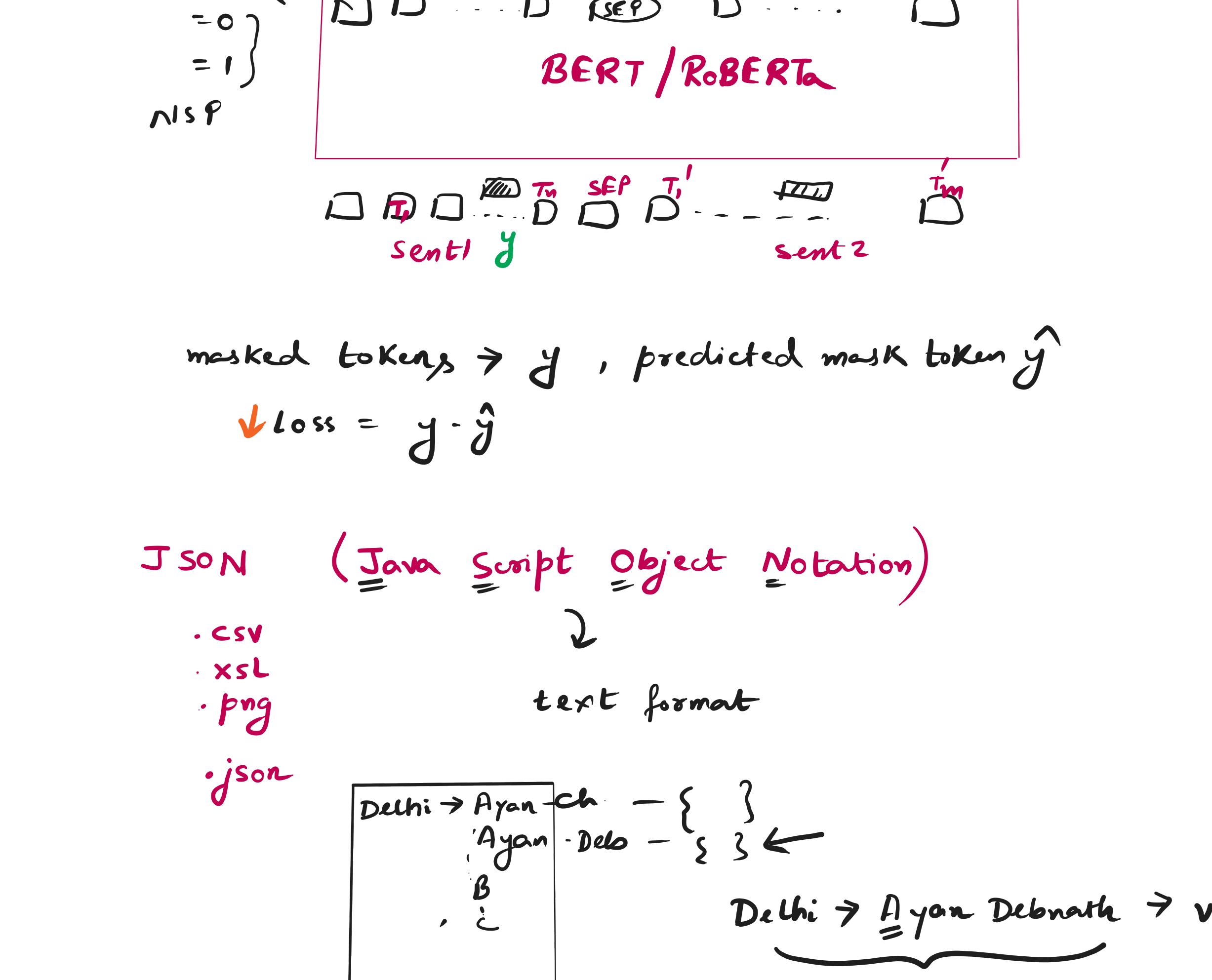


Day 20, 27<sup>th</sup> Jan 2024

NLP - DSM 2.0 - by mentor Dr. Ayan Debnath,  
IIT Delhi + Harvard University Alumni

YouTube:  
[https://www.youtube.com/@ad\\_academy](https://www.youtube.com/@ad_academy)

