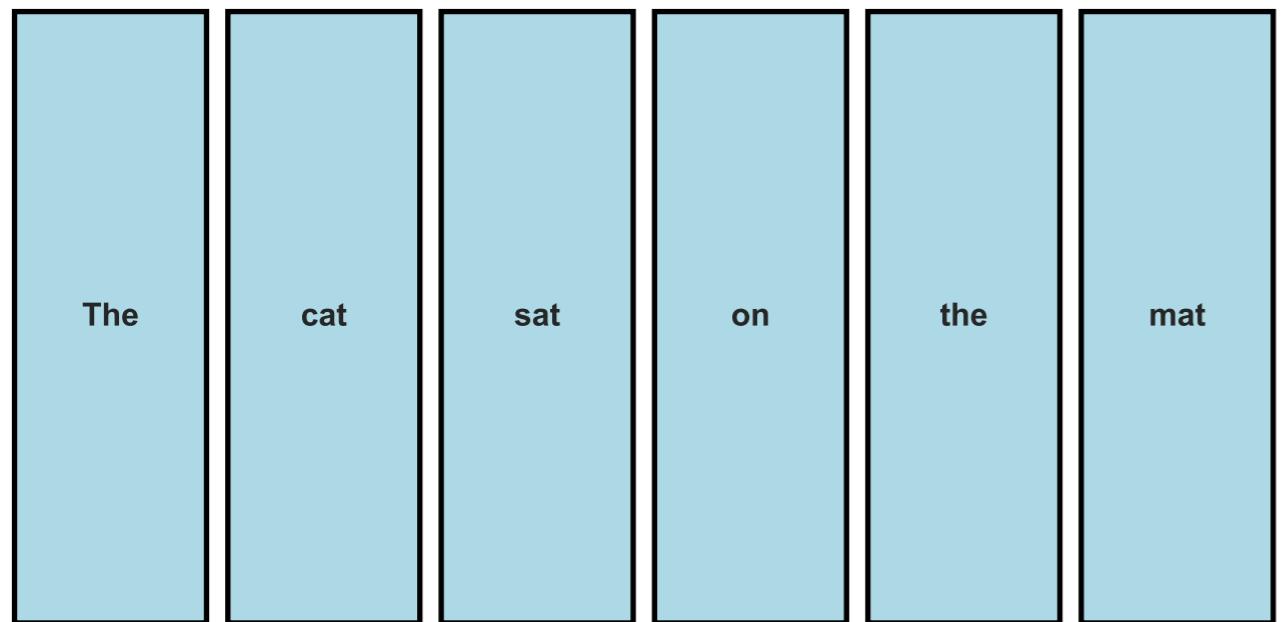
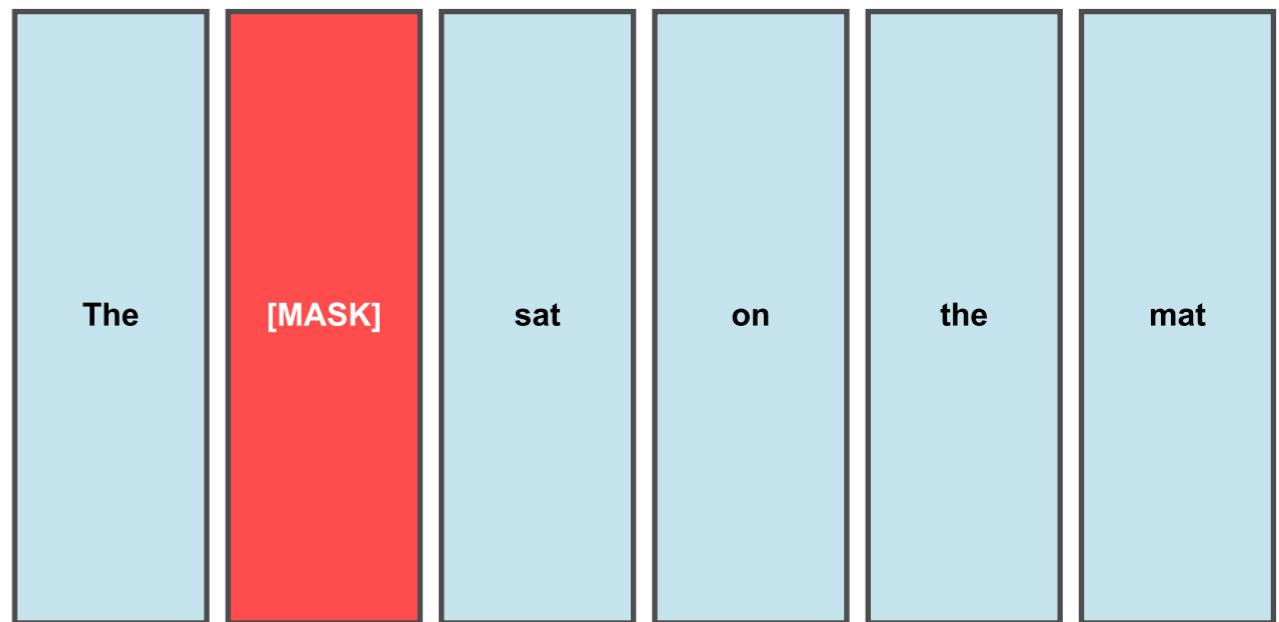


