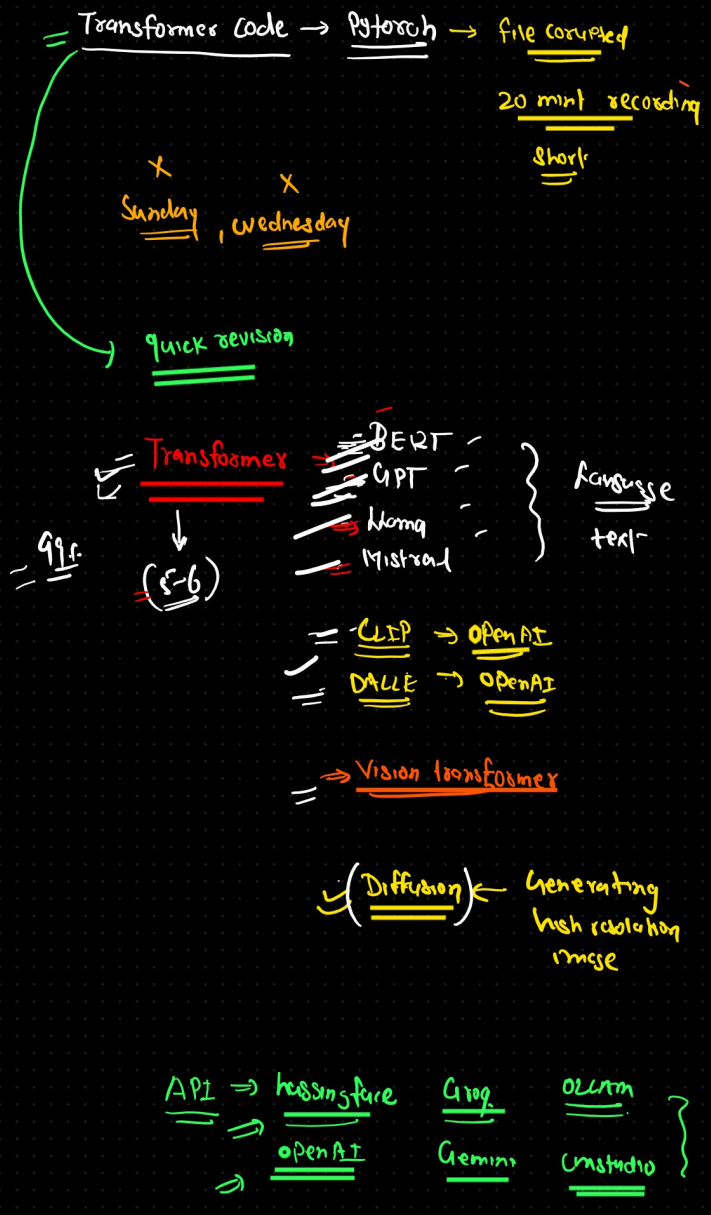
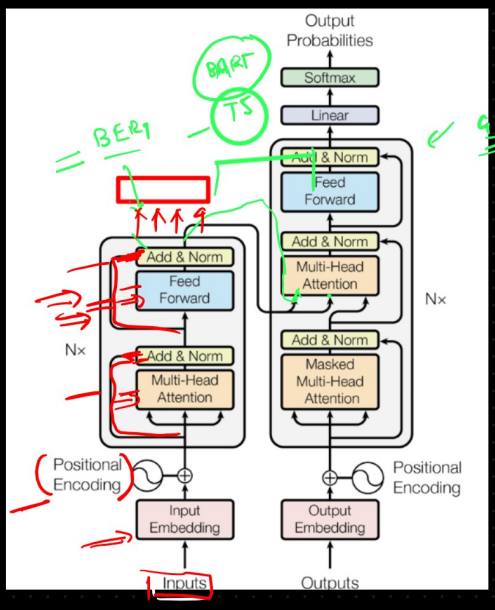
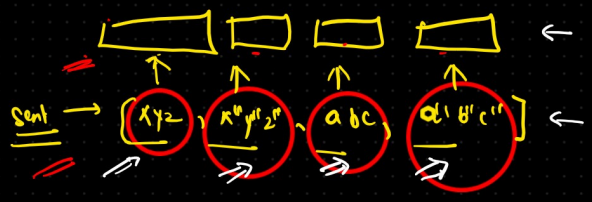

Transformer code → Pytorch → file corrupted
20 mint recording
short
X
Sunday
X
Wednesday
quick revision
Transformer
BERT
GPT
Llama
Mistral
Language
text
99%
(5-6)
CLIP → OpenAI
DALLE → OpenAI
→ Vision Transformer
Diffusion
Generating
high resolution
image
API ⇒ huggingface
Groq
Ollama
openAI
Gemini
Lmstudio





