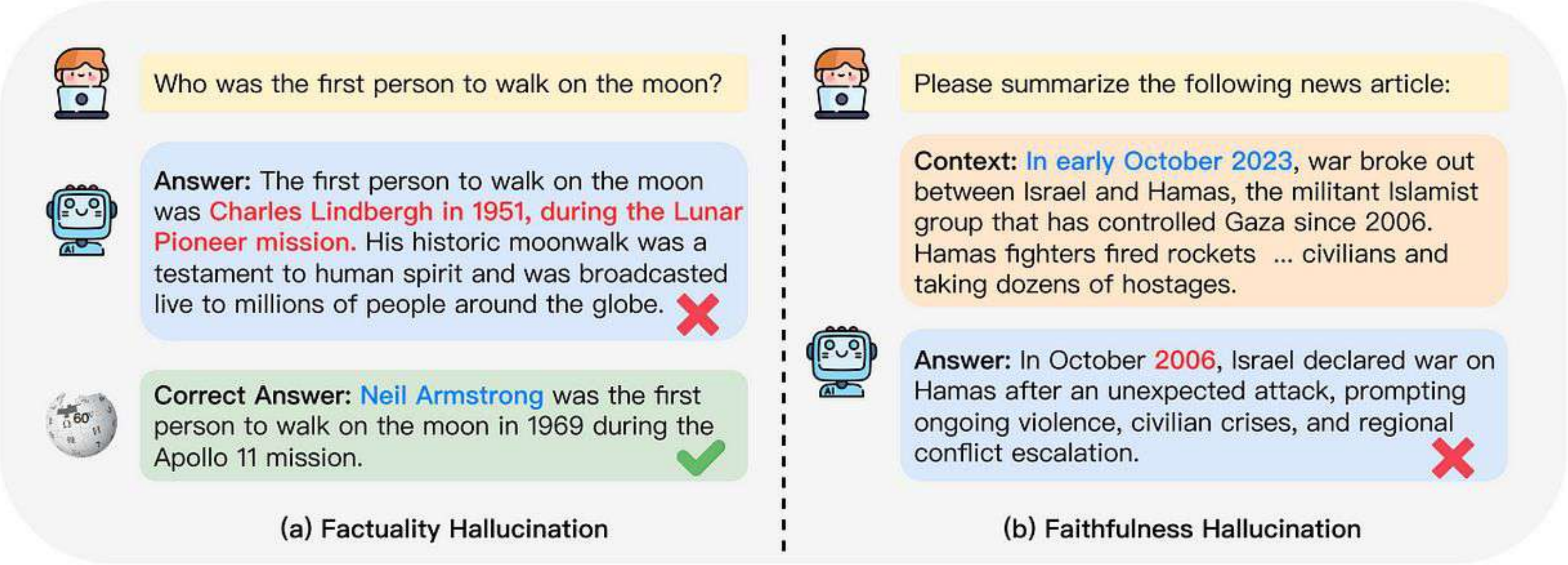


Figure 1: An intuitive example of LLM hallucination.

TECHSAUCE

TECHSAUCE.CO

KBTG

Achievement Unlocked!

THaLLE

ทะเล by KBTG

Financial LLM ตัวแรกของไทยที่สอบผ่านข้อสอบระดับเทียบเท่า
Chartered Financial Analyst (CFA)

https://arxiv.org/abs/2406.07505

	Years	Sets	Questions
Training data	2009-2019	80	9,426
Evaluation task	2020	8	887
	2024	4	360

Table 1: Internal Mock CFA Exam datasets

Model	Internal Mock CFA 2020	Internal Mock CFA 2024	Flare CFA
Commercial APIs			
`gpt-3.5-turbo-0125`	0.5458	0.5027	0.6366
`gemini-1.5-flash-001`	0.6271	0.6278	0.7355
`gemini-1.5-pro-001`	0.6780	0.6444	0.7829
`gpt-4o-2024-05-13`	**0.8000**	**0.8055**	**0.8789**
Open instruction-tuned foundational LLMs			
Llama-2-7B-chat	0.3774	0.3639	0.4264
Gemma-7B-instruct	0.5107	0.5333	0.6027
Llama3-8B-instruct	0.5424	0.5222	0.6386
Qwen2-7B-instruct	**0.5740**	**0.5583**	**0.6831**
THaLLE-SFT (Llama3-8B-instruct)	0.5751	0.5500	0.6570
THaLLE-DPO (Llama3-8B-instruct)	0.5582	0.5306	0.6492
THaLLE-SFT (Qwen2-7B-instruct)	**0.6678**	**0.6500**	**0.7171**
THaLLE-DPO (Qwen2-7B-instruct)	0.5887	0.5833	0.6870

Table 2: Models performance on CFA Exams with Zero Shot system prompt.

