Team 007

Working in NLP in the Age of Large Language Models

Maleke Chaker
Fedi Koubaa
Nour Chaker
Abir Belhedi

PLAN

1. Key Idea of NLP

- Tokenization
- Text Normalization
- Embeddings

2. Foundational Roots of NLP

- RNN / LSTM / XLSTM
- Attention / Transformers
- Large Language Models
- Fine-tuning
- Prompt Engineering
- Tooling Ecosystem

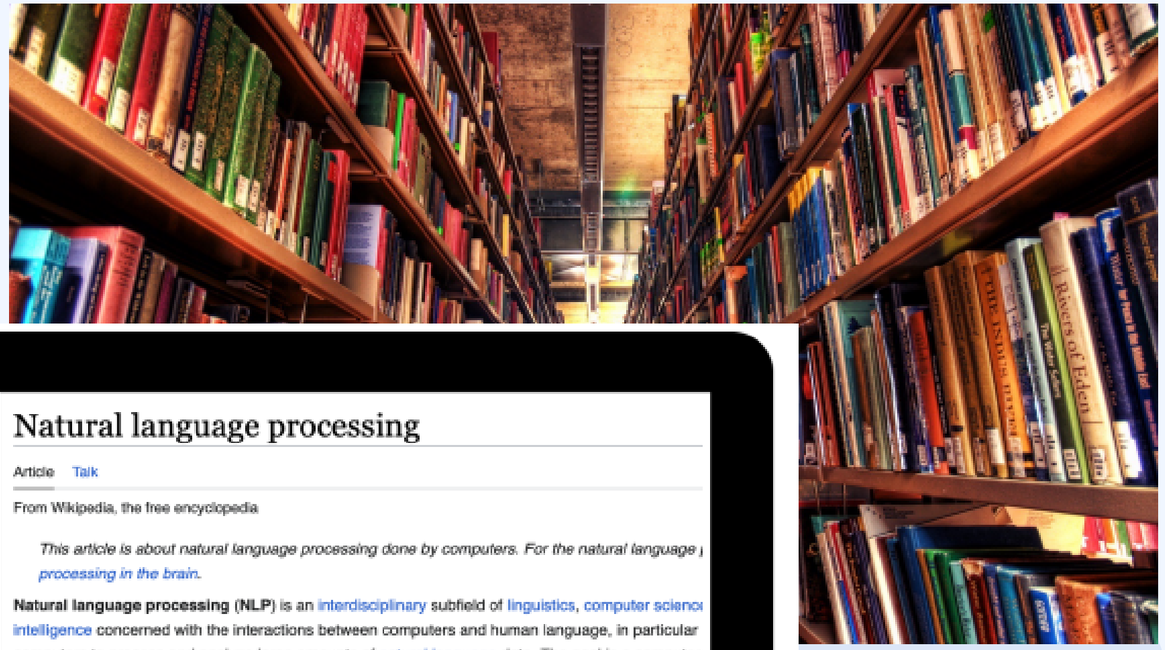
3. Advanced Architectures

- Retrieval-Augmented Generation
- LLM based agents
- Multi-Modal Language Models

4. Our Demo

1. Key Idea of NLP

Text data is everywhere!



To make sense of the vast amount of text data around us, we turn to Natural Language Processing (NLP)

What is NLP?

A field of AI focused on enabling computers to understand, and generate human language.

Key Idea of NLP

With a sufficiently large corpus of text data, models can learn the patterns of language

What can you do with NLP?

Natural Language Understanding

Text Classification

Information Extraction / Named Entity Recognition

Sentiment Analysis

Natural Language Generation

Machine Translation

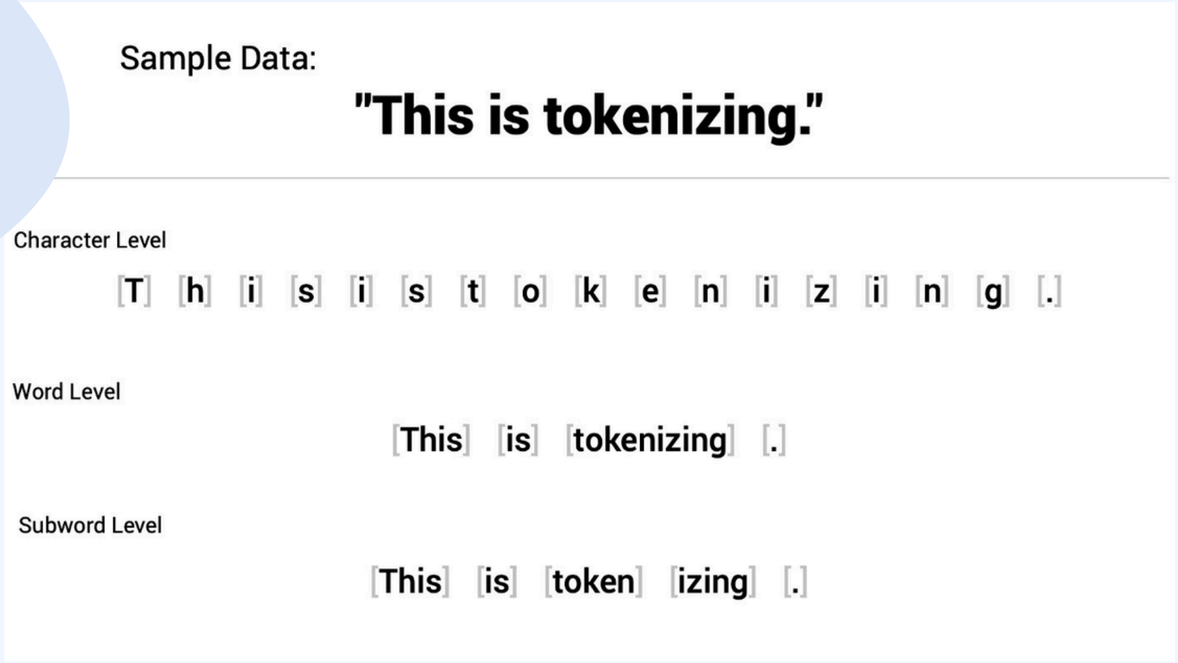
Text Summarization

Chatbots

NLP Foundations

Tokenization

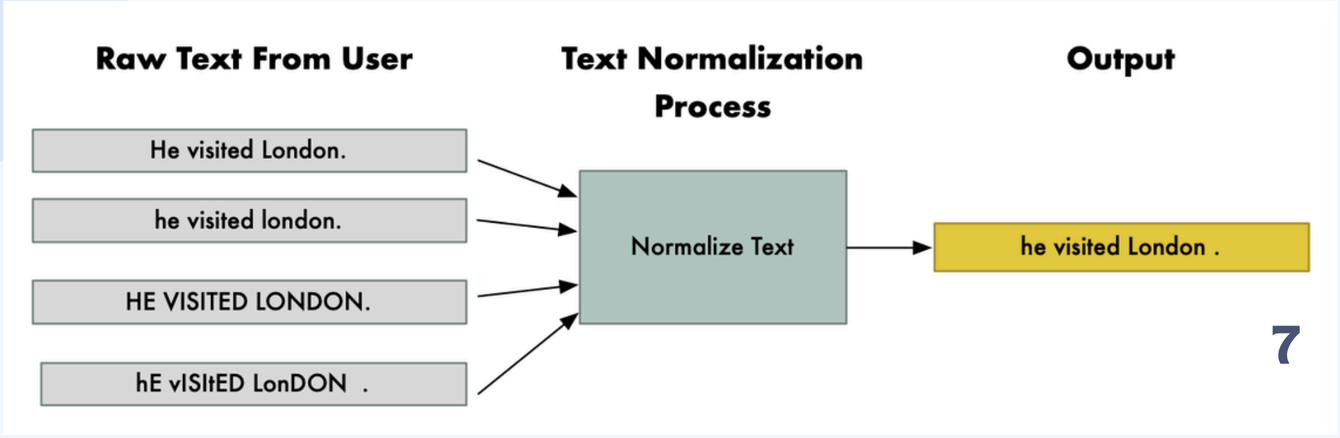
Breaking down text into individual words or phrases (tokens)