X
(CNN + BERT)
ViT

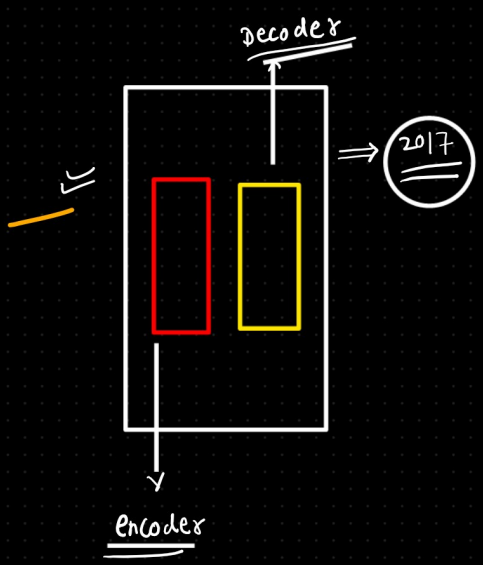
(GPT, LLama, Mistral)

(CLIP)

Image

(DaMe)

ViT



NLP ⇒ ULMFIT ← LSTM

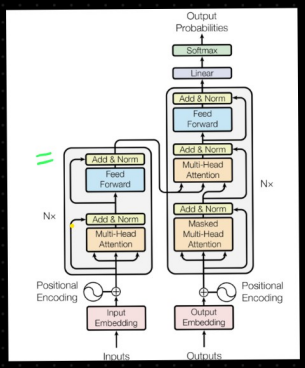
(Transfer Learning)

(2017-2018)

+ (1 Unsupervised (Pretraining)
2 Supervised (Fine tuning))

⇒ 3 RLHF → (OpenAI)
ChatGPT

Bidirectional encoder Representation from transformer



Encoder ⇒ RNN

Self attention

Self attention

Bi-direction I am Sunny Savita

Positional encoding

Self attention

layer norm, feed forward

BERT ⇒ 12 Layer ⇒ hidden layer size ⇒ 768 } FF
(110M) 24 layer ⇒ hidden layer size ⇒ 1024
340 M

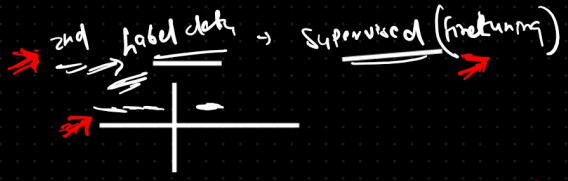
1 MASK Language modelling
2 Next sent pred

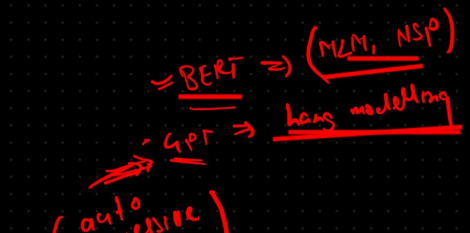
BERT
- 1 unsupervised Pretraining ← (base model) ← Wikipedia
- 2 Supervised Fine tuning → specific task
  - → text classification
  - → Summarization
- → Language modelling (generation)
- NER

