

Early history

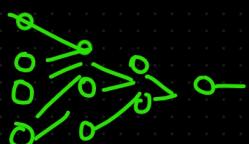
Medieval Apps

modern AI



## History of AI

ANN



CNN

Convolution + ANN

Generative Model

1

Img to Img  $\rightarrow$  (GAN, Autoencoder)

2

Text to Text  $\Rightarrow$  LLM, LM

3

Img to Text

4

Text to Img

RNN

ANN + Feed loop

BIG

GenAI  $\Rightarrow$  LLM

CNN + RNN (LSTM, GRU)

Diffusion, transformer

## Language Modelling

CNN  
efficient  
real  
dense

Computer vision  
Yolo, SSD, Faster RNN

Image classification

object Detection

obj Segmentation

.. tracking

OCR

## Transfer Learning

Functioning

Requirements

Previous task and implementing  
by the team.

teenage

cycle

① balance

② traffic

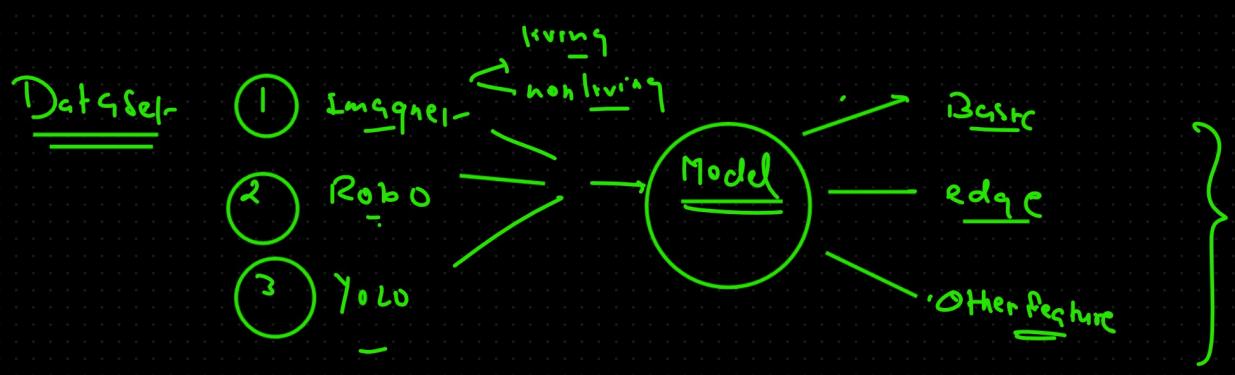
③ double riding

adult

motorcycle

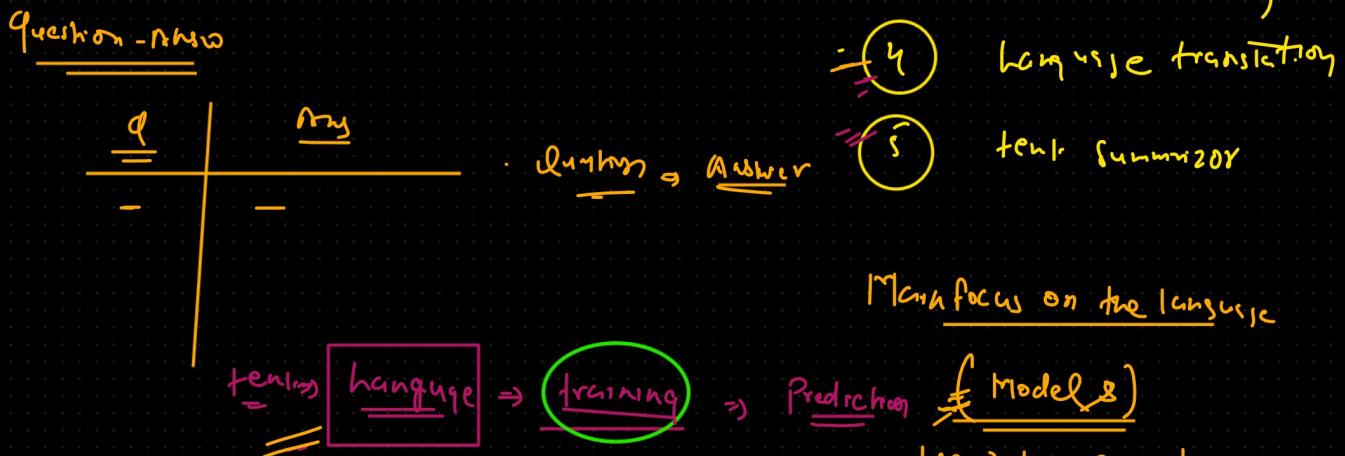
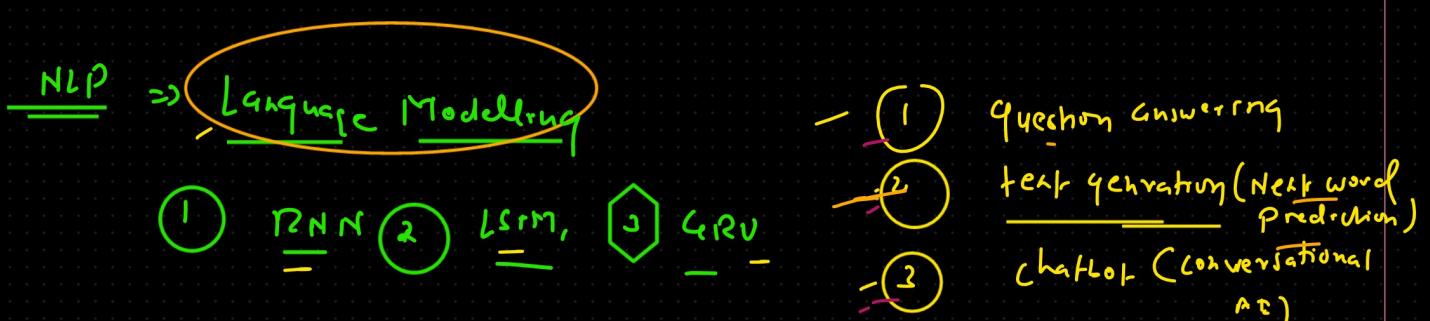
↑↑↑

Territory  
Accelerator