Text Normalization

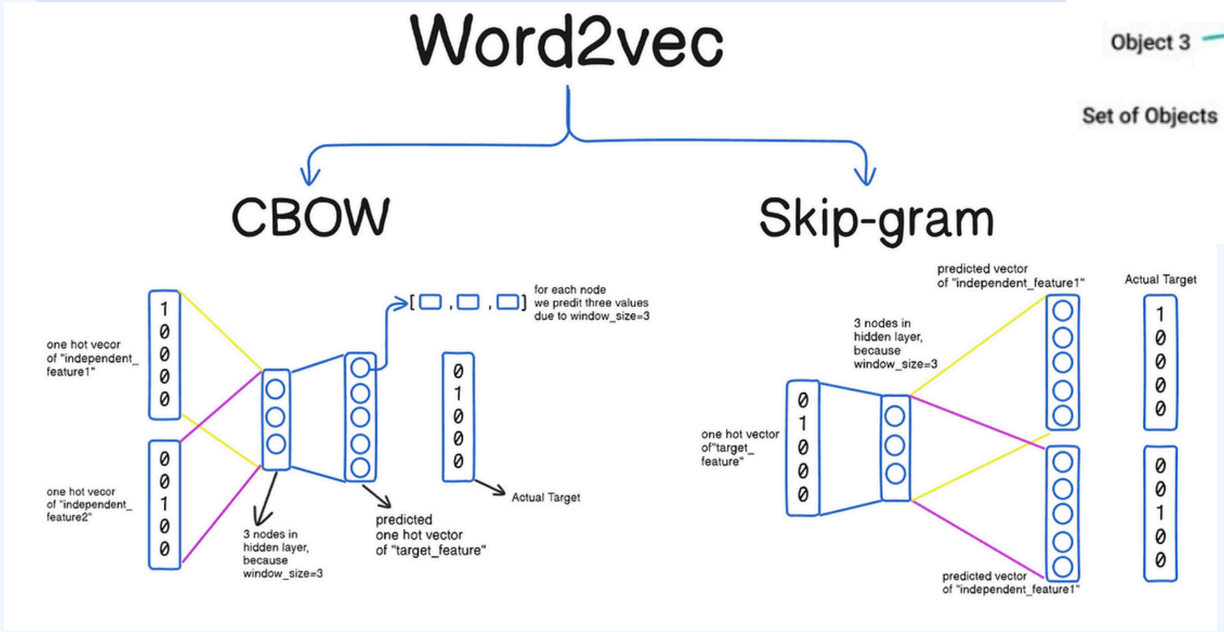
Cleaning and standardizing text:

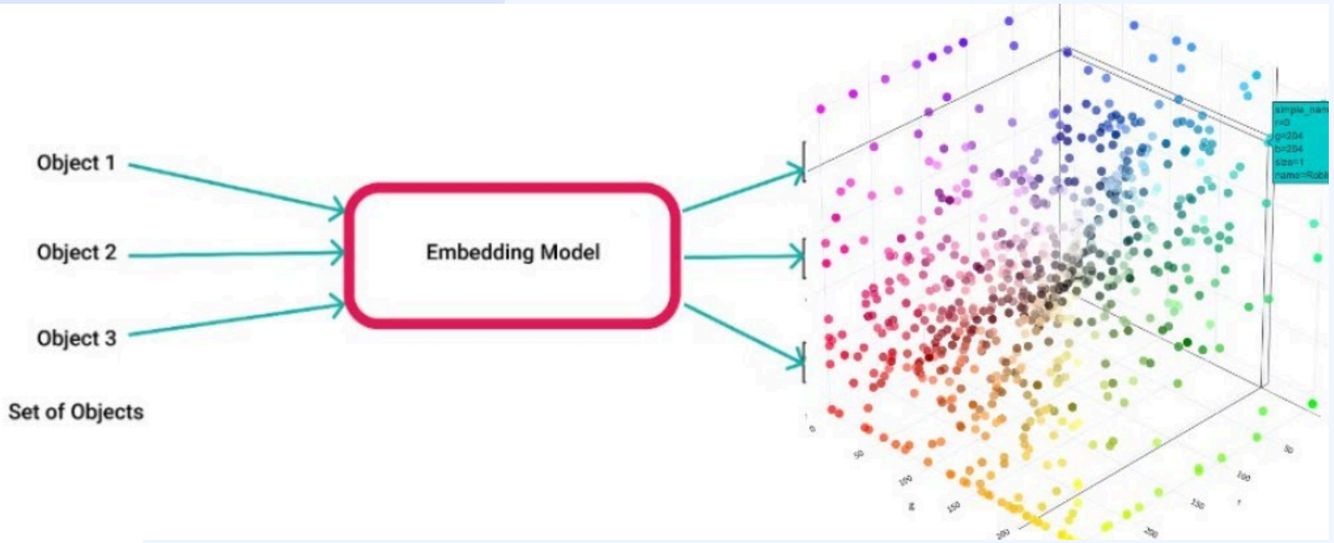
- Lowercasing
- Stemming
- Lemmatization



Embeddings

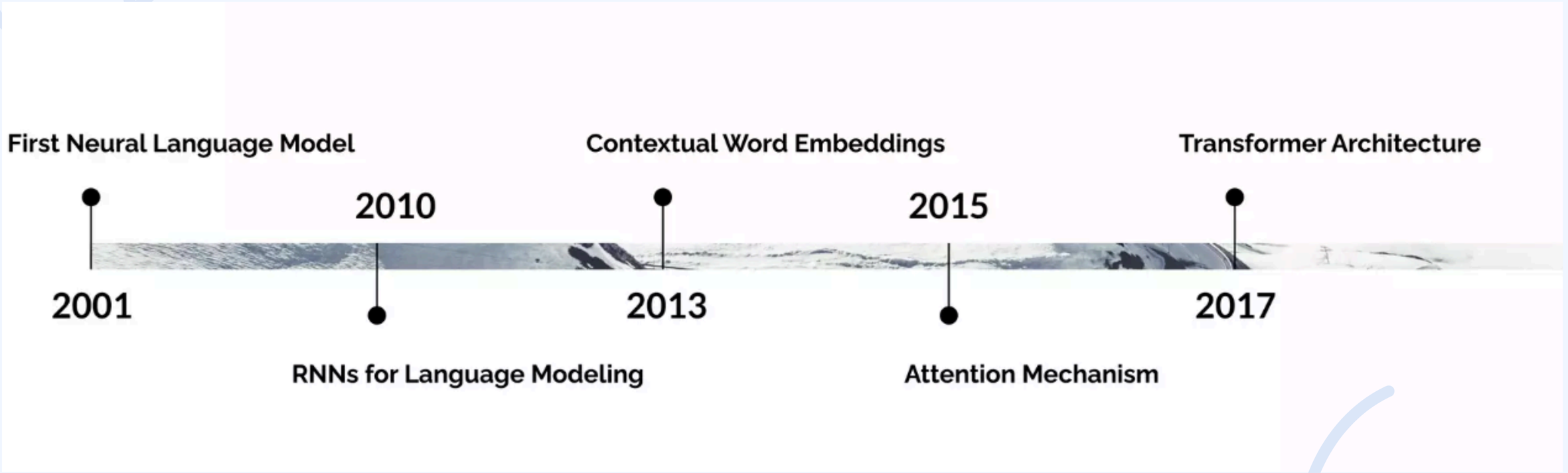
Representing words or phrases as dense vectors .
Semantically similar words have vectors that are close together.