# BERT Masked Language Modeling Process

Step 1: Original Sentence

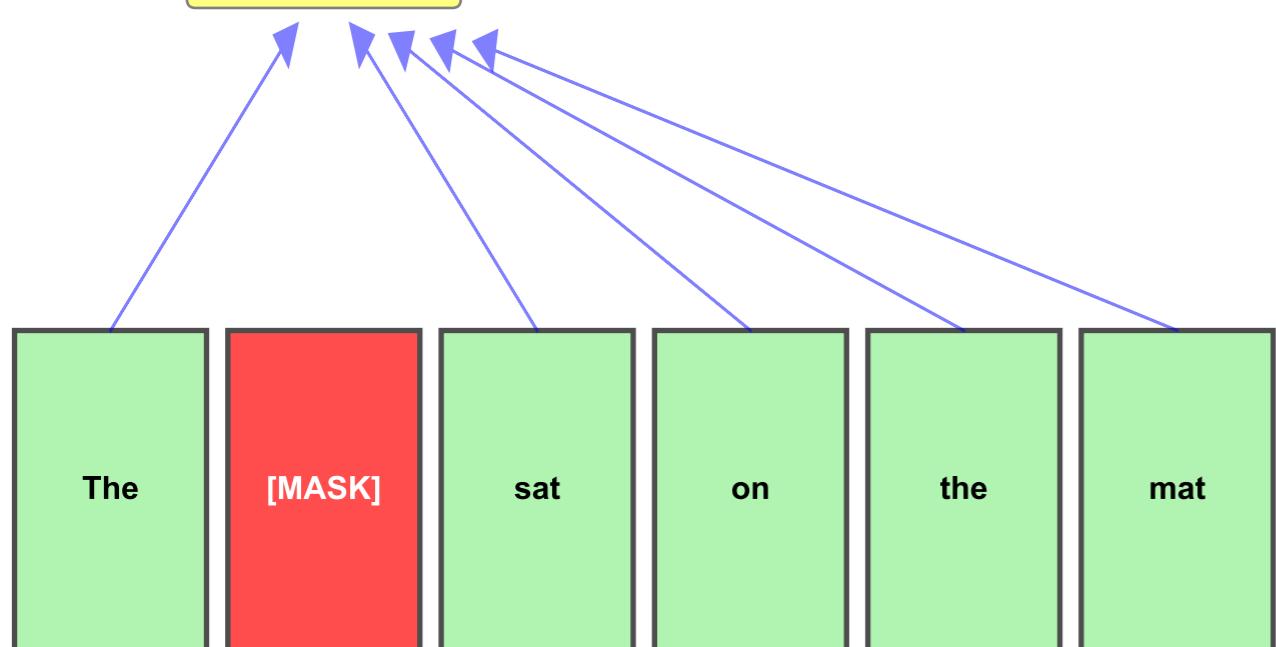


Step 2: Apply Masking (15% randomly)



Step 3: Bidirectional Context

All words contribute  
to prediction



Step 4: Predict Masked Token