If was supervised learning same model  $\Rightarrow \{ + \text{FOD}$  Determin  $\}$   $\Rightarrow$  Do it alone

Generative  $\Rightarrow$  GAN  $\Rightarrow$  Image  
 $\rightarrow$  Autoencoder, Variation



Supervised Language modelling  
Unsupervised Language Modelling  $\Rightarrow$  which

How we are teaching  
 What will be the best task for the model

<u>1987</u>	<u>1997</u>	<u>2007</u>	<u>2014</u>	<u>2016</u>	<u>2017</u>
<u>RNN</u>	<u>LSTM</u>	<u>GRU</u>	<u>Encoder-Decoder</u>	<u>Attention</u>	<u>Transformer</u>

2018  
Bert

2019  
CPT

1000's of Model

## Mapping



One to one mapping



one to many



many to one



many to many (synchronized & asynchronized)

same  
length

Diff  
length

## Show the image

### Example of One to one mapping

1

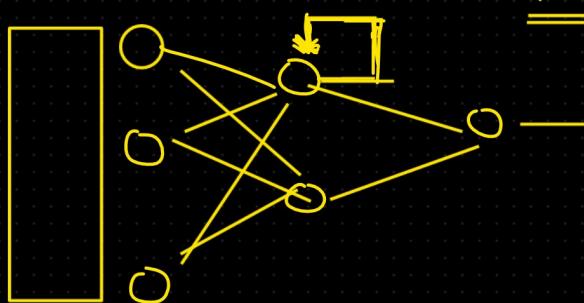
RNN

Picture



Example

sentence



RNN  $\Rightarrow$  time stamp

Problem  $\Rightarrow$  longer sentence [sustain] [vanishing gradient | exploding gradient]