2. Foundational Roots of NLP

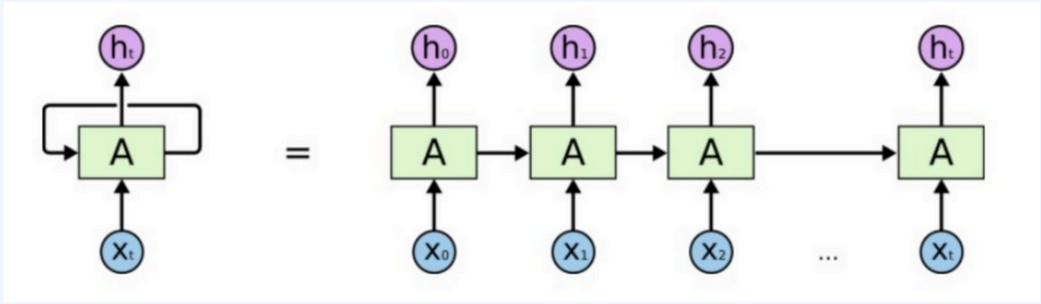
Foundational Roots of NLP



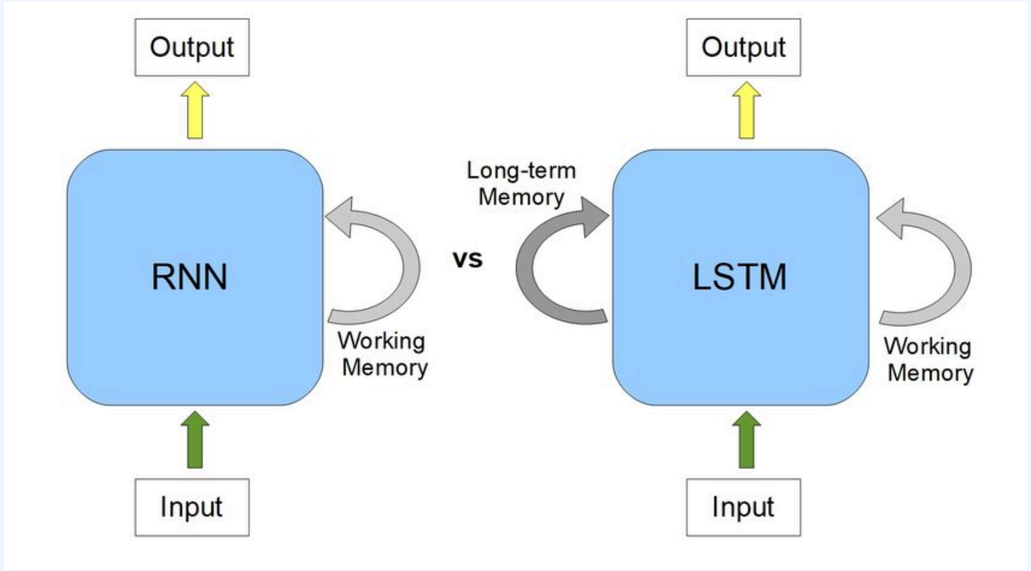
1

Recurrent Neural Networks

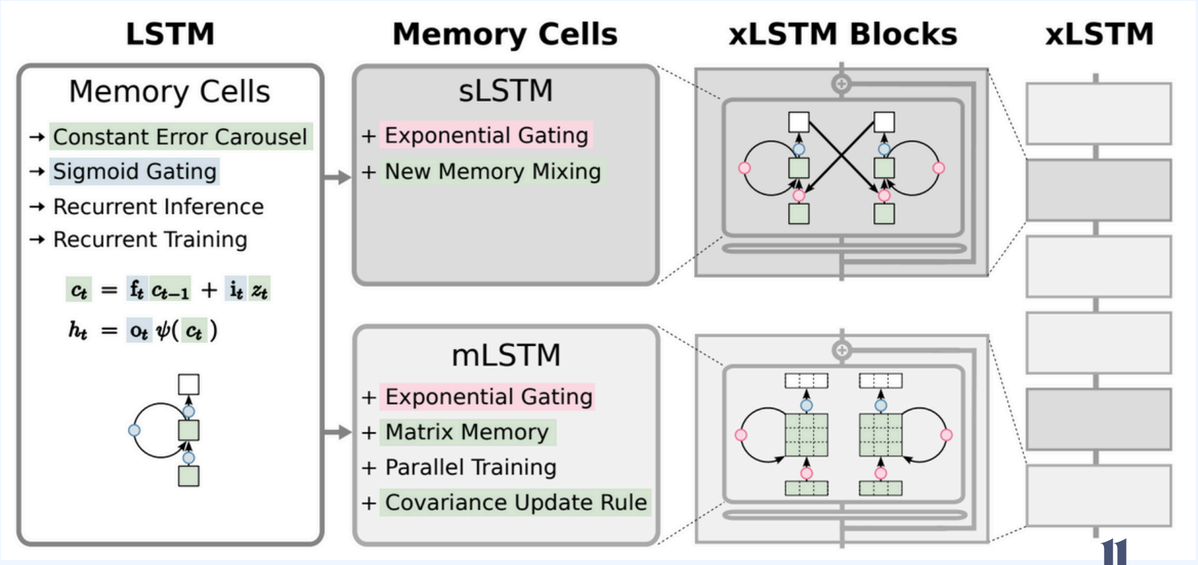
Limitations of RNNs:
Difficulty handling long-range dependencies in text.



LSTM



Extended LSTM



2

Attention Mechanism

Allows the model to focus only on the most relevant parts of the input text

Transformers

A new architecture based on:

- **Attention Mechanism**
- **Self-Attention**: Relates different positions of a single sequence to compute a representation.
- **Encoder-Decoder Structure**

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

3

Large Language Models (LLMs)

Transformer-based models trained
on massive amounts of text data.

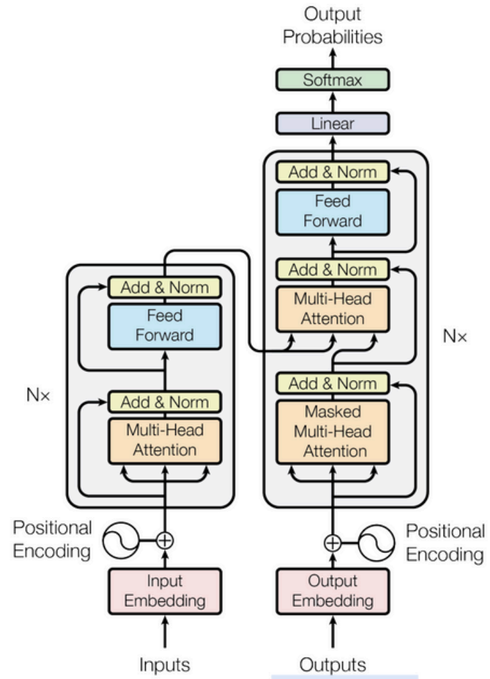
LLM Applications

Question Answering	Code Generation
Chatbots	Text Summarization
AGI (Artificial General Intelligence)	Creative Writing

Model architectures

BERT

Encoder

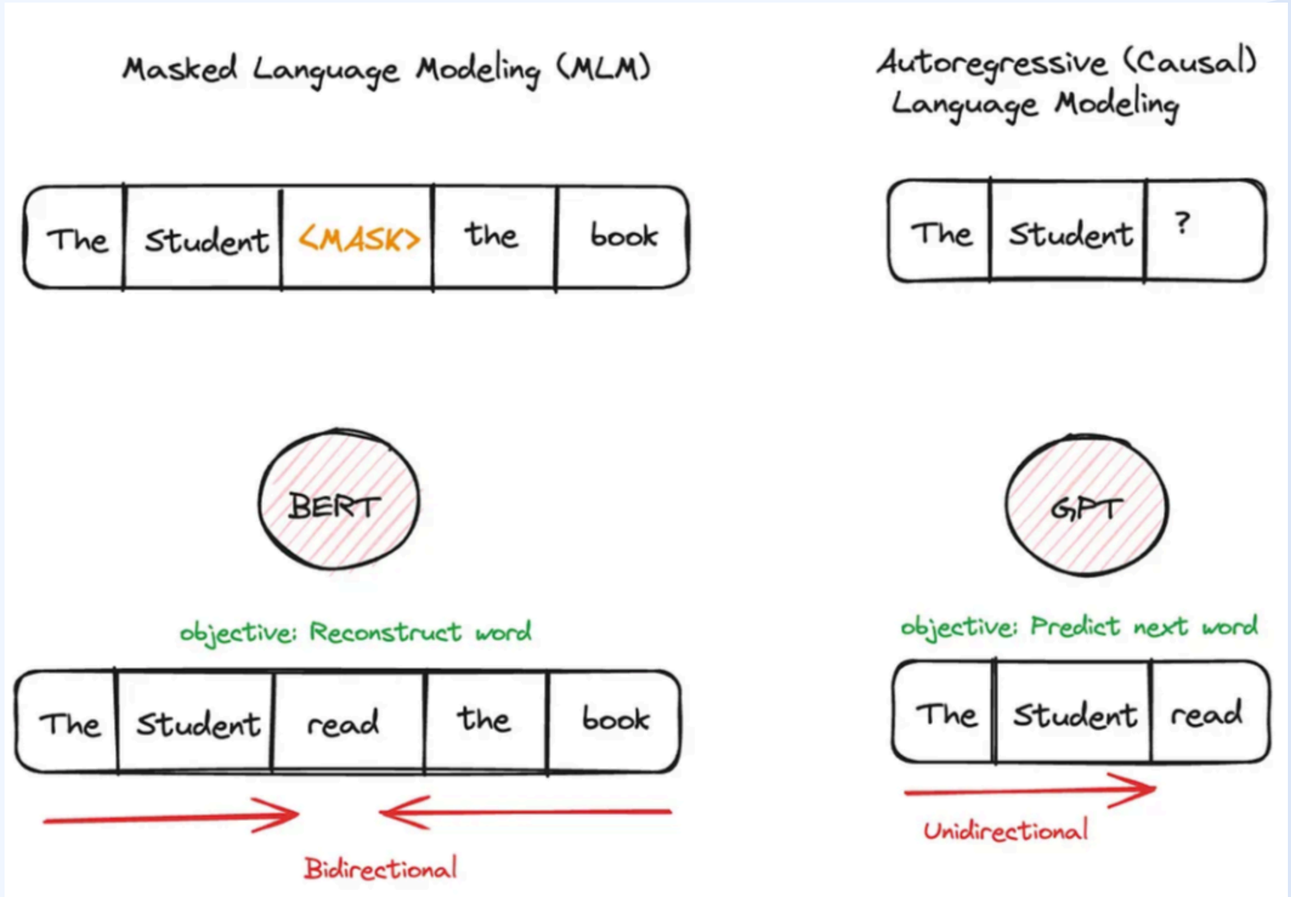


GPT

Decoder

3 BERT vs GPT

	BERT	GPT
Model Type	Encoder Only	Decoder Only
Direction	Bidirectional	Unidirectional (left-to-right)
Pre-training Objective	Masked language modeling (MLM)	Autoregressive (casual) language modeling
Fine-tuning	Task-specific layer added on top of the pre-trained BERT model	Providing task-specific prompts using few-shot or one-shot adaptation and adapting the model's parameters
Use Case	Sentiment Analysis Named entity Recognition Word Classification	Text generation Text completion creative writing
Original Organisations	Google AI	OpenAI



