Transformer

Encoder → **BERT**

Encoder and Decoder — Big Bird, BART, T5, M2M-100

Decoder (Text generation) → **GPT**

BERT

Masked Language modelling (MLM)

I went to shop for [MASK]

I am [MASK] so I bought food

Next sentence Prediction (NSP)

I am hungry. [   ]

I went to restaurant. [   ]

DistillBERT

Light weight

meant to deploy in production Environment

- 60% faster then BERT
- 40% lesser memory then Bert

97% retain Puformance of Bert

GPT

Generative Pretrained Transformer

Predect the next word from previous word
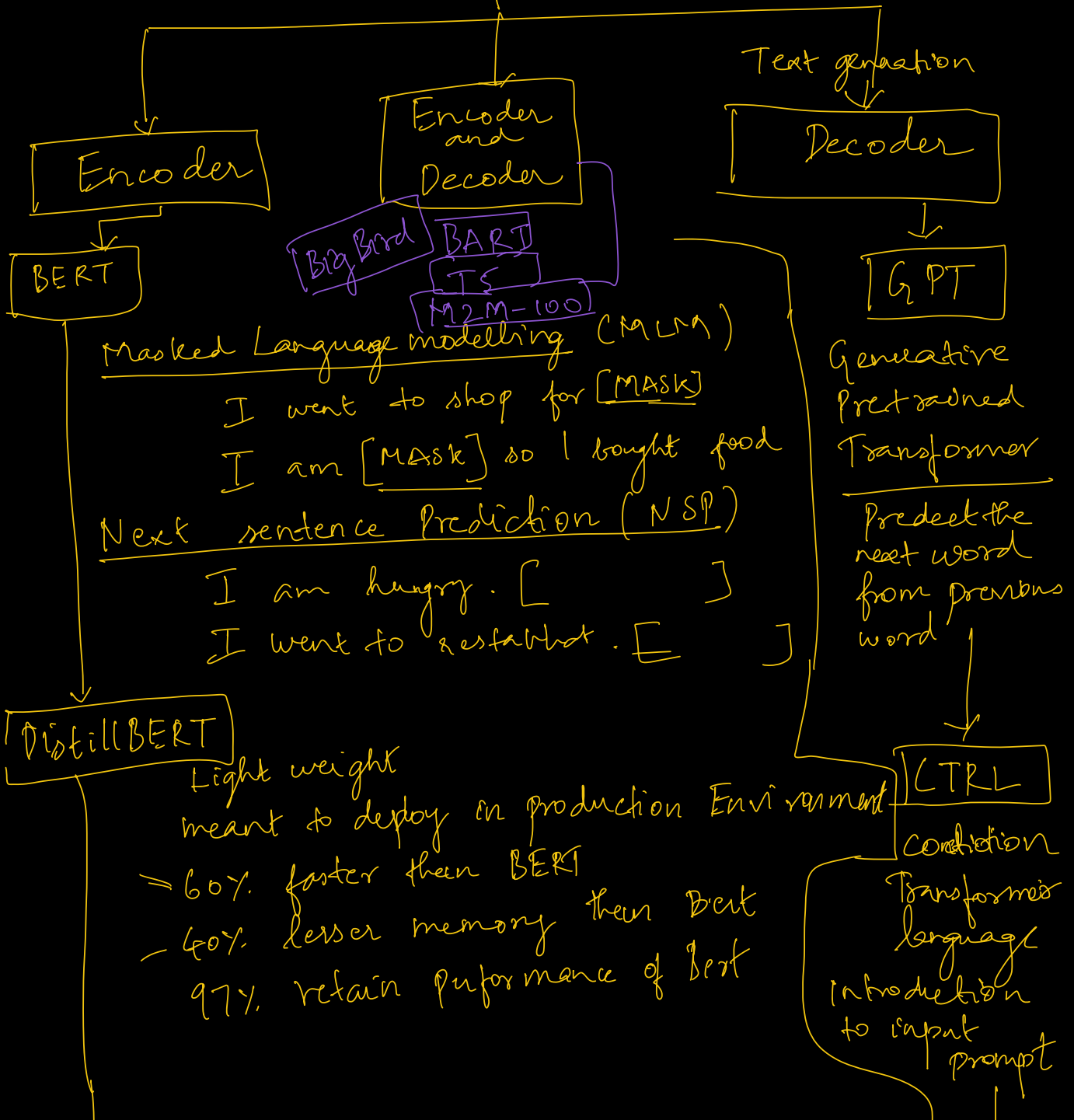
CTRL

Condition Transformer language introduction to input prompt

ROBERTa – Robustly Optimized Bert
approach
Trained longer Sentences
Droped the NSP

GPT-2
long datset
acurate
coherent
text

XLM ⟶ XLM - RoBerta

GPT-3
100 times of
GPT-2

AIBERT
A life Bert for
Self supervised language
representation.

175 billion
parameter
Few short learning
Not opensouree

ELECTRA

GPT Neo

GPT - J

GPT 3.5

DeBERTa

GPT 4 ⟶ Dumb

GPT 5 ⟶