Solution  $\Rightarrow$  LSTM

Memory  $\rightarrow$  Long (short) term memory [Long sentence]

Small  $\uparrow$  very  $\downarrow$   
(gradation) (conversion)

Sunny, decoder

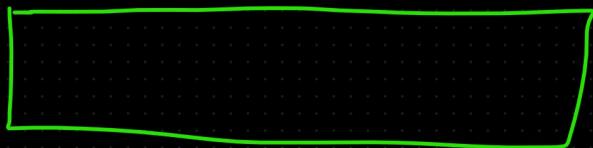
Optimized

Dandeep was the greatest teacher and Cleopatra now has -  
great tutor Dandeep  
Sunny

Memory unit

Architecture

of LSTM



Query  $\Rightarrow$  {Update  
Add  
Remove}

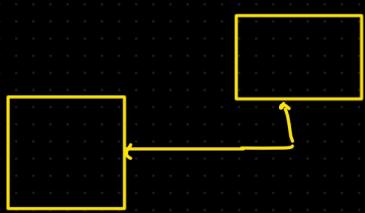
GRU  $\Rightarrow$  Architecture

Diff GRU | LSTM

4  
 Encoder Decoder & Encoder-decoder with attention  
 $\Rightarrow$  {Asynchronous task}  
 Translation(M to M)

Architecture of E&D