Fine-tuning LLMs

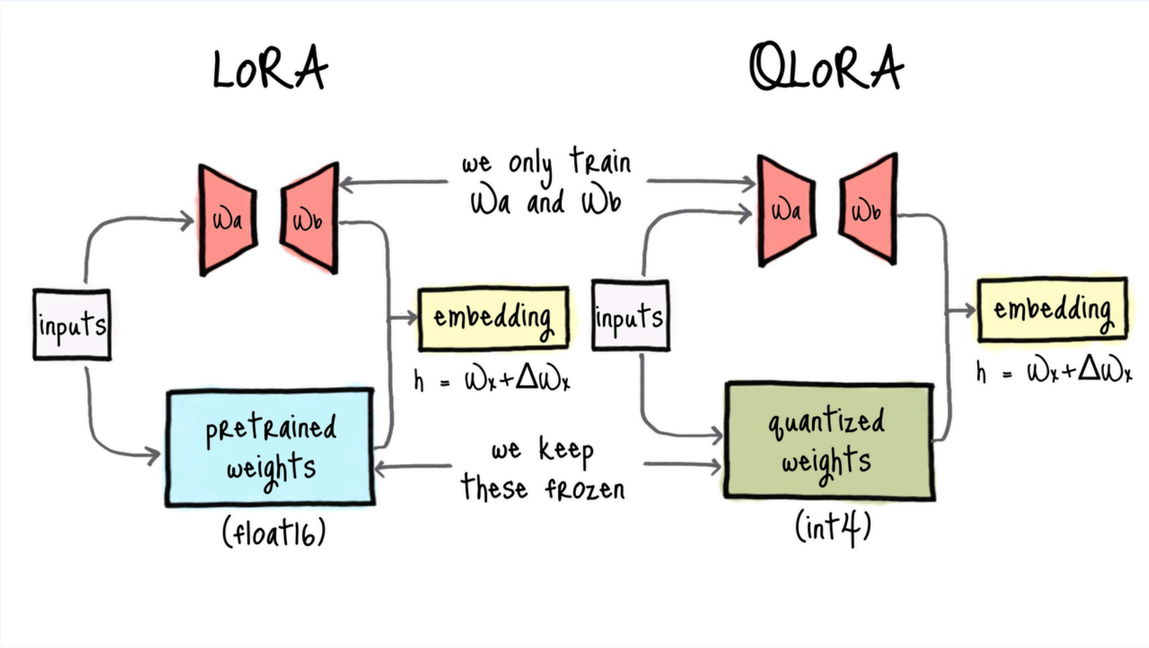
Adapting pre-trained LLMs to specific tasks.

LoRA (Low-Rank Adaptation)

Fine-tunes only a small subset of weights, reducing memory usage.

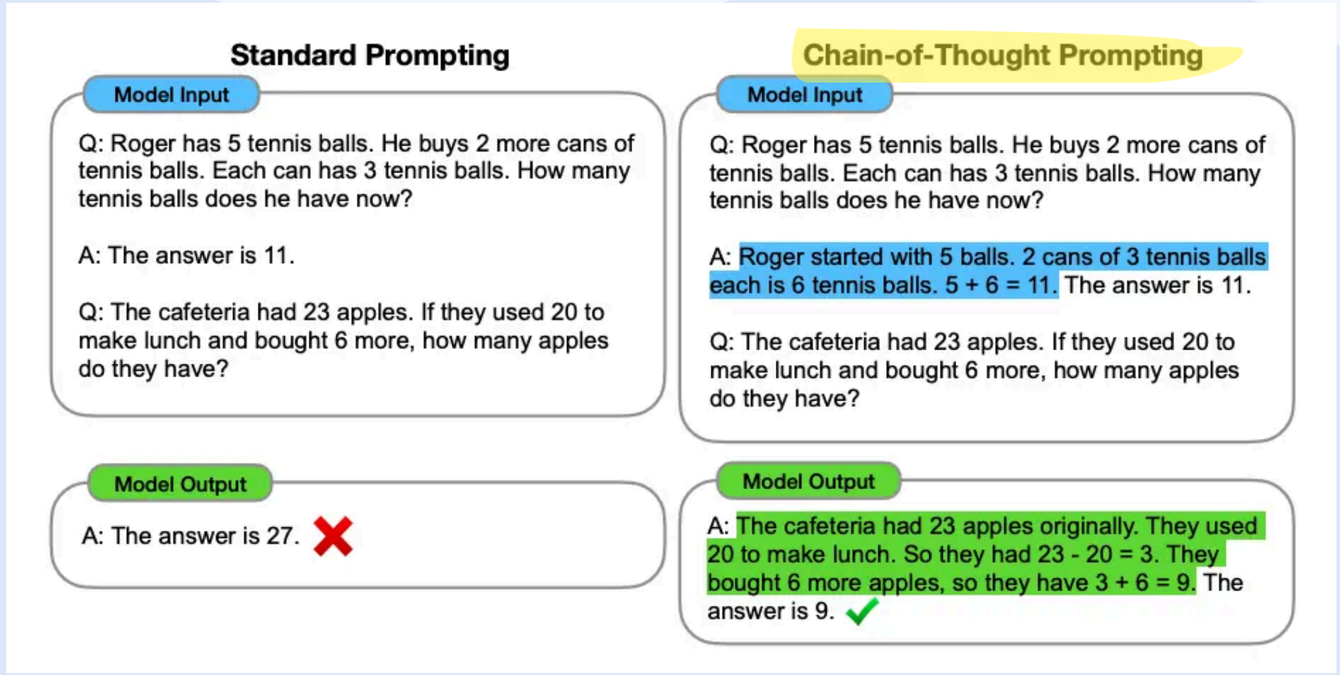
QLoRA (Quantized LoRA)

Uses 4-bit quantization (Compresses the model) to reduce memory while fine-tuning a small subset of weights.

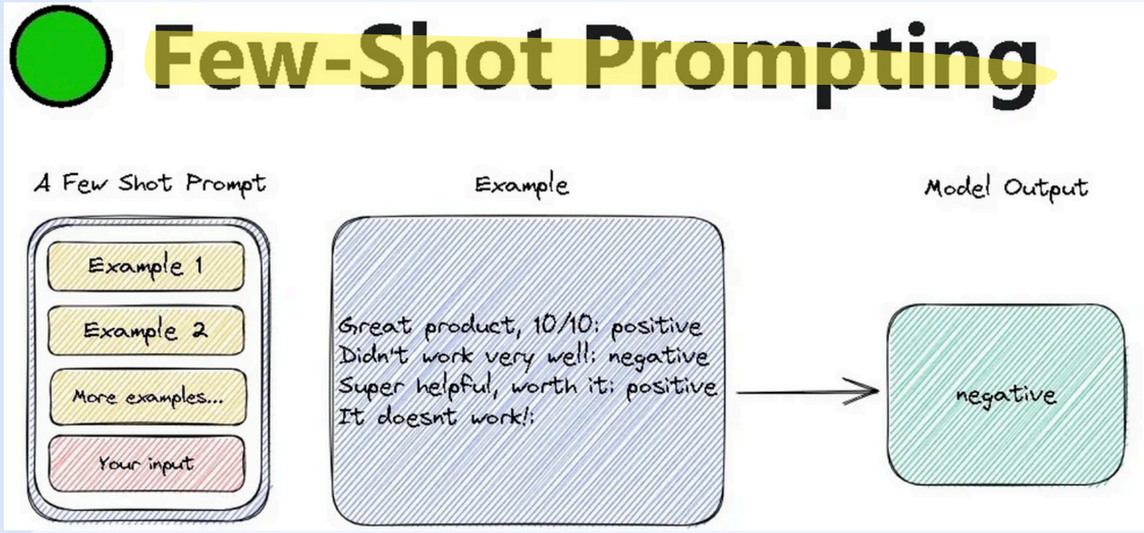


Prompt Engineering

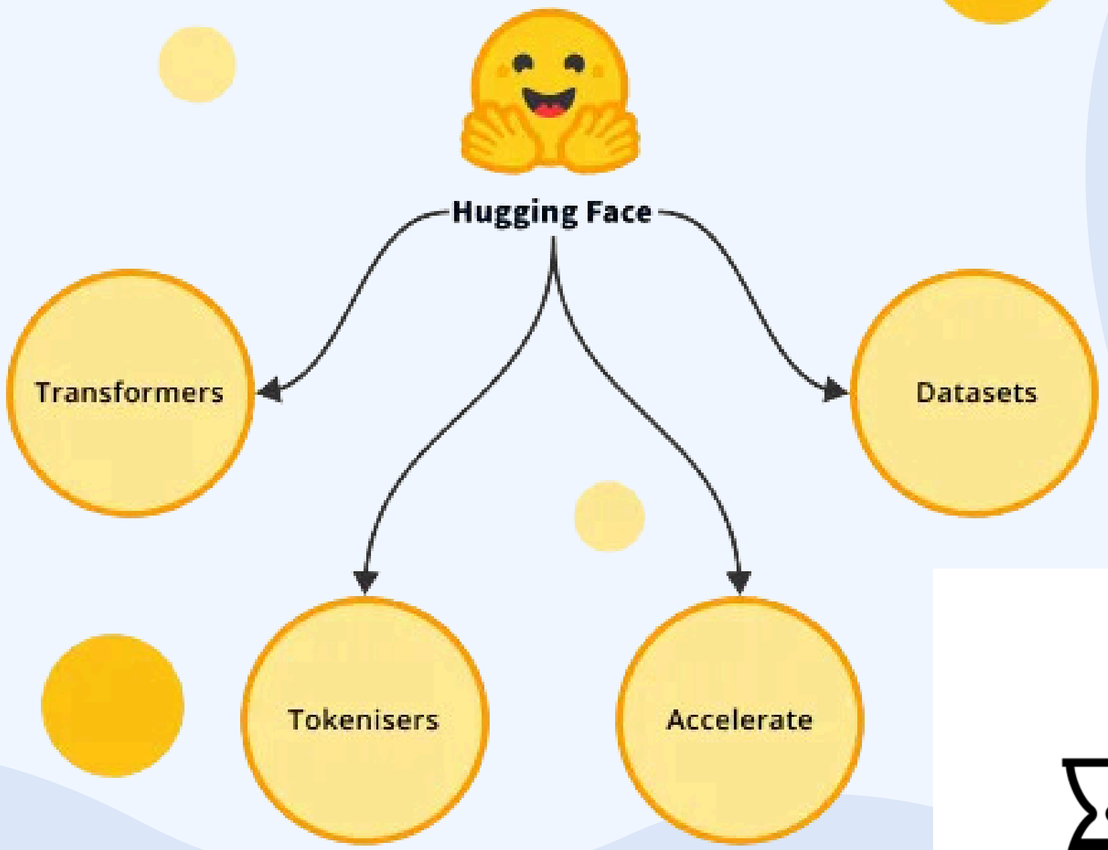
Crafting effective prompts to guide LLM behavior.