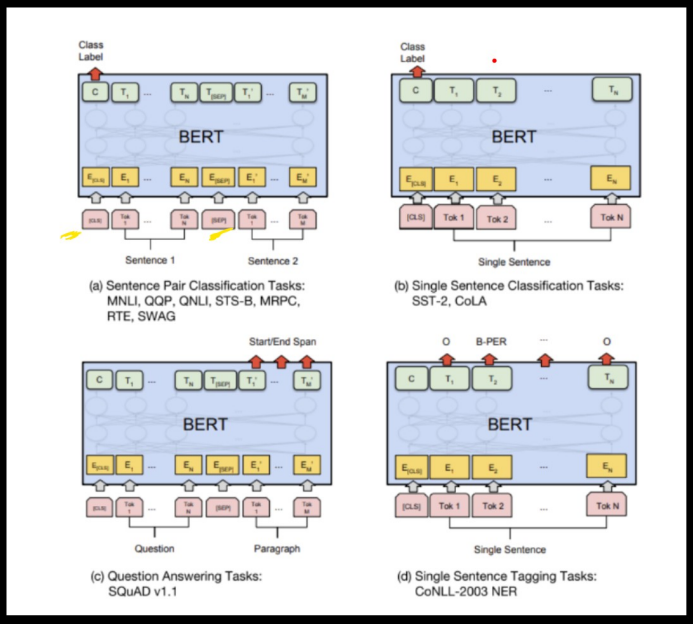
(Base) ⇒ specific

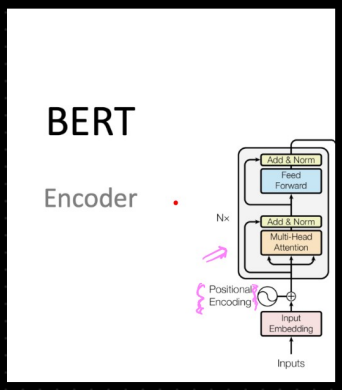
Data ⇒ Labelling (books, wiki, reddit) very huge data

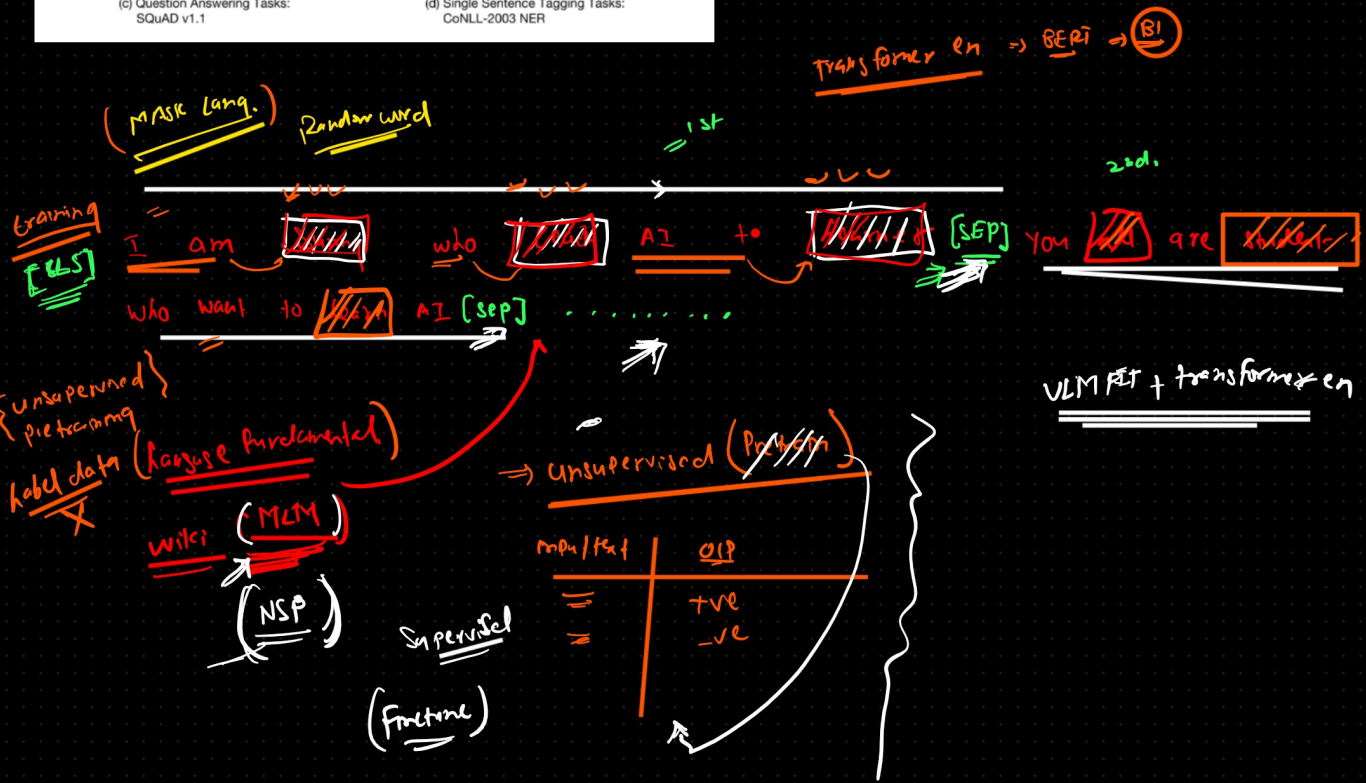
unsupervised (Pretraining) ⇒ BERT ⇒ (MLM, NSP)
1st
(text generation) GPT ⇒ language modelling ⇒ (next word pred)
(Llama, Mistral) ⇒ GPT decoder