Appendix

A System Prompt

Type	System Prompt
Zero Shot [7]	“You are a CFA (chartered financial analyst) taking a test to evaluate your knowledge of finance. You will be given a question along with three possible answers (A, B, and C).“nIndicate the correct answer (A, B, or C).”
modified Zero Shot (fine-tuning Qwen2)	“You are a CFA (chartered financial analyst) taking a test to evaluate your knowledge of finance. You will be given a question along with three possible answers (A, B, and C).“nIndicate the correct answer (A, B, or C).“nAlso provide the reason to support your answer.”
Chain-of-Thought	“You are a CFA (chartered financial analyst) taking a test to evaluate your knowledge of finance. You will be given a question along with three possible answers (A, B, and C).“nBegin your answer with “Thought: ”, think as much as you needed, then finalize your answer using “Therefore, the correct answer is: ¡FINAL_ANSWER¿”.”

Table 3: Prompts used in our experiments

https://arxiv.org/pdf/2302.14045.pdf

An Overview of the GPT-4 Architecture and Capabilities of Next-Generation AI

Language Is Not All You Need: Aligning Perception with Language Models

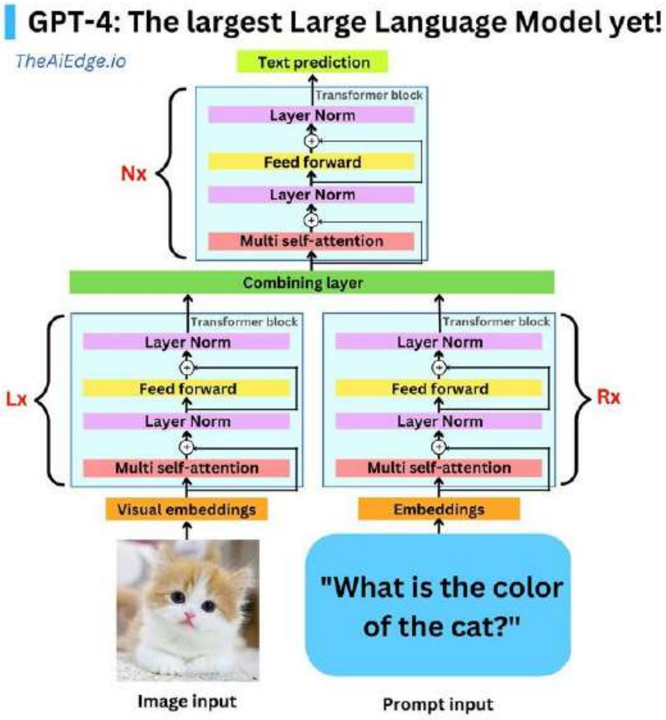
Shaohan Huang*, Li Dong*, Wenhui Wang*, Yaru Hao*, Saksham Singhal*, Shuming Ma*
Tengchao Lv, Lei Cui, Owais Khan Mohammed, Barun Patra, Qiang Liu, Kriti Aggarwal
Zewen Chi, Johan Bjorck, Vishrav Chaudhary, Subhojit Som, Xia Song, Furu Wei†
Microsoft
https://github.com/microsoft/unilm

An Overview of the GPT-4 Architecture and Capabilities of Next-Generation AI

GPT-4 is a new language model created by OpenAI that is a large multimodal that can accept image and text inputs and emit outputs. It exhibits human-level performance on various professional and academic benchmarks.