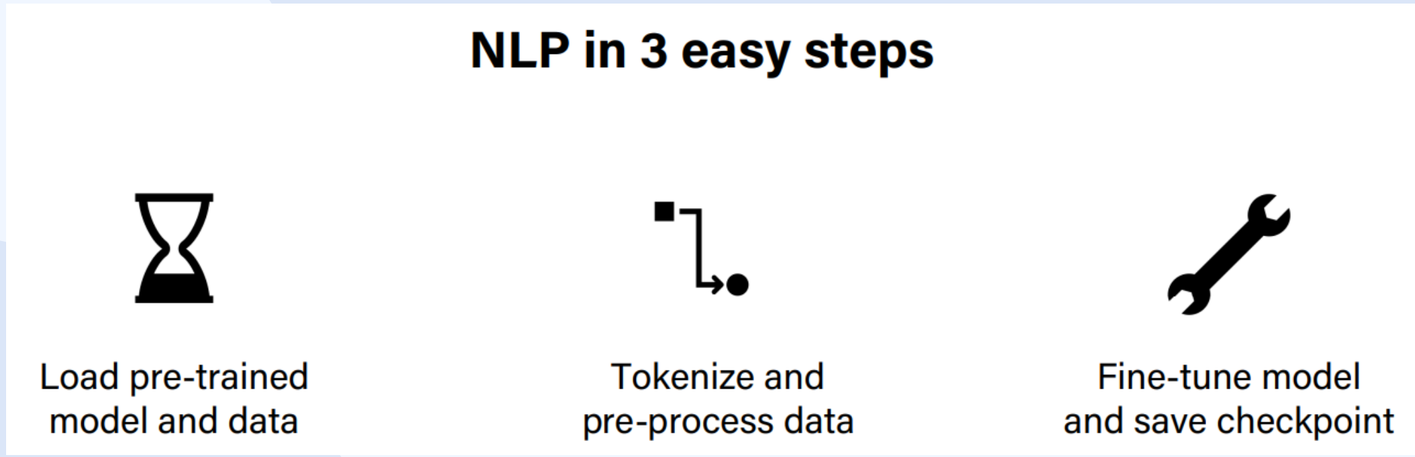
Guiding the model to reason step-by-step.



Learning from a handful of examples.



GettingStarted with Hugging Face



User -friendly resource to help you get started with NLP .

It's a hub for Models/datasets for variety of different tasks.

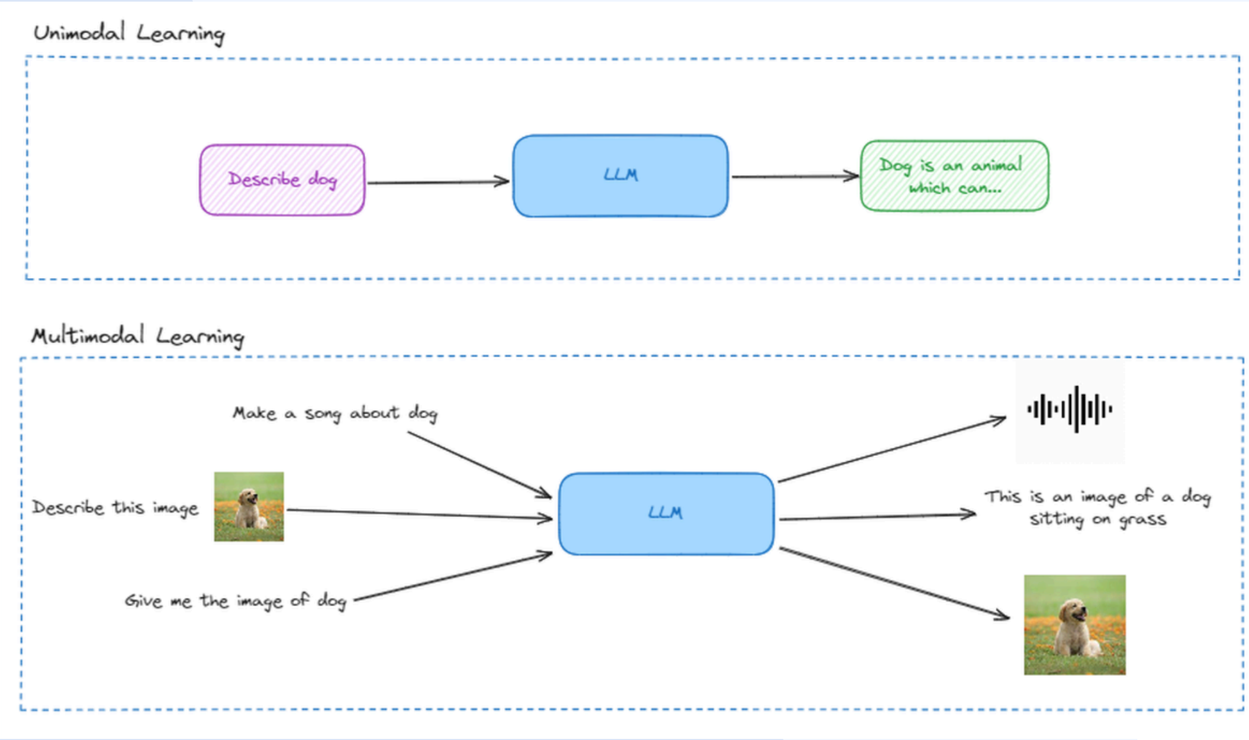
Tooling Ecosystem

To support this new emerging ecosystem of LLMs, a whole host of new tools have emerged, and some existing solutions have found new life. Some prominent areas of novel tooling include:

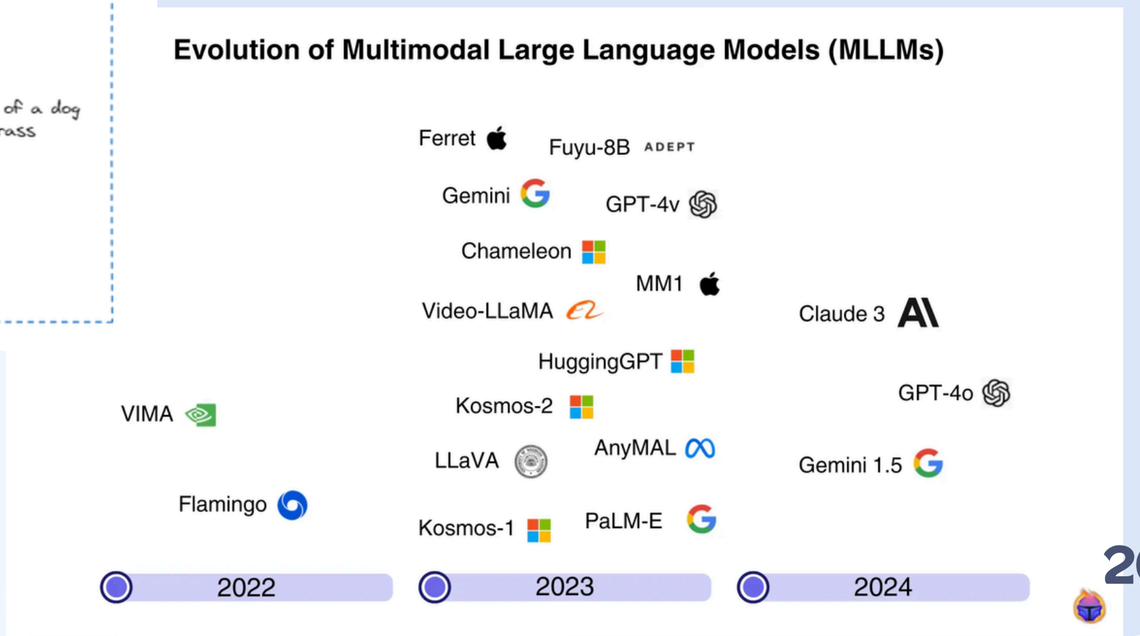
- **Chain creation and management**
- **Vector data stores**
- **LLM training and inference platforms**
- **LLM testing and monitoring suites**
- **Labeling platforms**