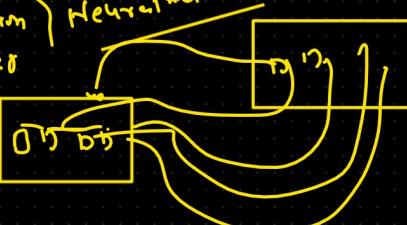
$\Rightarrow$  E&D with Attention



{Context  
Content  
3 words  
computation}

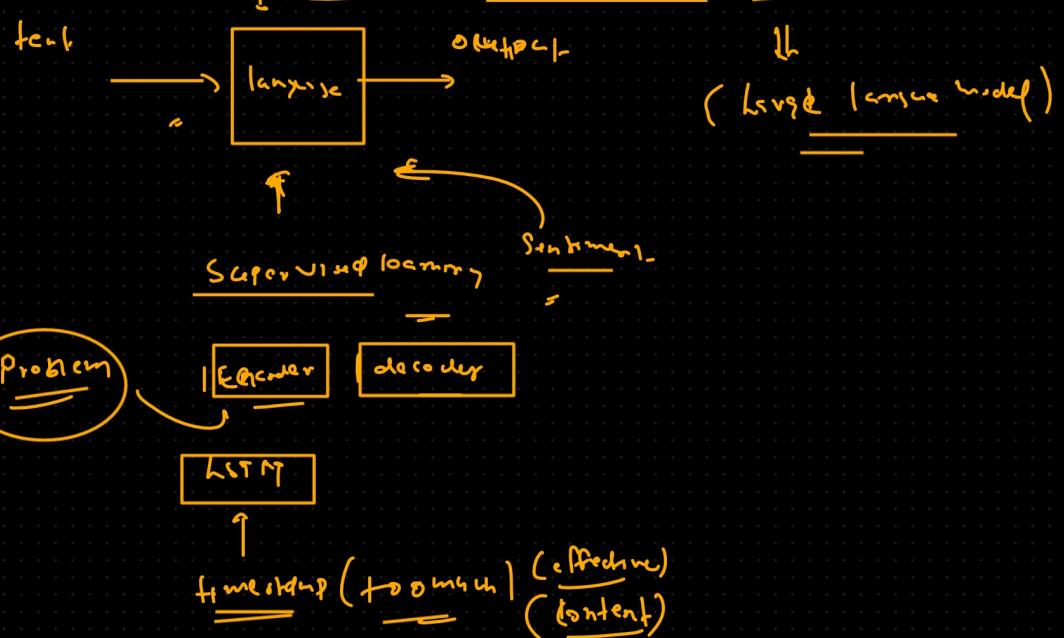
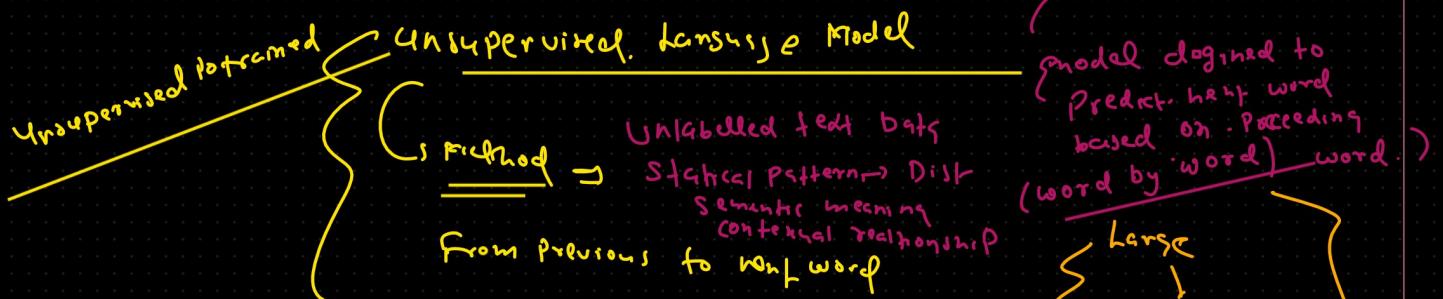
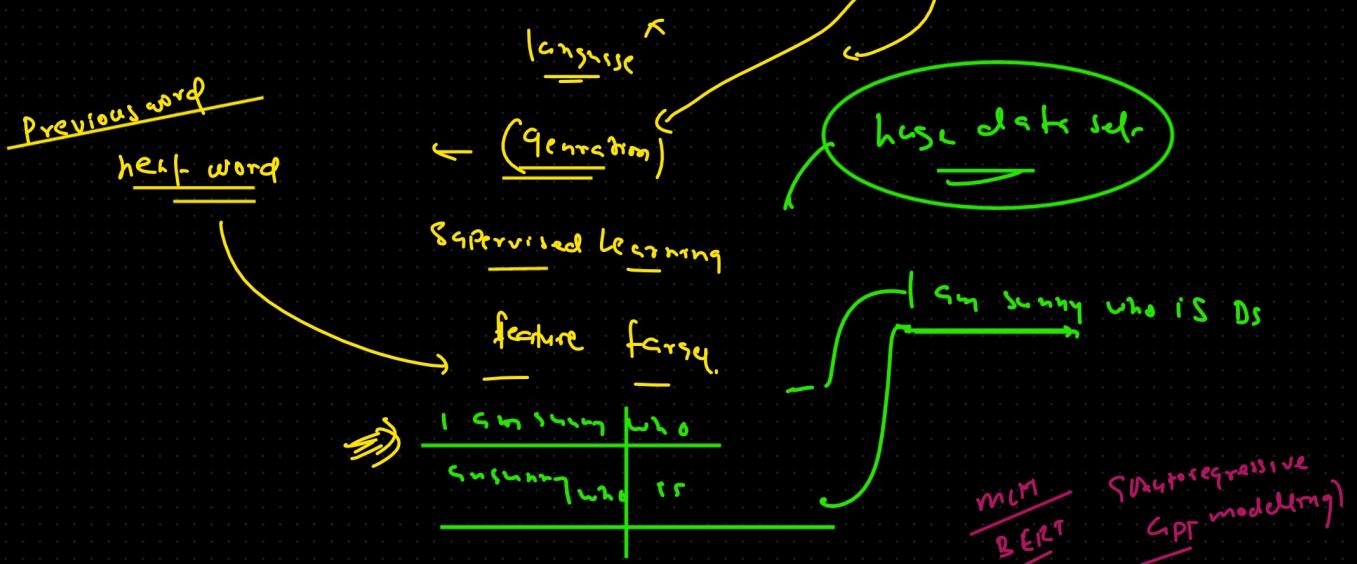
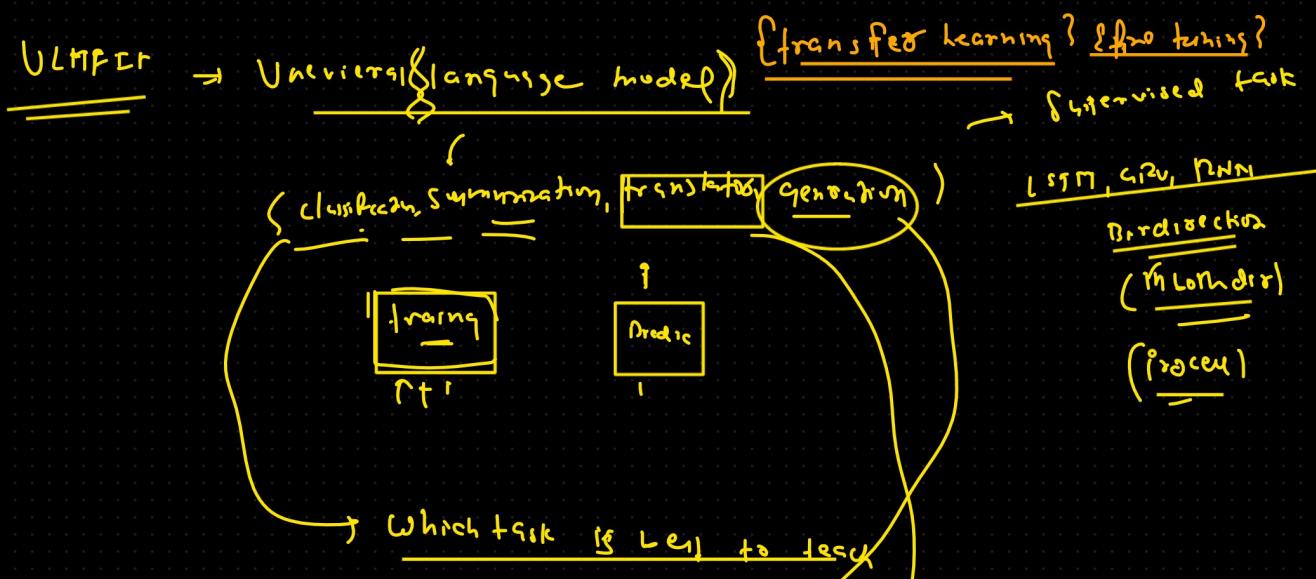
Special layer on top