LinkedIn:  
<https://www.linkedin.com/in/ayan-debnath/>



masked tokens  $\rightarrow \hat{y}$ , predicted mask token  $\hat{y}$

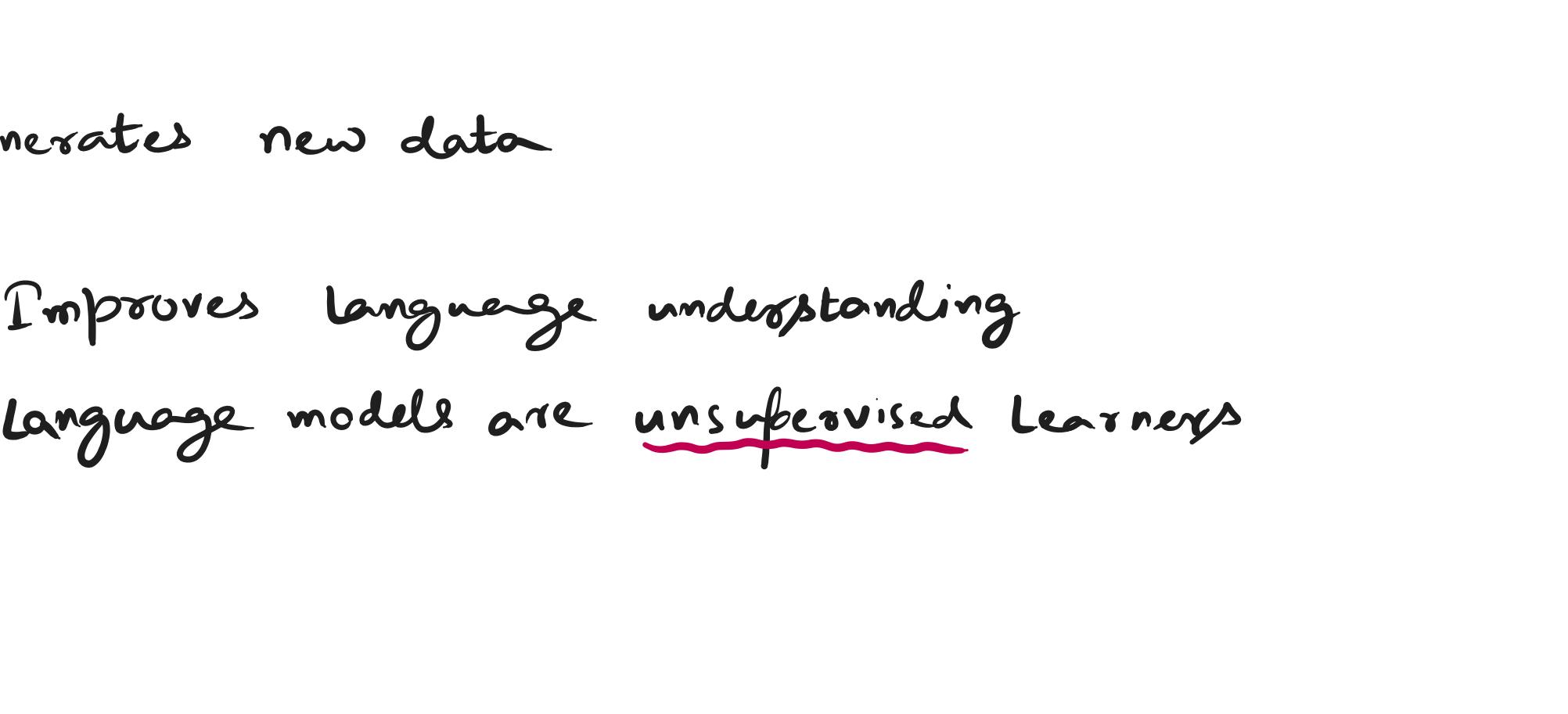
$$\text{Loss} = y \cdot \hat{y}$$

JSON (Java Script Object Notation)