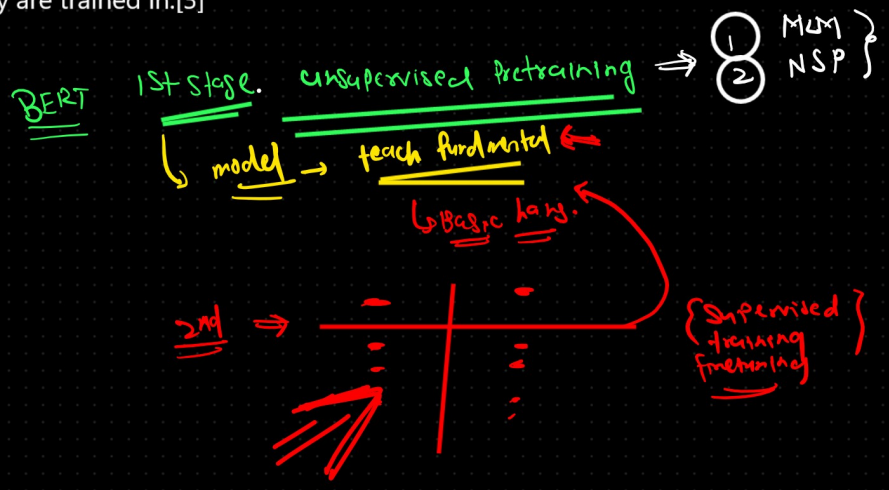

Class Label
BERT
Sentence 1
Sentence 2
(a) Sentence Pair Classification Tasks: MNLI, QQP, QNLI, STS-B, MRPC, RTE, SWAG
Class Label
BERT
Single Sentence
(b) Single Sentence Classification Tasks: SST-2, CoLA
Start/End Span
BERT
Question
Paragraph
(c) Question Answering Tasks: SQuAD v1.1
O B-PER O
BERT
Single Sentence
(d) Single Sentence Tagging Tasks: CoNLL-2003 NER


BERT
Encoder
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Positional Encoding
Input Embedding
Inputs

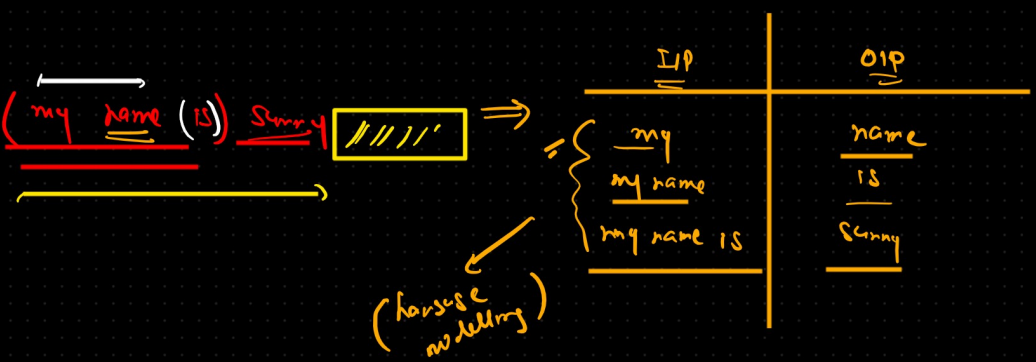


A large language model (LLM) is a type of machine learning model designed for natural language processing tasks such as language generation. LLMs are language models with many parameters, and are trained with self-supervised learning on a vast amount of text.

The largest and most capable LLMs are generative pretrained transformers (GPTs). Modern models can be fine-tuned for specific tasks or guided by prompt engineering.[1] These models acquire predictive power regarding syntax, semantics, and ontologies[2] inherent in human language corpora, but they also inherit inaccuracies and biases present in the data they are trained in.[3]



A large language model (LLM) is a type of machine learning model designed for natural language processing tasks such as language generation. LLMs are language models with many parameters, and are trained with self-supervised learning on a vast amount of text.



1. Bi-directional Attention => strong contextual understanding

Pretrain on massive Wiki data (MLM, NSP)

↳ Sentiment analysis, NER, Summarization

Encoder based architecture.

