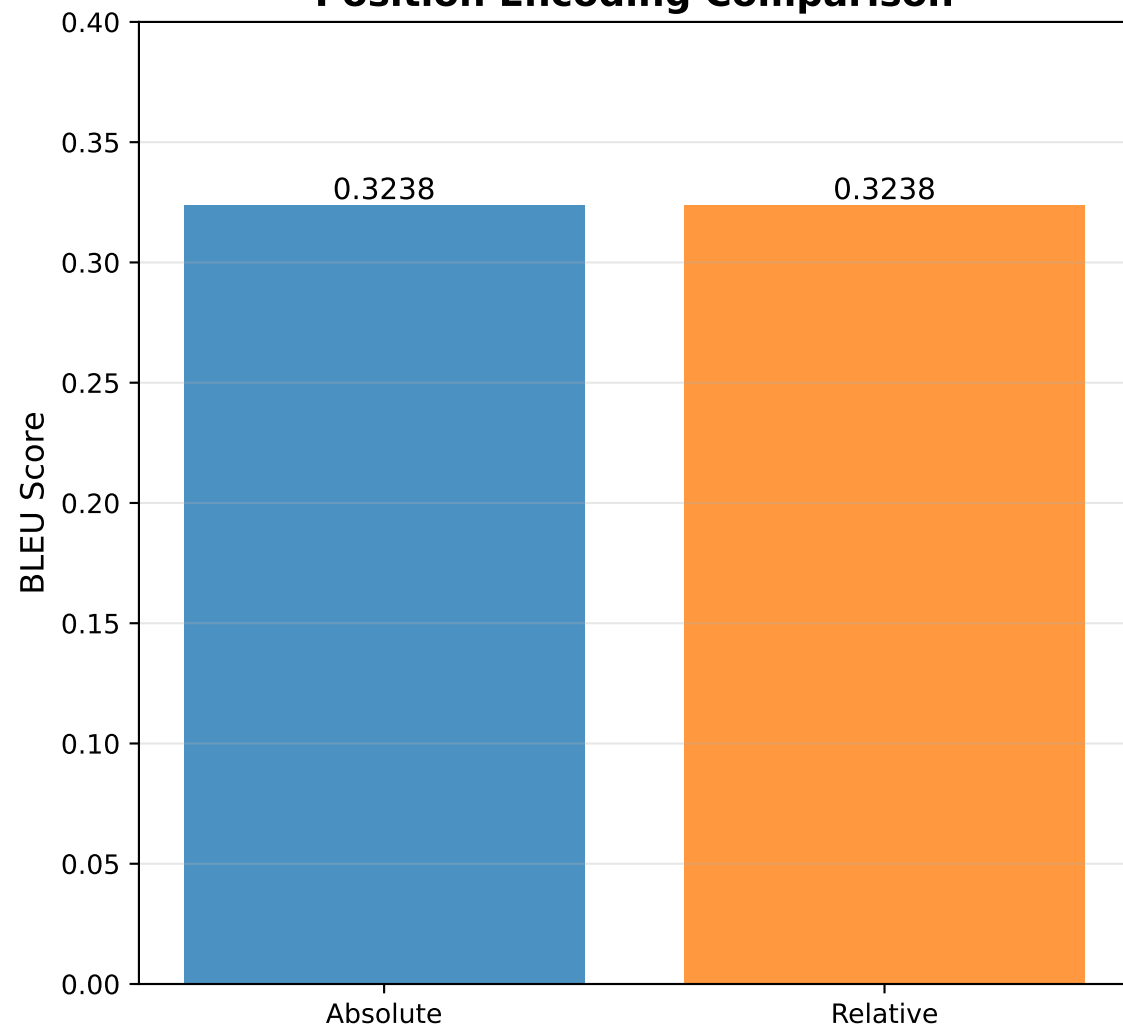
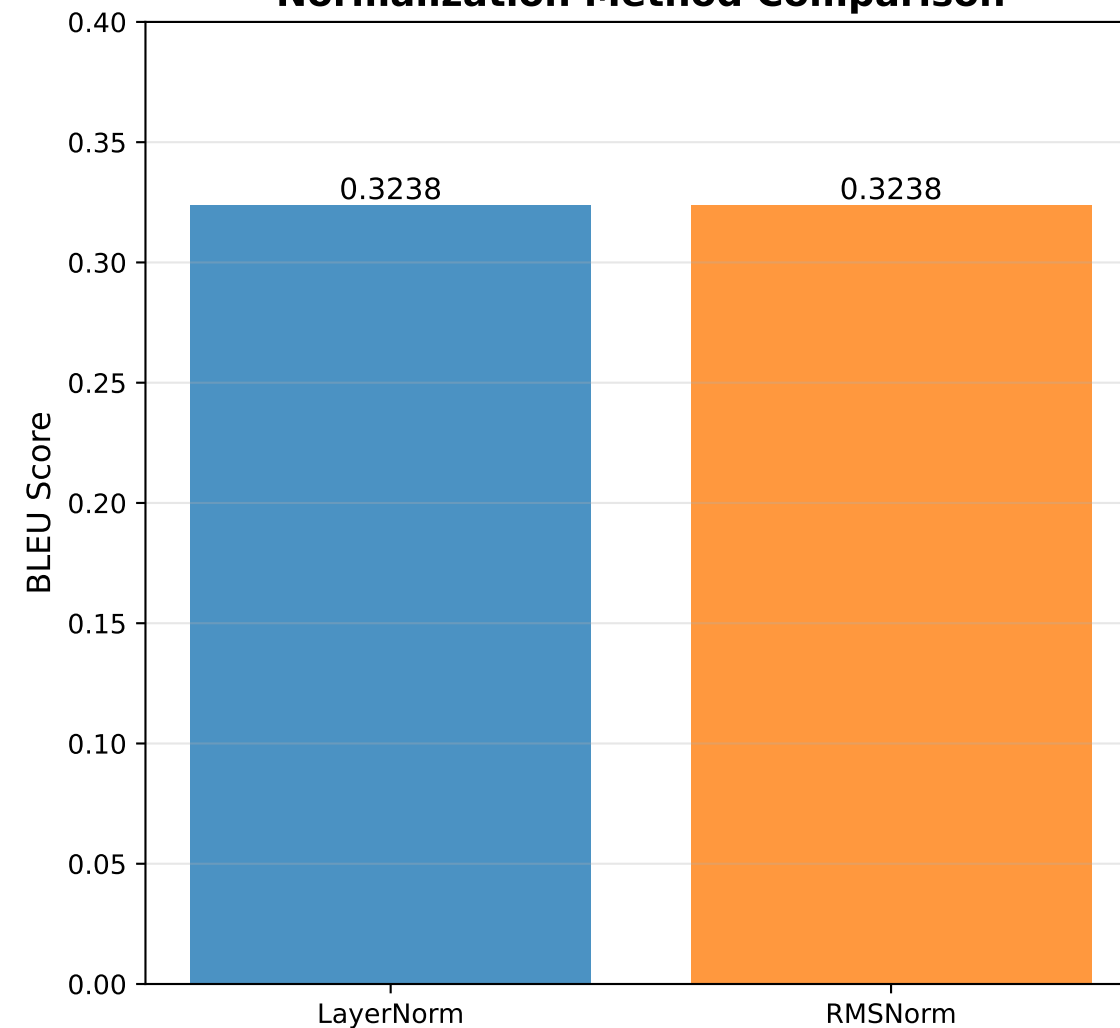