.csv .xsl .png

.json

text format



→ Dictionary, data is stored in the format of key - value pair

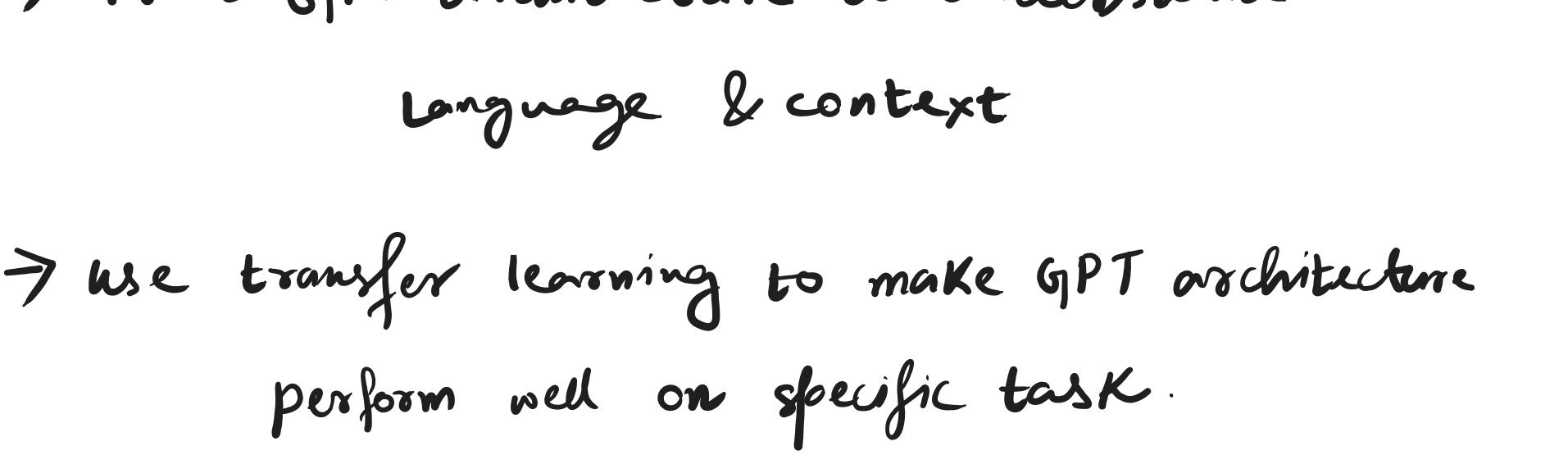
Q & A



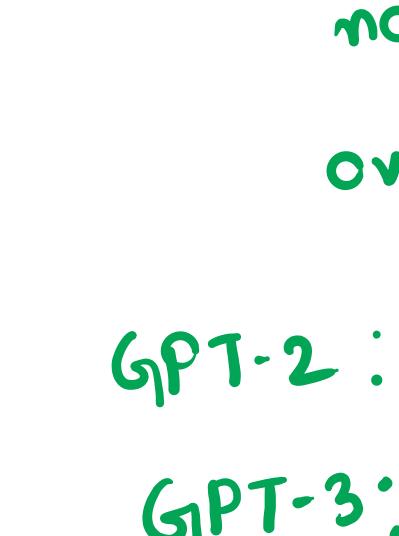
Q → ---

A → ---

passage → ---



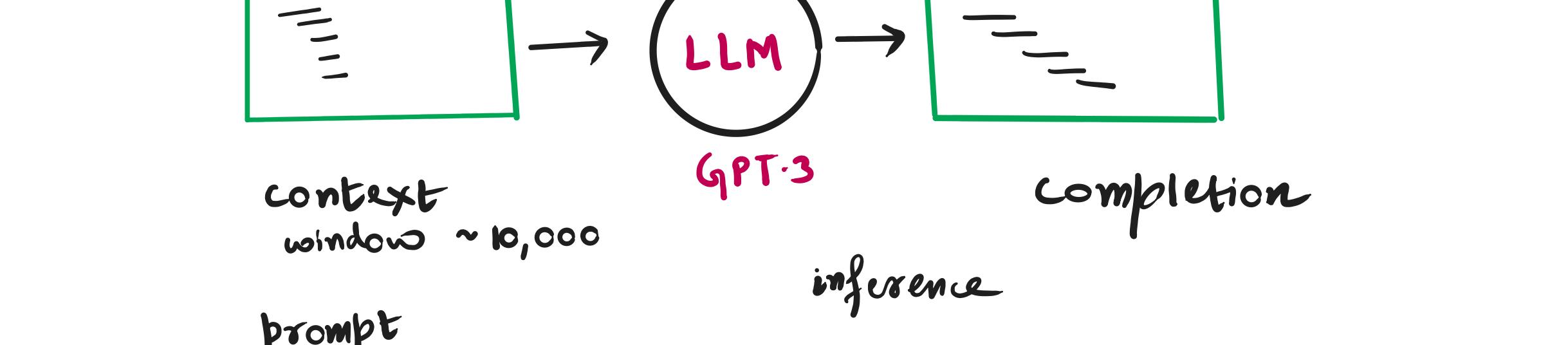
GPT (Generative Pre-trained model)



Generates new data

↳ Improves language understanding

language models are unsupervised learners



hyper-parameters

Q) How can you make LLMs with less data?

Transfer learning

Pre-training → Train GPT architecture to understand language & context

Fine-tuning → use transfer learning to make GPT architecture perform well on specific task.

Tissues with fine tuning

no. of trainable parameters ↑↑

overfitting happens easily

GPT-2: 1.5 billions parameter

GPT-3: 175 billions parameter

Prompt Engineering



context window ~10,000

inference

completion

prompt

In-Context Learning

① Zero-shot learning

Task: Classify the restaurant

LLM

---

e.g.: a=2  
Ex-1 b=3  
c=a+b

eg →  $2x+3x^2 = 5x^3$

② One shot learning

Task: Classify the restaurant

Ex-1: The food is good

SA: +ve review

SA: +ve review

Q: The food is not tasty

Q: The food is not tasty

A: -ve review

③ Few shot learning

Task: —

Ex-1: The movie is fantastic

SA: +ve

Ex-2: The food is not good

SA: -ve

Q: The chair is not working

Task: —

Ex-1: The movie is fantastic

SA: +ve

Ex-2: The food is not good

SA: -ve

Q: The chair is not working

A: -ve

④ Interview : ① Stemming & Lemmatization

↳ what?

( diff? )

↳ applies / examples

② Data Processing

↳ coding → FAANG

data cleaning

lowercasing

data reading

③ Embedding → TF-IDF

( BOW )

why better

④ RNN

Nvidia

⑤ LSTM

VGP ★★

weight update, chain rule

Activation fn

Gradient Descent

Loss fn

advance

⑥ Transformer

BERT / GPT / LLM → fine tuning

https://youtu.be/Pi6c8bZFXII?si=mSSoxLL32pqe7VTD

prompt engineering