2. GPT => unidirectional attention

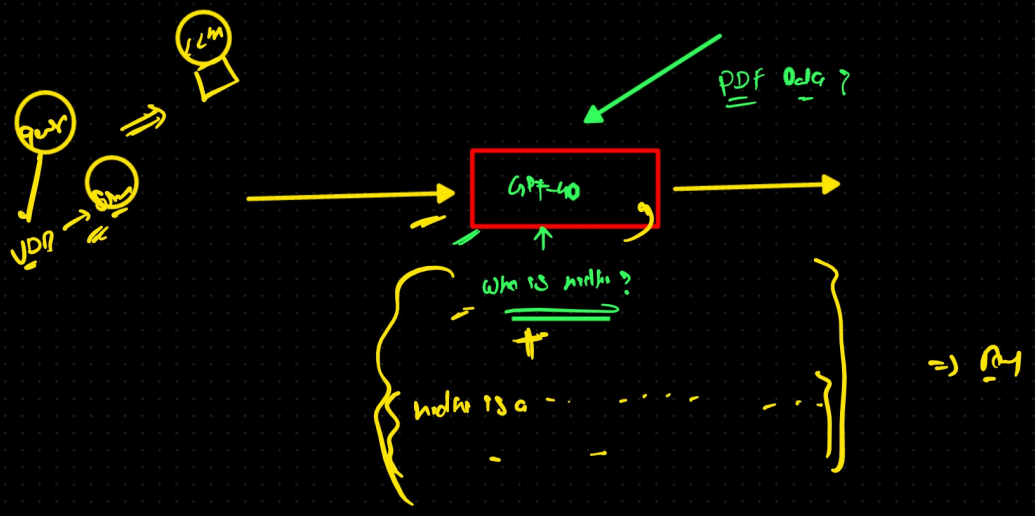
Autoregressive modelling => (Pretrained)

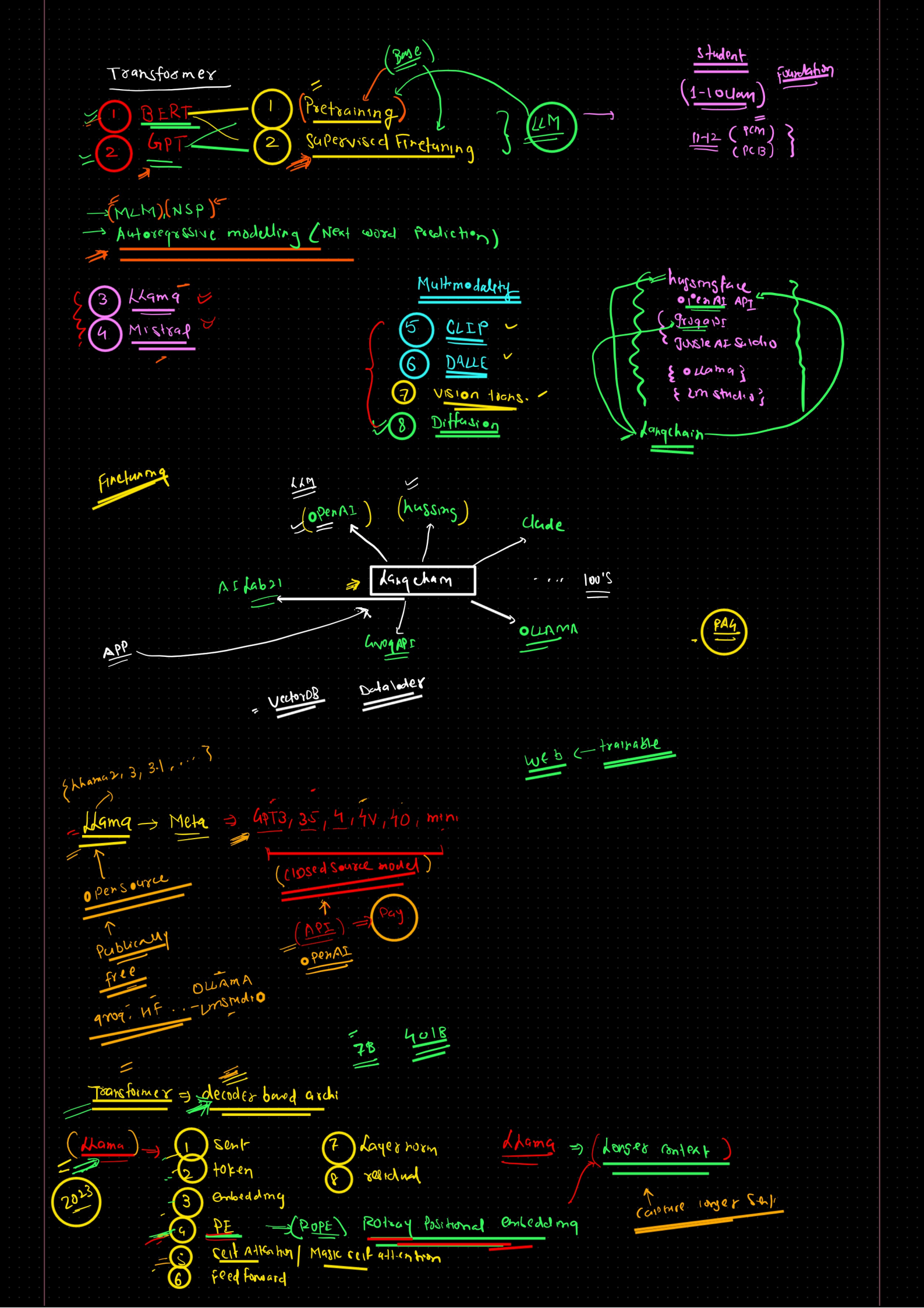
(next word pred)

