BERT Pre-training: Two Main Tasks

① Masked Language Model
MLM - 15% masking

Input:

[CLS] The cat [MASK] on the mat

↓
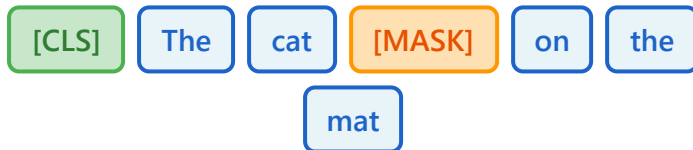
Predict: "sat"

Uses **bidirectional context**
to predict masked tokens

② Next Sentence Prediction
NSP - Binary classification

Input:

A: The man went to the store.

B: He bought a gallon of milk.

↓

IsNext ✓

[CLS] token for
sentence-level representation

📝 **Tokenization**

WordPiece
30K vocabulary

🏷️ **Special Tokens**

[CLS] [SEP] [MASK] [PAD]

⚡ **Training**

4 days
on 4-16 Cloud TPUs