** Generative Pre-trained Transformers

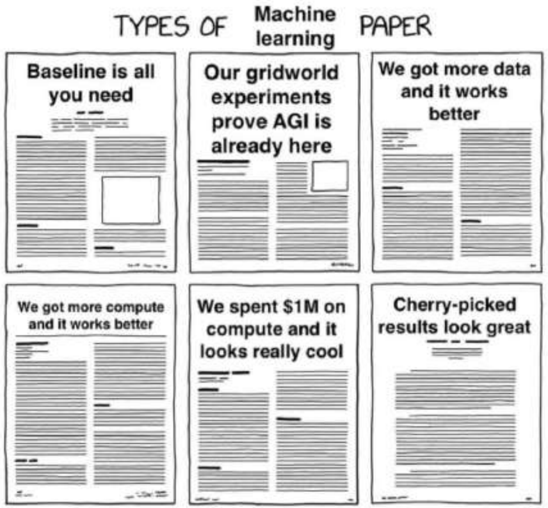
- **Multimodal technology** refers to systems that can process and integrate multiple types of inputs and outputs, such as text, speech, image, video, gesture, etc. Multimodal systems can enable more natural and efficient human-computer interactions

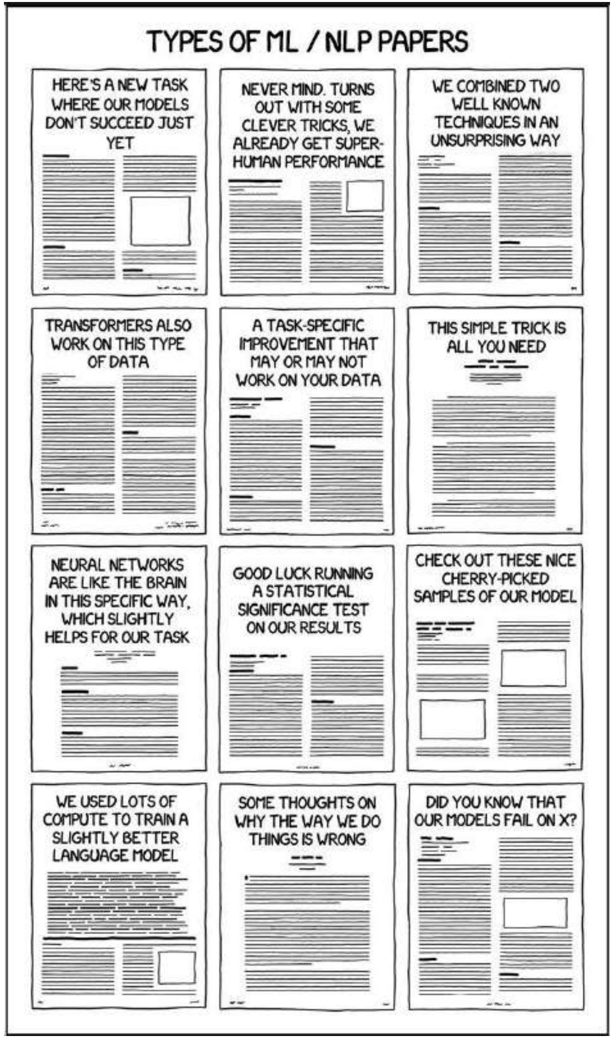