decoder based archi ← transformer

massive (book, wiki, social platform) (reddit)

text generation => Gpt 3.5 :) Chatgpt




Transformer
1 BERT
2 GPT
1 Pretraining
2 Supervised Finetuning
Base
LLM
Student
(1-10Llam)
Foundation
11-12 (PCM) (PC13)
(MLM) (NSP)
Autoregressive modelling (Next word prediction)
3 Llama
4 Mistral
Multimodality
5 CLIP
6 DALLE
7 Vision trans.
8 Diffusion
huggingface
OpenAI API
Groq API
Google AI Studio
Ollama
LM studio
langchain
Finetuning
LLM
(OpenAI)
(hugging)
Claude
AI lab21
langchain
100's
APP
Groq API
OLLAMA
RAG
VectorDB
Dataloader
{Llama2, 3, 3.1, ...}
Llama → Meta
GPT3, 3.5, 4, 4V, 4o, mini
(closed source model)
Web ← trainable
Open source
Publically free
groq, HF ... OLLAMA LMstudio
API
Pay
OpenAI
7B
40B
Transformer ⇒ decoder based archi
(Llama)
2023
1 Sent
2 token
3 Embedding
4 PE
5
6 feed forward
(ROPE) Rotary Positional embedding
Self Attention / Magic self attention
7 Layer norm
8 residual
Llama ⇒ (longer context)
Capture longer seq

Supervised

dialog:

question	answer

mistral AI

Mistral ⇒ French

↳ 7B, 70B

↳ decoder based model

↳ efficient computational perf

Benchmark

reduce redundancy

clustering

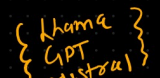
FFNN

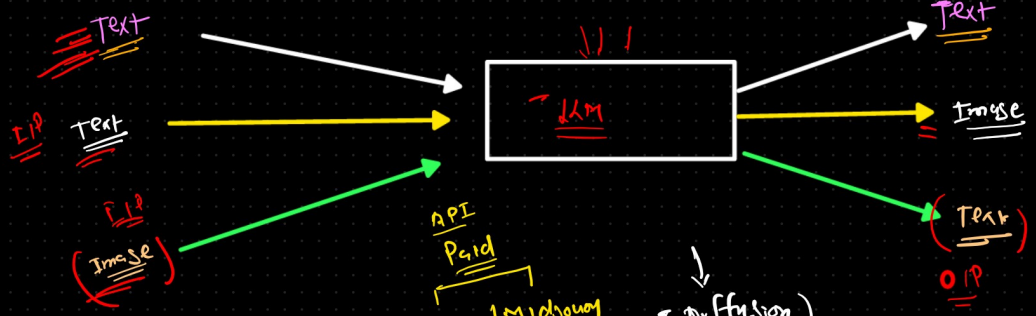
Masked self attention

(grouped) query attention

Sliding window attention

{long seq.}





API

Paid

Dalle / Midjourney

Diffusion

Text

Bird are flying

Image

Image

Probability

(Moving, Pendings)

= CNN + LSTM ⇒ 2014-2015

CNN + BERT ⇒ 2017-2018, 19

(MM)

Open AI

CLIP

VIT

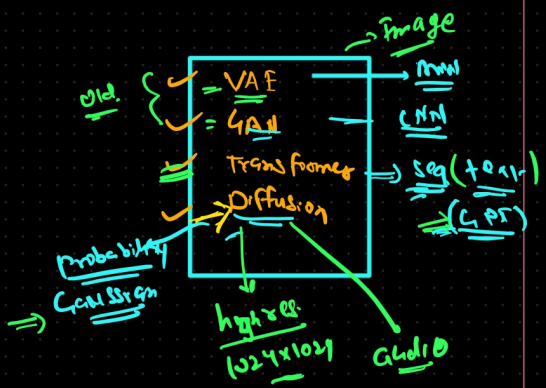
Image

text

Birds are flying

OCR ⇒ PDF

Chatbot ⇒ (text)



CLIP ⇒ Transformer ⇒ (encoder)