3.Advanced Architectures

Advanced Architectures
Multi-Modal Language Models

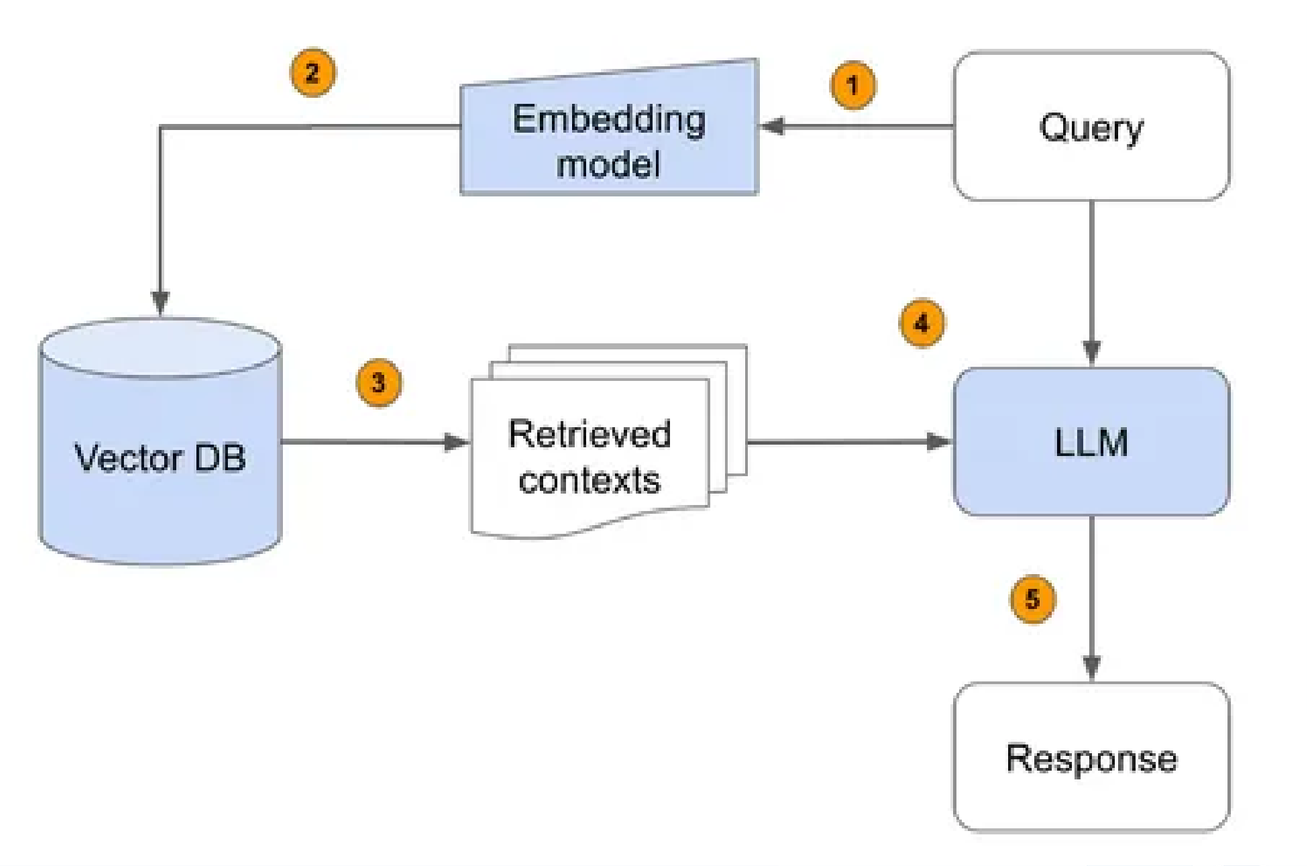


Advanced AI models that combine text-based language models with other modalities such as images, audio, or video to enhance their understanding and generation capabilities.



Advanced Architectures

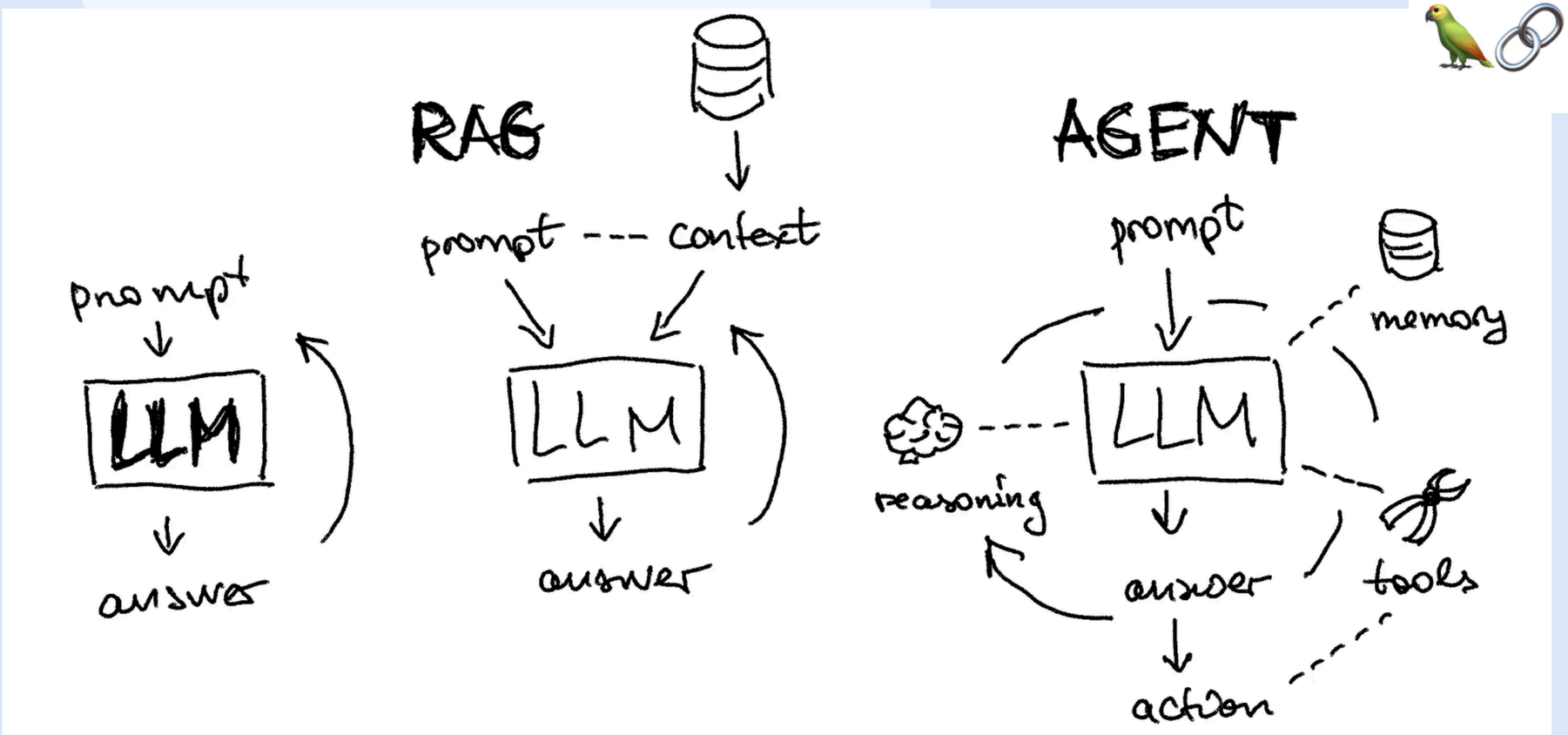
Retrieval-Augmented Generation



Enhances LLMs by retrieving relevant context from external knowledge sources. This improves LLM accuracy and contextual understanding, leading to better responses.

Advanced Architectures
LLM based agents

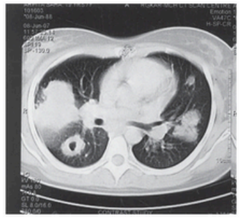
LLM serves as the main controller or "brain" that controls a flow of operations needed to complete a complex task or user request.



4. Our Demo

Our Demo

Medical Visual Question Answering



Q: What does the ct scan of thorax show?
A: bilateral multiple pulmonary nodules

Q: Is the lesion associated with a mass effect?
A: no

Organ System

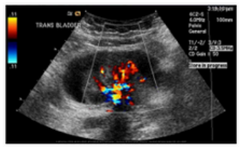
Q: What is the organ system?
A: Gastrointestinal

Object/Condition Presence

Q: Is there gastric fullness?
A: yes

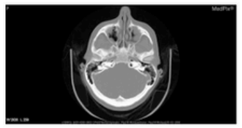
Positional

Q: What is the location of the mass?
A: head of the pancreas



Modality

Q: what imaging method was used?
A: us-d - doppler ultrasound



Plane

Q: which plane is the image shown in?
A: axial

Given a medical ***image*** and a clinically relevant ***question*** in natural language, the medical VQA system is expected to predict a plausible and convincing ***answer***.

