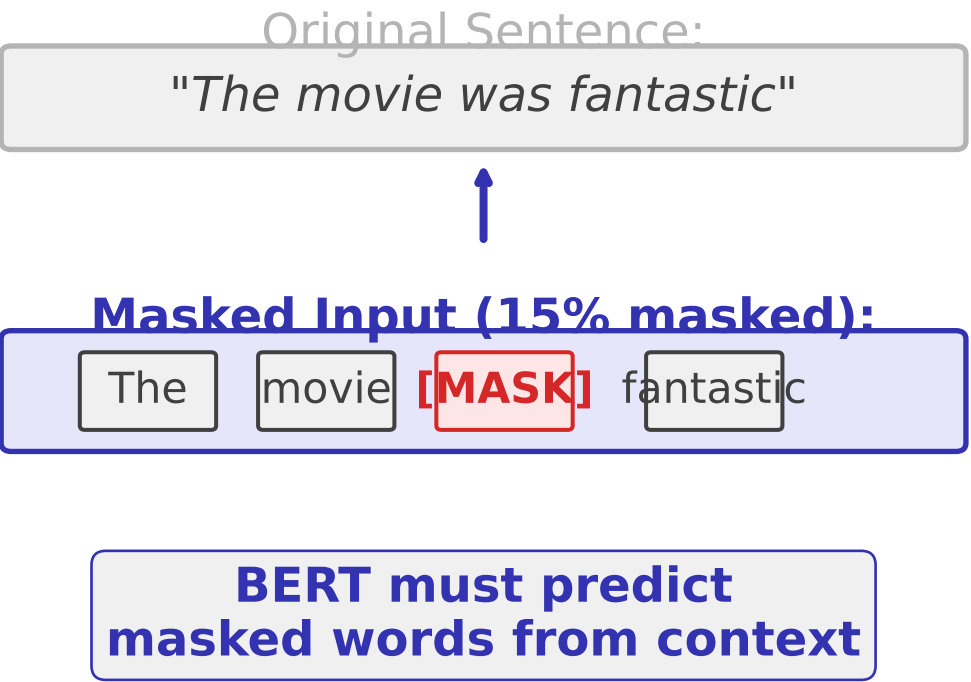
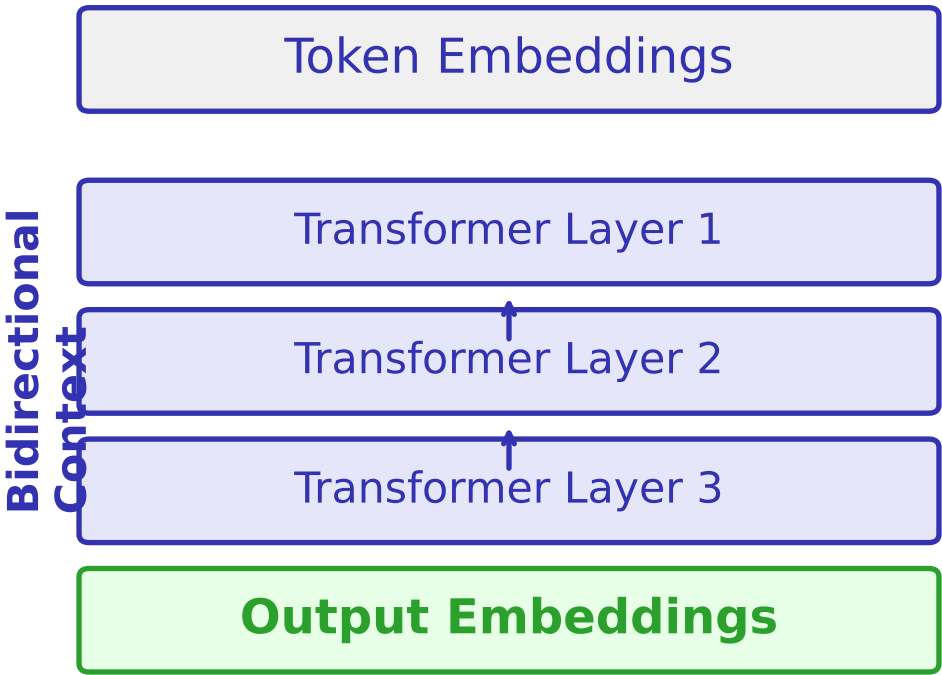


Stage 1: Pre-training with Masked Language Model (MLM)

Step 1: Mask 15% of Tokens



Step 2: BERT Processes Input



Step 3: Predict Masked Token

