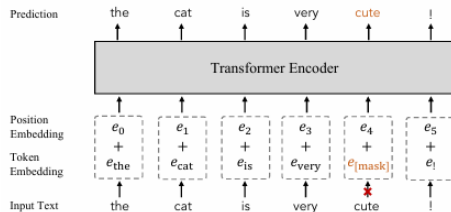


# Masked Language Modeling - Generative



Randomly masks a subset of text tokens and reconstructs them based on the unmasked tokens.

## Loss Formulation

$$\mathcal{L}_{\text{MLM}} = -\frac{1}{B} \sum_{i=1}^B \log f_{\phi}(x_i^T \mid \hat{x}_i^T)$$

where  $x_i^T$  is the masked token set and  $\hat{x}_i^T$  is the unmasked set.