Our idea is based on This paper:

Medical Visual Question Answering: A Survey

Zhihong Lin[a], Donghao Zhang[b], Qingyi Tao[c], Danli Shi[d], Gholamreza Haffari[e], Qi Wu[f], Mingguang He[g] and Zongyuan Ge[b,h,i,*]

[a] Faculty of Engineering, Monash University, Clayton, VIC, 3800 Australia
[b] eResearch Center, Monash University, Clayton, VIC, 3800 Australia
[c] NVIDIA AI Technology Center, 038988, Singapore
[d] State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-Sen University, Guangzhou, 510060 China
[e] Faculty of Information Technology, Monash University, Clayton, 3800, VIC, Australia
[f] Australian Centre for Robotic Vision, The University of Adelaide, Adelaide, SA 5005, Australia
[g] Eye Research Australia, Royal Victorian Eye and Ear Hospital, East Melbourne, VIC, 3002 Australia
[h] Airdoc Research, Melbourne, VIC, 3000 Australia
[i] Monash-NVIDIA AI Tech Centre, Melbourne, VIC, 3000 Australia

ARTICLE INFO

ABSTRACT

Challenges

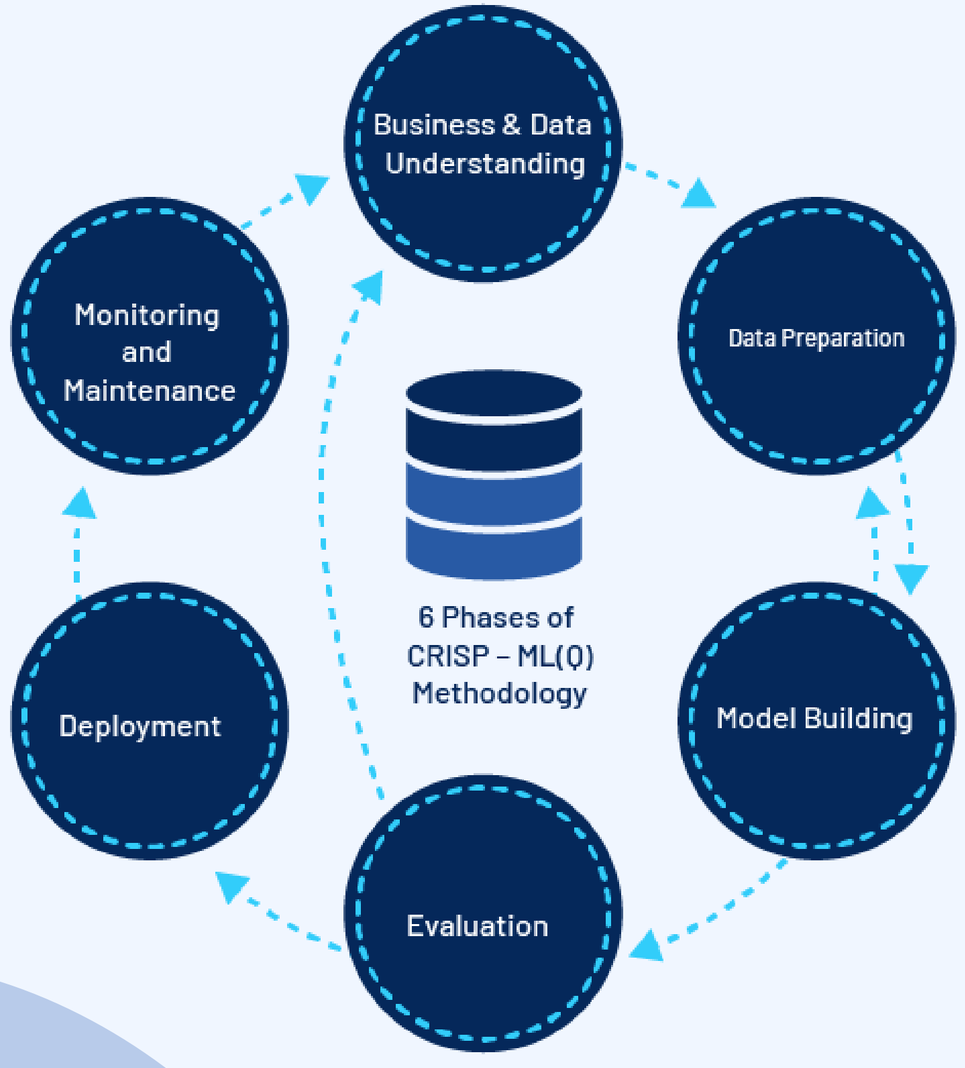
1. Data is tough
2. Details matter
3. Medical knowledge is key
4. Real-world needs

Purpose

1. Supporting clinical decisions
2. Increasing efficiency
3. Offering second opinions
4. Providing automated consultations
5. Delivering reliable online information

Methodology CRISP-DM

CRoss Industry Standard Process for Data Mining

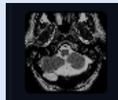
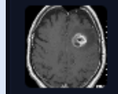
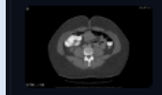
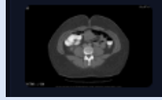


Data Preparation

```
DatasetDict({
    train: Dataset({
        features: ['image', 'question', 'answer'],
        num_rows: 27900
    })
    validation: Dataset({
        features: ['image', 'question', 'answer'],
        num_rows: 5580
    })
    test: Dataset({
        features: ['image', 'question', 'answer'],
        num_rows: 1396
```

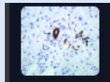
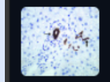
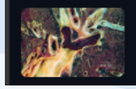
VQA-RAD:
with 2.2k rows

Radiology-focused VQA dataset with clinically-realistic questions about head, chest, and abdomen images.

	is the lesion causing significant brainstem herniation?	no
	how was this image taken?	mri
	what is the condition of the patient	blind loop syndrome
	what abnormality is seen?	blind-ending loop of bou cecum

Path-VQA:
with 19.7k rows

Pathology VQA dataset using textbook/library images with questions designed for pathologist certification exams.

	where are liver stem cells (oval cells) located?	in the canals of he
	what are stained here with an immunohistochemical stain for…	bile duct cells and
	what do the areas of white chalky deposits represent?	foci of fat necrosis
	is embolus derived from a lower-extremity deep venous thrombus lodged…	yes

The joint embedding framework

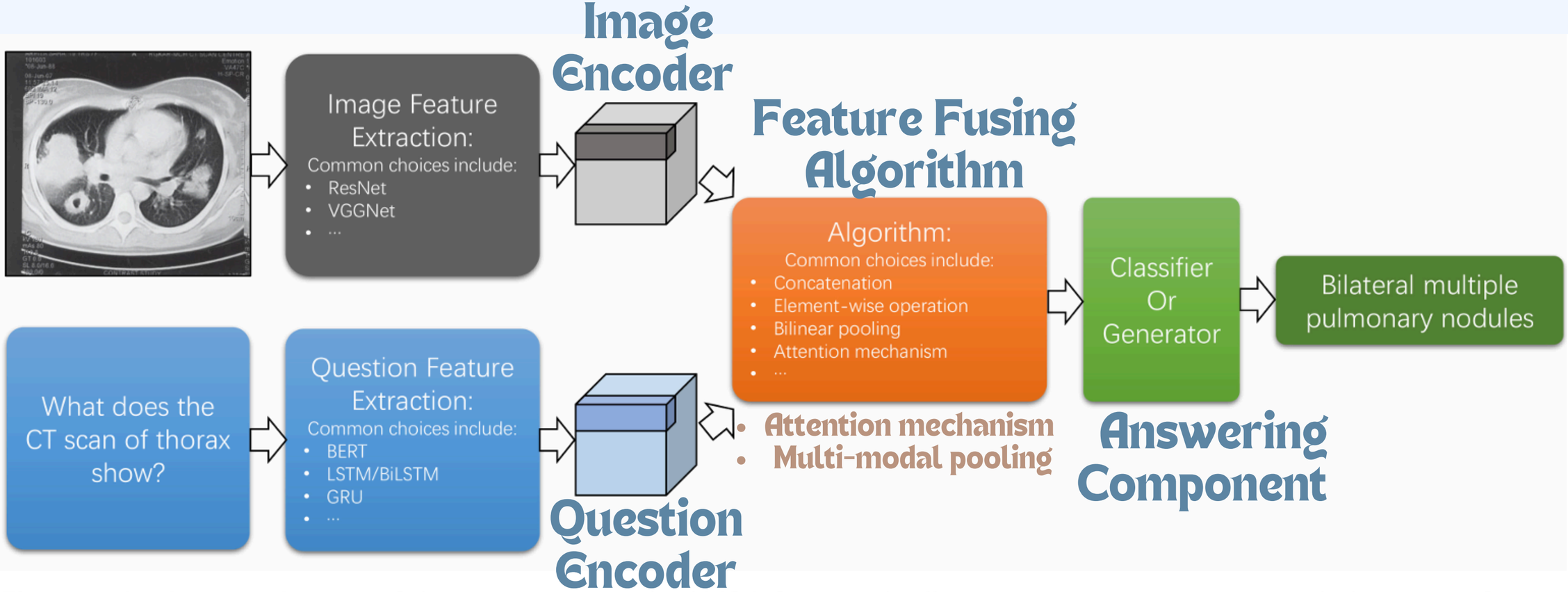
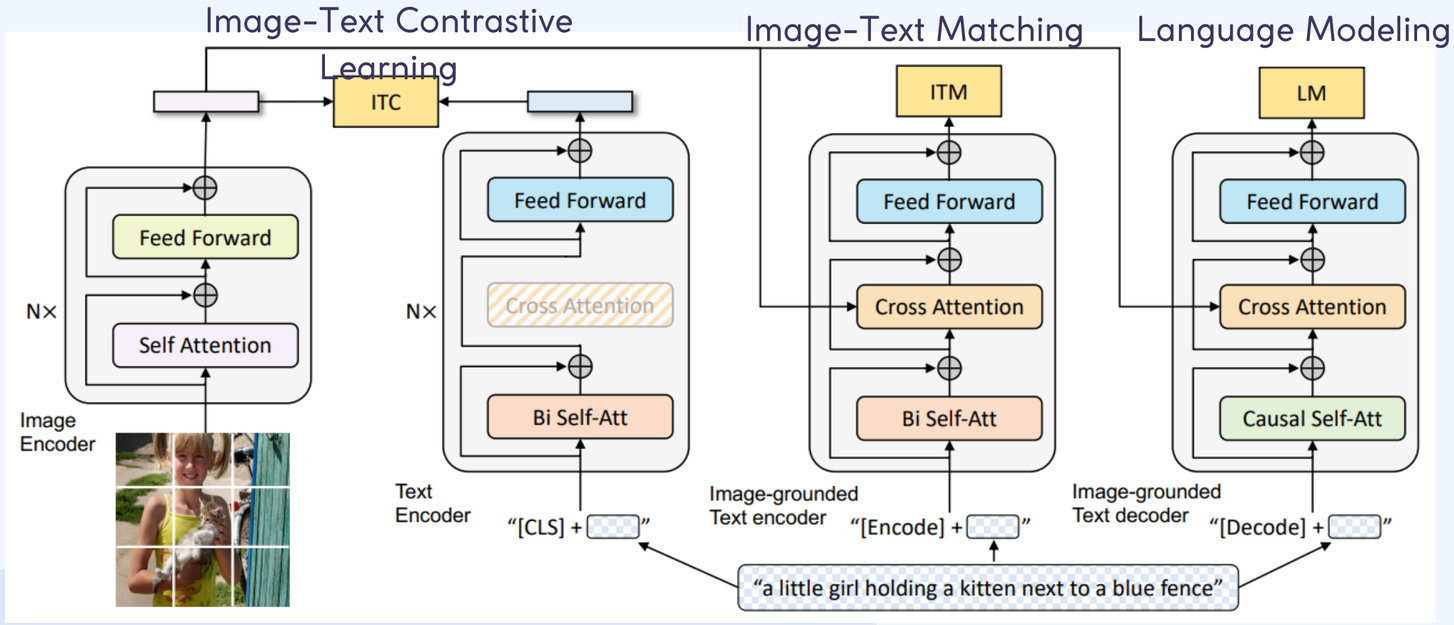