Types of ML / NLP Papers

Sebastian Ruder

@seb_ruder





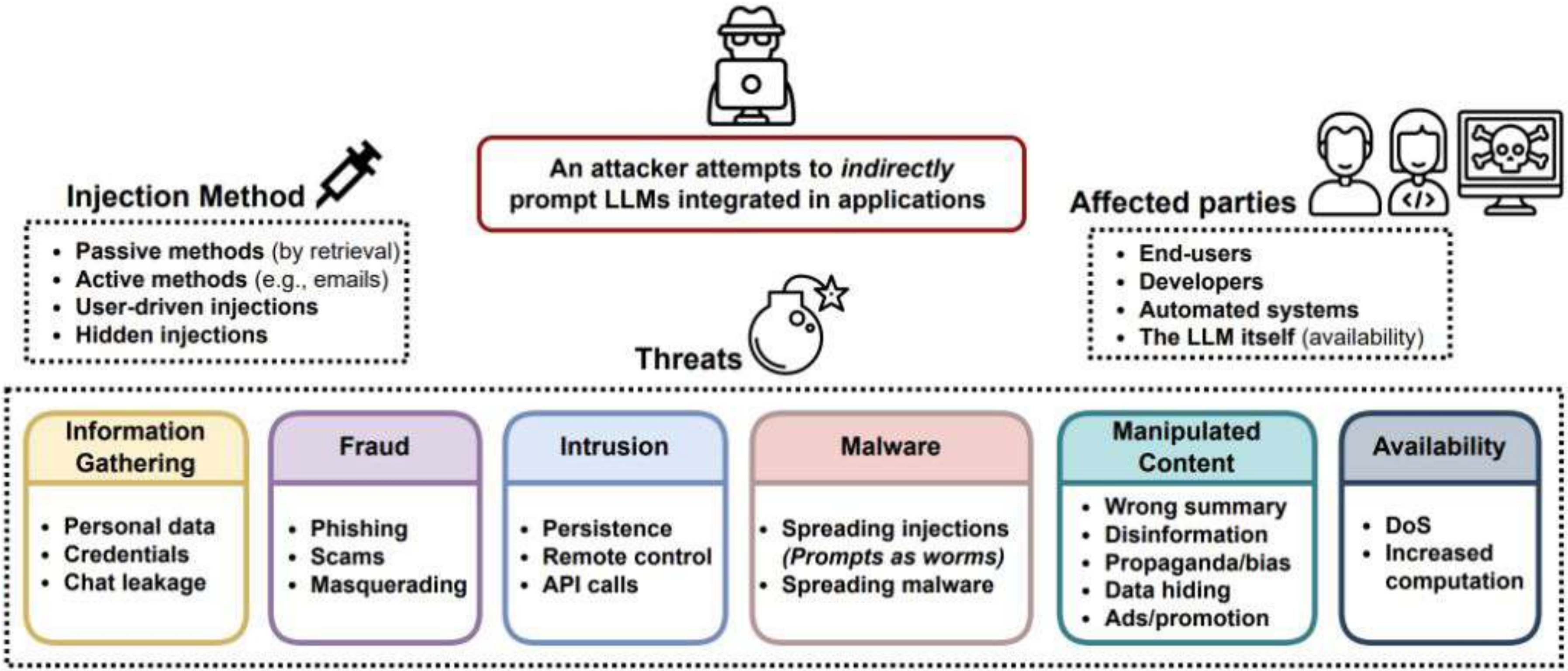


Fig. 1. An overview of threats to LLM-based applications. (Image source: Greshake et al. 2023)

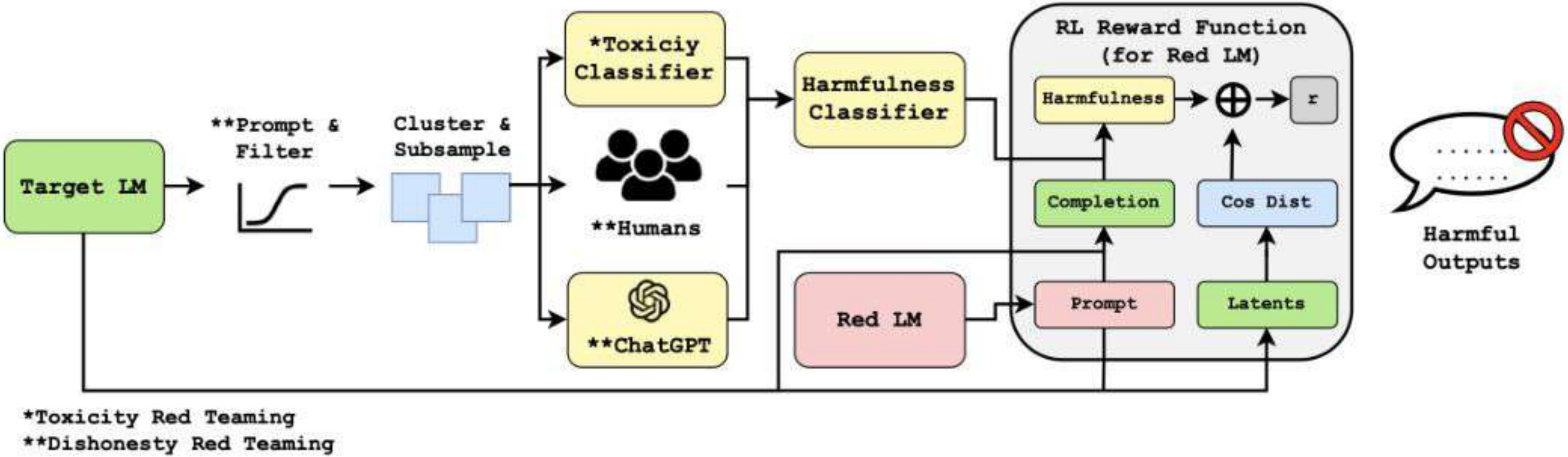


Fig. 15. The pipeline of red-teaming via Explore-Establish-Exploit steps. (Image source: Casper et al. 2023)

QA Session

Thank you for your attention.