# Transformer Model Architecture Ablation Study

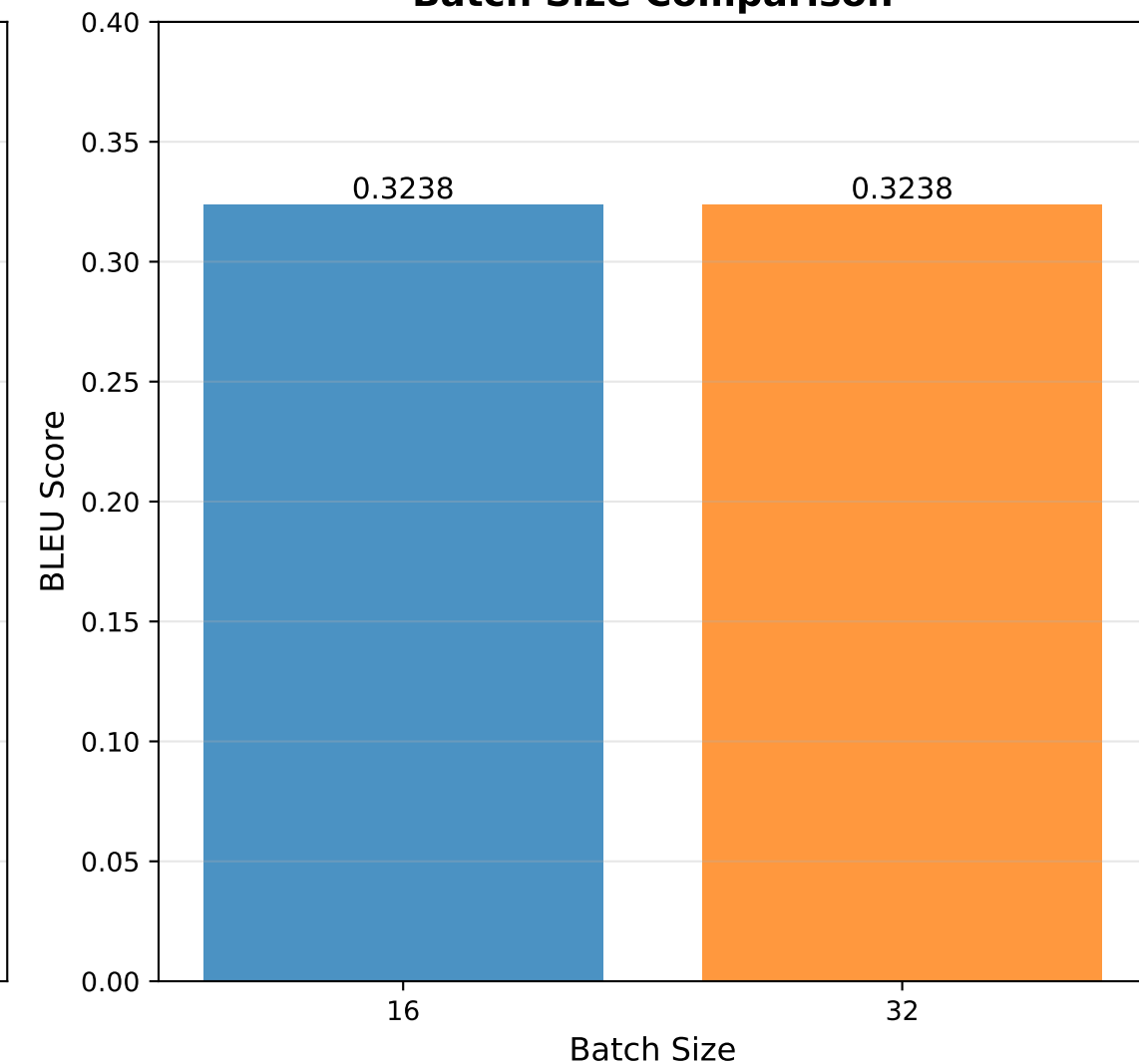