Figure 3: A schematic diagram of the mainstream medical VQA frameworks illustrating main components, including image feature extraction, question feature extraction, feature fusion, and answering component.

Our fine-tuned Model

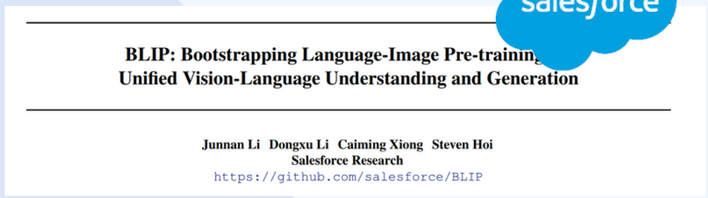
BLIP



Key Features:

- **Multimodal Mixture:** Combines image and text processing.
  - Image Transformer from **ViT-B**
  - Text Transformer from **BERT base**
  - Decoder to generate the answer
- **Powerful captioning :** Generates context-aware and descriptive captions for images.
- **Three pretraining Losses:** ITC, ITM, LM to enhance image-text alignment.
- **Filtering noisy data:** Removes irrelevant image-text pairs to improve training quality.
- Trained on **massive image-text data** (COCO, Visual Genome, etc.)

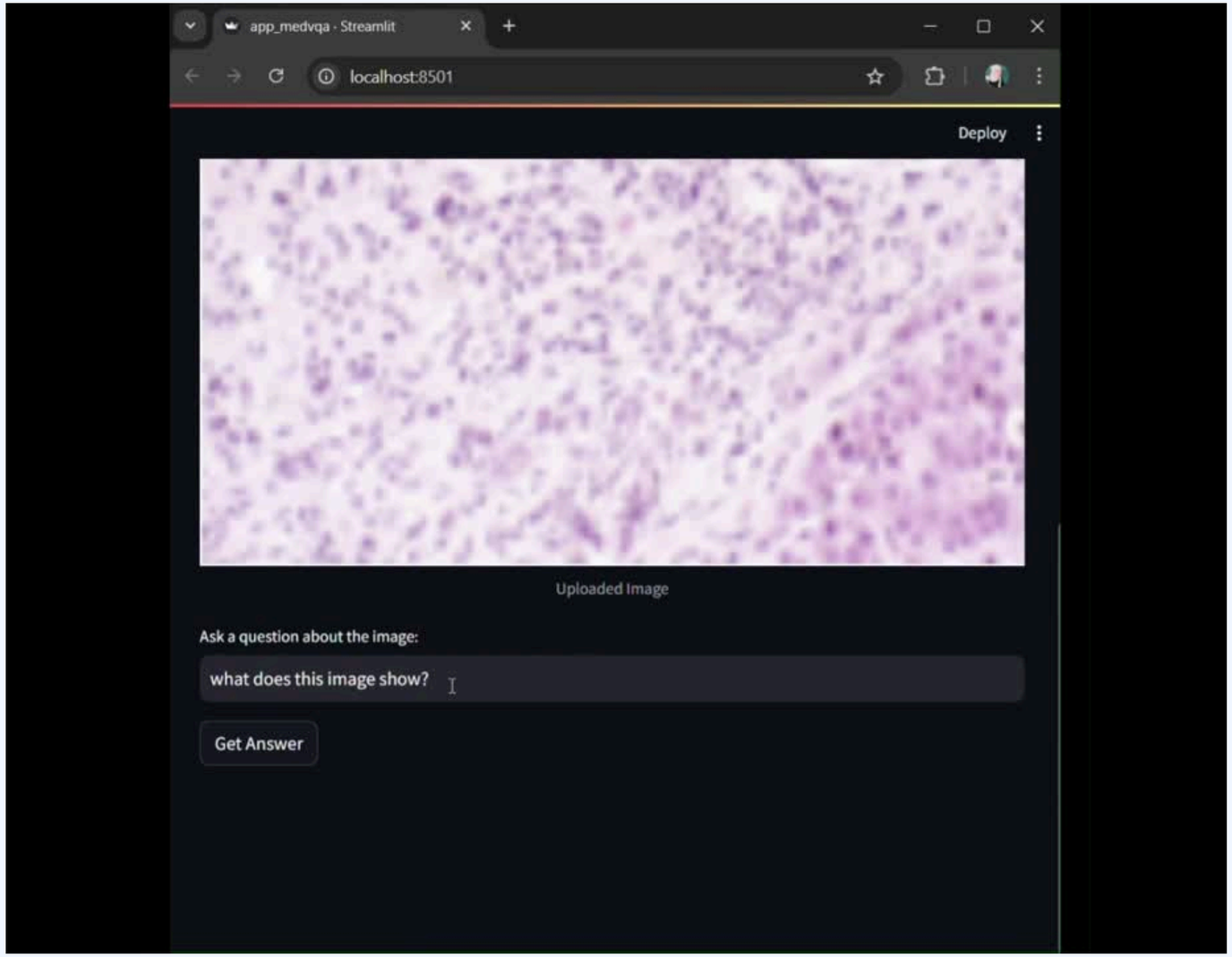
Original Paper:


BLIP: Bootstrapping Language-Image Pre-trainin
Unified Vision-Language Understanding and Generation

Junnan Li Dongxu Li Caiming Xiong Steven Hoi
Salesforce Research
https://github.com/salesforce/BLIP

Inference Examples

- Created streamlit interface



Obstacles Found

```
OutOfMemoryError: CUDA out of memory.
```

Despite employing common memory reduction techniques—including :

1. **smaller batch sizes**
2. **gradient accumulation**
3. **mixed precision (fp16)**
4. **model parallelism across two GPUs**

training remained challenging on free platforms like **Kaggle T4**, **Colab TPU**, and **AWS SageMaker Studio Lab**.

- This highlights the substantial memory demands of Multi-Modal model training.

Perspectives

- **More Data Collected**

Figure_path string	Question string	Answer string	Choice A string	Choice B string	Choice C string	Choice D string	Answer_label string
PMC1064097_F1.jpg	What is the uptake pattern in the breast?	Focal uptake pattern	A:Diffuse uptake pattern	B:Focal uptake pattern	C:No uptake pattern	D:Cannot determine from the information given	B
PMC1064097_F2.jpg	What radiological technique was used to…	Mammography	A: Mammography	B: CT Scan	C: MRI	D: X-ray	A
PMC1064097_F4.jpg	Where were the microcalcifications…	Behind the nipple	A:Behind the nipple	B:Above the nipple	C:Below the nipple	D:Around the nipple	A

- Converted the QCM based dataset called **PMC-VQA** into a simplified format consisting solely of question-answer pairs.
- Translated chinese data (questions & answers) in **Slake Dataset** to english

Figure_path string	Question string	Answer string
PMC1064097_F1.jpg	What is the uptake pattern in the breast?	Focal uptake pattern
PMC1064097_F2.jpg	What radiological technique was used to confirm the diagnosis?	Mammography
PMC1064097_F4.jpg	Where were the microcalcifications located in the mammography image?	Behind the nipple

Perspectives

1. RAG for Contextual Accuracy

- Ensures the model's responses are grounded in up-to-date medical context, improving diagnostic reliability.

2. LoRA Fine-Tuning for Efficient Adaptation

- Allows quick adaptation to specific medical domains (e.g., radiology), enhancing the model's accuracy in medical VQA tasks.

3. Using Med-PaLM for Enhanced Medical Context

- Med-PaLM is a Google-developed model fine-tuned on medical data, designed for healthcare-related NLP tasks.

References

- https://www.rcac.purdue.edu/files/training/NLP_101.pdf
- https://www.dataknobs.com/generativeai/10-llms/
- https://www.slideshare.net/slideshow/working-in-nlp-in-the-age-of-large-language-models/261151965#1
- https://arxiv.org/abs/2111.10056
- https://arxiv.org/pdf/2201.12086

Thank you for your attention