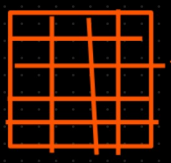
dual encoder

⇒ first encoder ⇒ Image

second encoder ⇒ text

One person standing

black white

 ⇒ Pixel ⇒ Numbers (0-255)

↳ Flatten Image ⇒ (1D array) ⇒ transformer encoder ⇒ contextual embedding

Text ⇒ embedding ⇒ Transformer encoder ⇒ contextual embedding

CLIP

Image

caption

Data (training) Transformer

⇒ Dual encoder

→ contextual embedding text

→ contextual embeddings image

(cosine Similarity)

+ve Similar

-ve Dissimilar

Testing

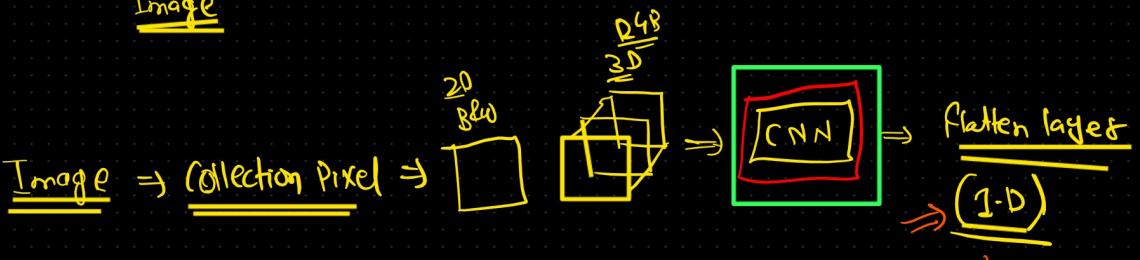
1. hundreds of millions Image-text pair
2. Dual encoder setup
   - Image encoder
   - text encoder
3. Similarity
   - Cosine Similarity
4. Zero-shot capabilities

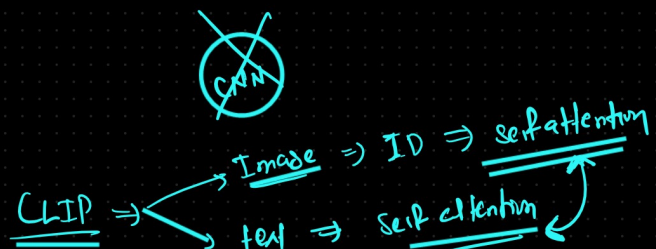
 → Description

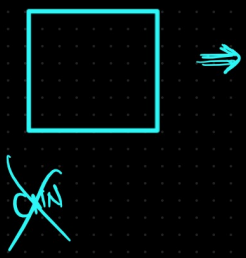
Vision and lang transformers (ViLT)