## Position Encoding Comparison



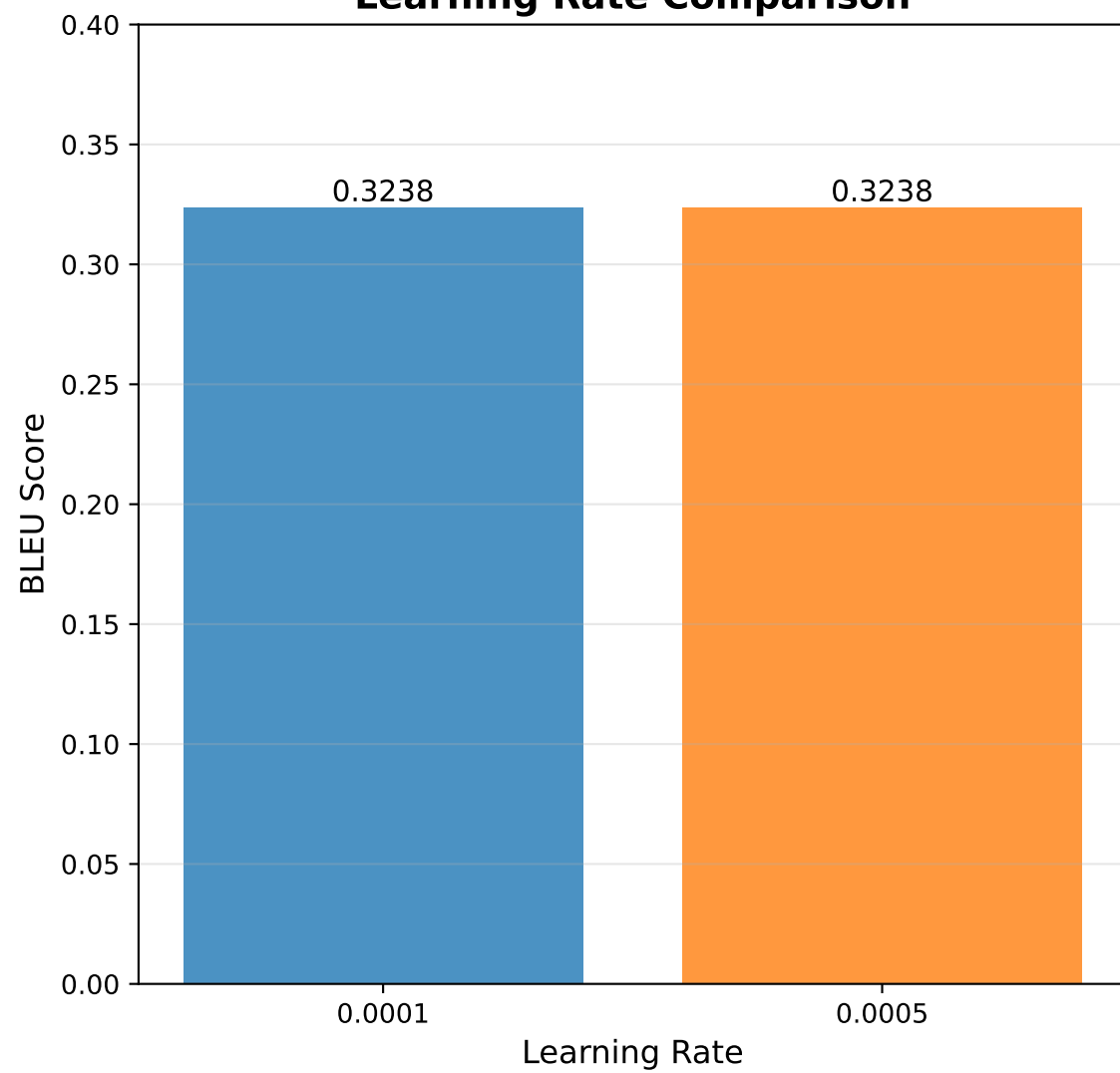
## Normalization Method Comparison



## Batch Size Comparison



## Learning Rate Comparison



## Model Size Comparison

