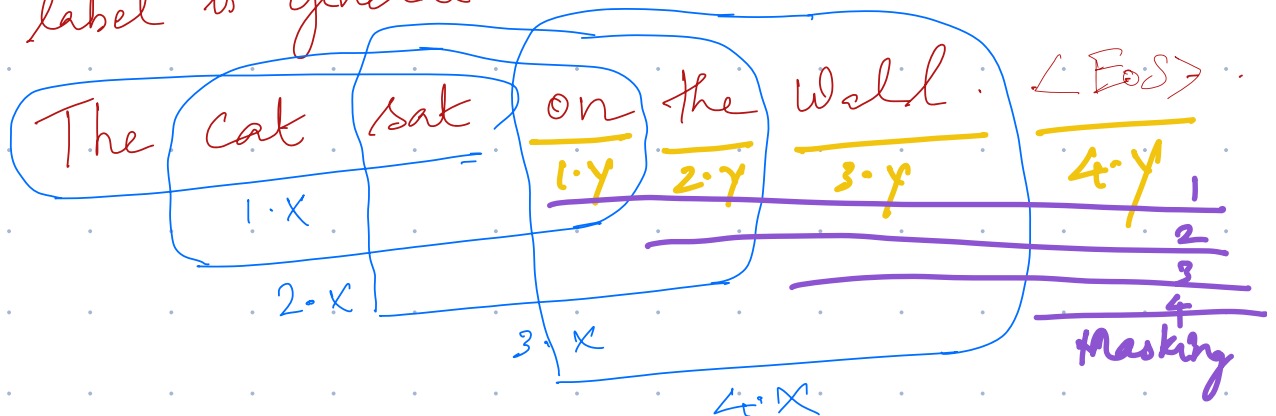GPT

BookCorpus Dataset — 7000 Books — ~6GB

1. Unsupervised - PreTraining

2. Supervised - FineTuning

1. Unsupervised PreTraining

↗ without manual Human Intervention
label is generated

The cat sat on the Well. <EoS>
$1 \cdot X$ $1 \cdot Y$ $2 \cdot Y$ $3 \cdot Y$ $4 \cdot Y$
$2 \cdot X$ 1
$3 \cdot X$ 2
$4 \cdot X$ 3
                                                      4
                                              Masking

Programmatically we are doing

2. Supervised Fine tuning

Universal language model Fine tune text (ULMFit)

1. Text classification
2. Text summarization
3. Text Comprehension
4. Choosing Right Answer

1. Text classification :

Task: Can you classify the following?

Content: The movie is fun to watch X

Classification : Positive Y

2. Text summarization

Task: Can you summarize the Para

Context: We started Serny NLP from
basics of prpr, em -- fo -- X

And then we learned LSTM, Transform, Decod, GPT and we are now knowng how the Gen AI work

Answer: We started learned NLP
Y_1 basics Y_3 to GPT for Gen AI Y_4
Y_5

Task: Choose the best answer?

Context: What is capital of India?

Str
a. London
b. Delhi
c. Mumbai
d. New york

X

answer b.) delhi
Y_2
Y_1

Parameters **117 Millions**

$\leftarrow$ weights and bias

Atte Keys, Query, Value, FeedForw W
liters

Preprocessing. BPE Tokenizer