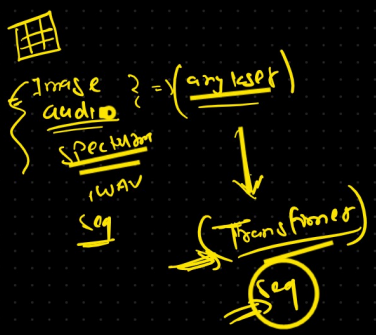
→ transformer based model ⇒ Description

Image

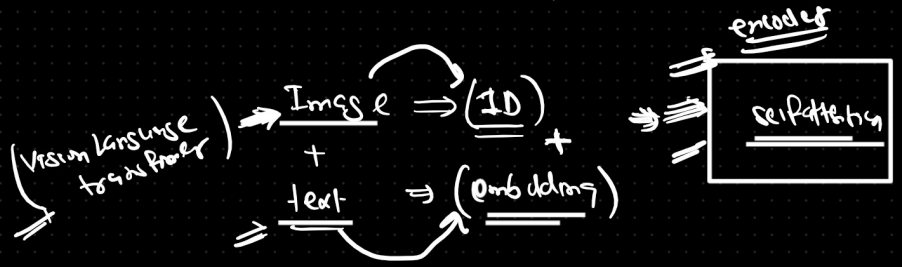








Vision transformation ⇒ ~~Pixel~~ Single encoder



Dall-E

Text ⟶ Image generation

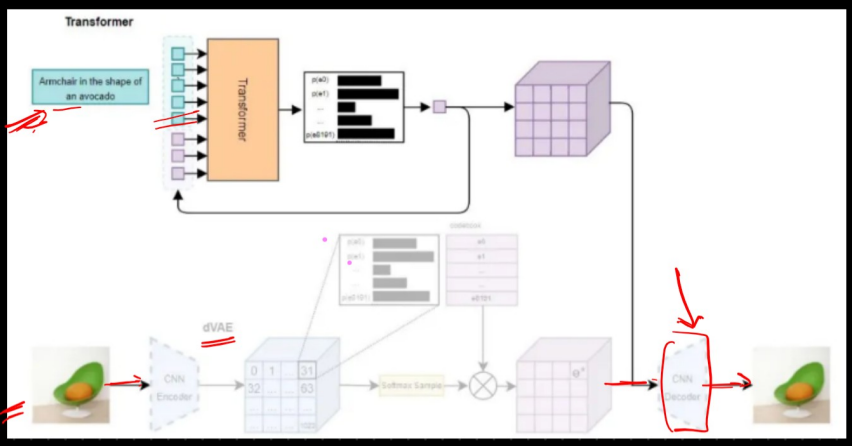
= GAN
= dVAE
⇒ Clerical ?

Transformer

DALLE
- Patch Image
- = Pixel
- = GPT
⇒ 128×128, 512×512

DALLE-2 ⟶ Diffusion
high resolution ⇒ HD ⇒ 1024×1024

Two Stage Process



training

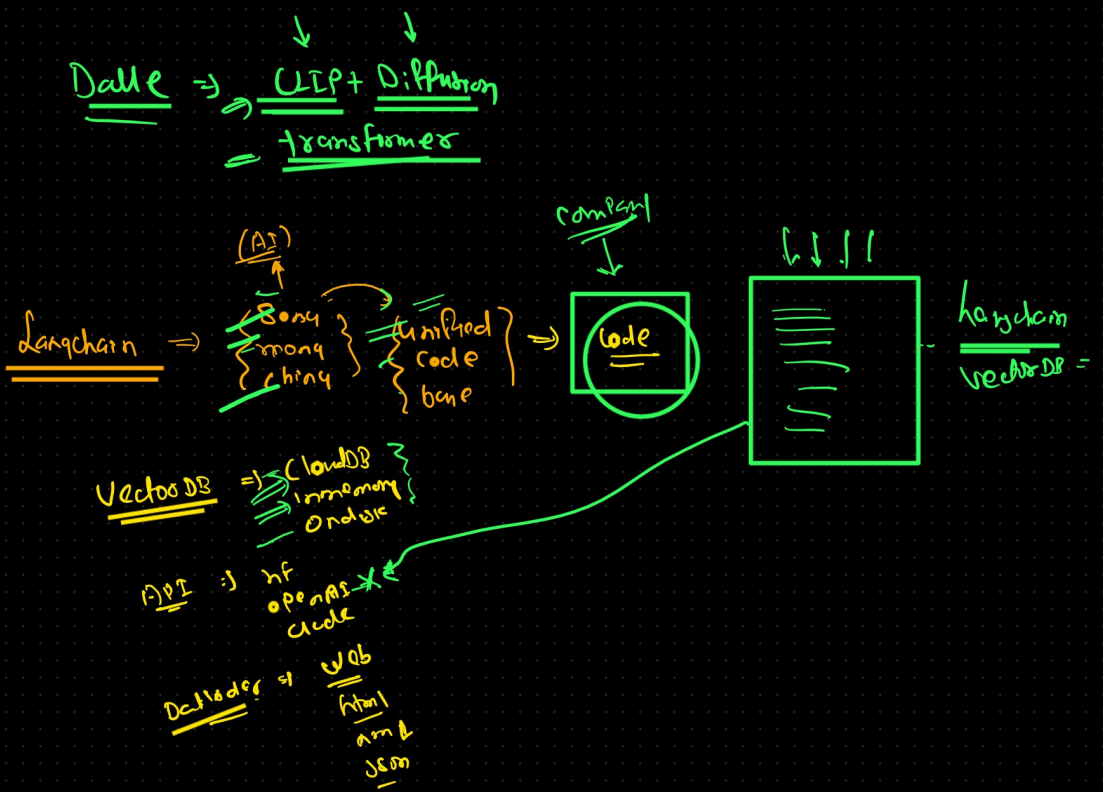
VAE (CNN)

2021
Dalle 1 ⇒ GPT like transformer arch.

Dalle 2 ⇒ CLIP + Diffusion
2022
↓
Limitation

training
Image
text
VIT
A
text & Img
relationship
encoder
CLIP
text & Image
1st
VAE

B
Image Gen
- autoregressive ⇒ GPT ⇒ Image Patching
- Diffusion ⇒ nd, hD


Dalle
CLIP + Diffusion
= transformer
(AI)
Langchain
Song
among
thing
unified
code
base
company
code
Langchain
VectorDB =
VectorDB
Cloud DB
in memory
Ondisk
API
hf
openAPI
claude
Doc loader
web
html
xml
json