{Attention} Neural network  
layer



{Each & every having everything}  
{Every hidden state}

Attention Picture



Transformer

Architecture of transformer

Encoder - Decoder

Attention

Self Attention

- 1 Remove LSTM-RNN
  - 2 USE self Attention →  
Positional encoding  
Parallel processing
  - 3 Normalization
  - 4 Attention
  - 5 Skip connection
- multihed attention  
(key, query, value)  
↳  $k, q, v \rightarrow$  weight matrix

- State of Art Performance

- Architecture = Universal

↳ Language  
↳ vision

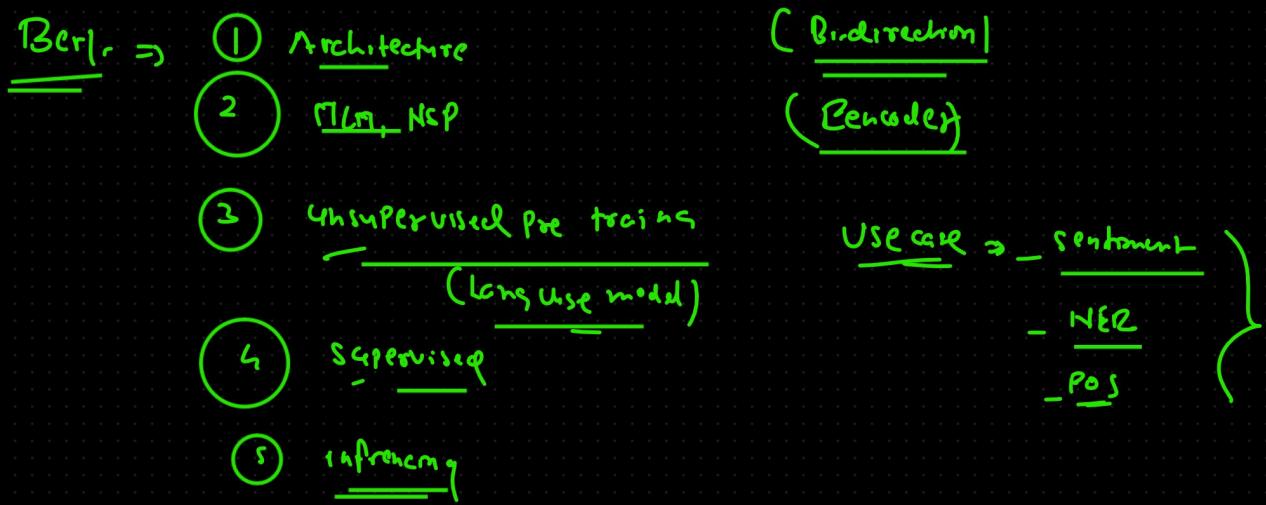
Write couple of Name here of Transformer

Transformer ⇒ Bert, GPT, BART, T5  
= = =

Nano ⇒ Transformer V

Diffusion model, multimodel ⇒

Research Paper ⇒



GPT  $\Rightarrow$  Generation (decoder)

Auto-regressive (LM)

GPT1, GPT2, GPT3, GPTV

---



---

GPT1: GPT: Improving Language Understanding by Generative Pre-Training

GPT2: GPT2: language models are unsupervised multitask learner

GPT3: GPT3: language model are fewshot learner

The term "generative" refers to the model's capability to generate coherent and contextually relevant text. "Pre-training" indicates that the model is initially trained on a large corpus of text data in an unsupervised manner before being fine-tuned for specific tasks.

During the pretraining phase, the model learns to predict the next word in a sequence, capturing language patterns and semantics.

In the context of the GPT-2 (Generative Pre-trained Transformer 2) paper, the term "unsupervised multitask learner" refers to the model's ability to perform a variety of natural language processing (NLP) tasks without task-specific supervision during the pretraining phase. GPT-2 is designed as a large-scale language model that is pretrained on a diverse corpus of

In the context of the GPT-3 research paper, the term "few-shot learner" refers to the model's ability to perform a task or answer questions with only a few examples or prompts provided during inference. GPT-3 is known for its remarkable ability to generalise from a small number of examples or demonstrations given to it at runtime.

just tell the size

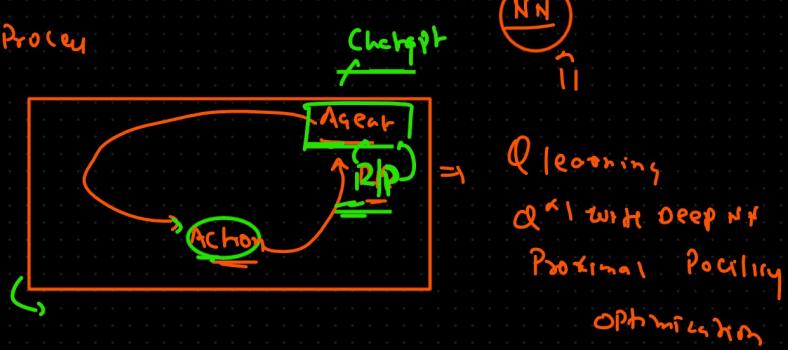
Context window size

fine tuning

please date

## Chatgpt training

- 1 Show the image
- 2 Explain the training process
- 3 RHFL  
↳ Q2



## Implacks

- 1 Ethical
- 2 Maintaining memory context
- 3 Dialogue specific (conversation)
- 4 RHFL (iterative feedback)
- 5 Brace error → minimization

## Generative NL

