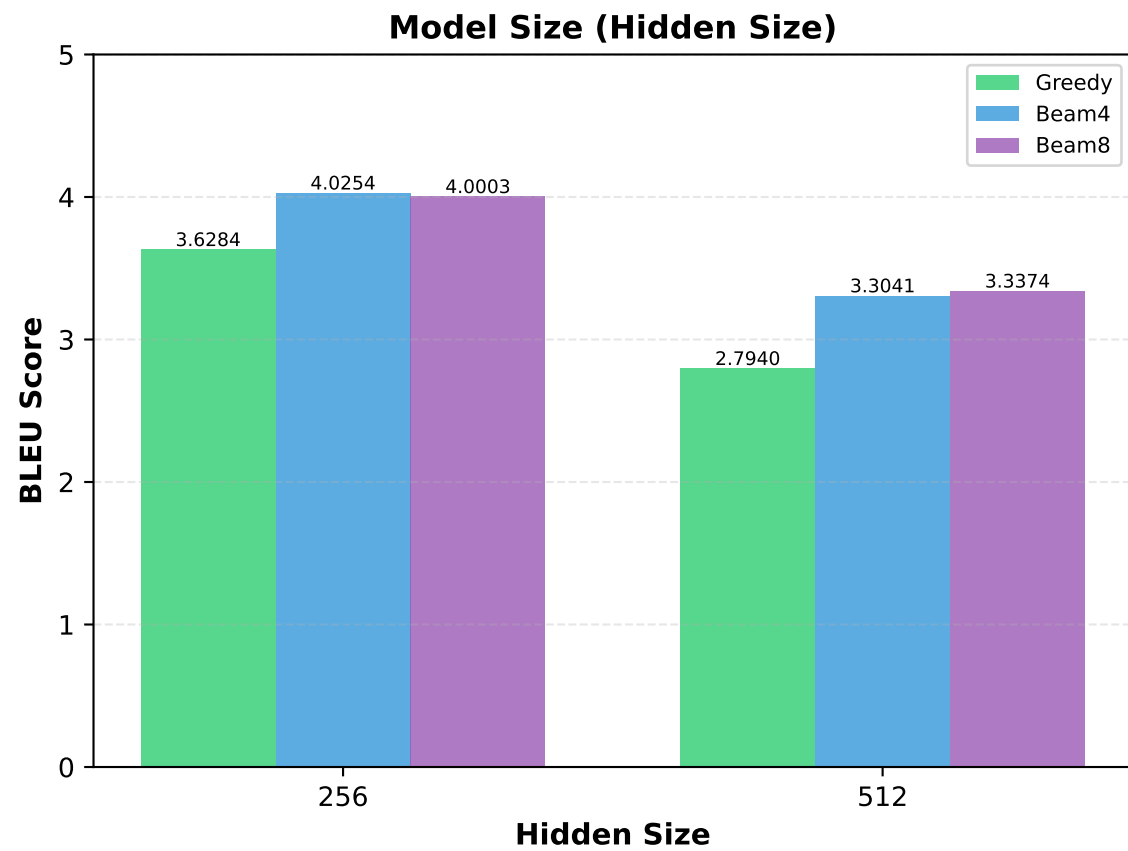
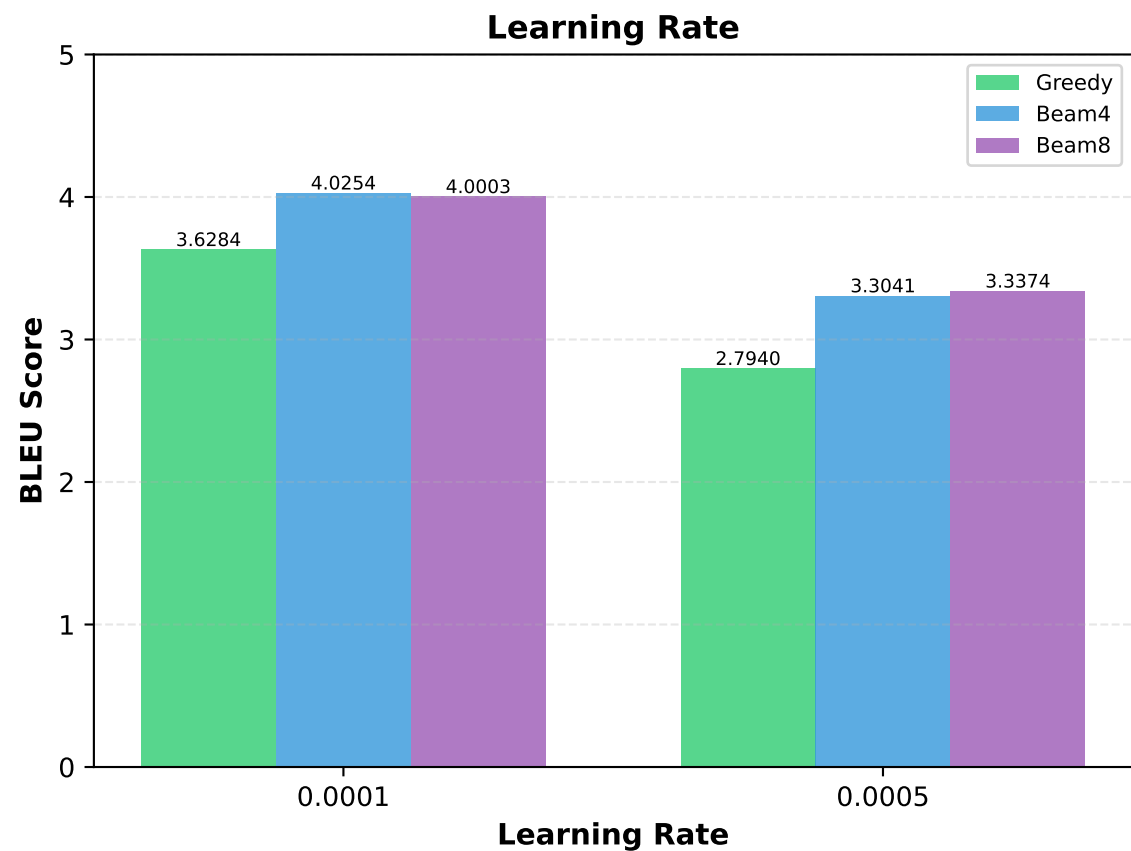
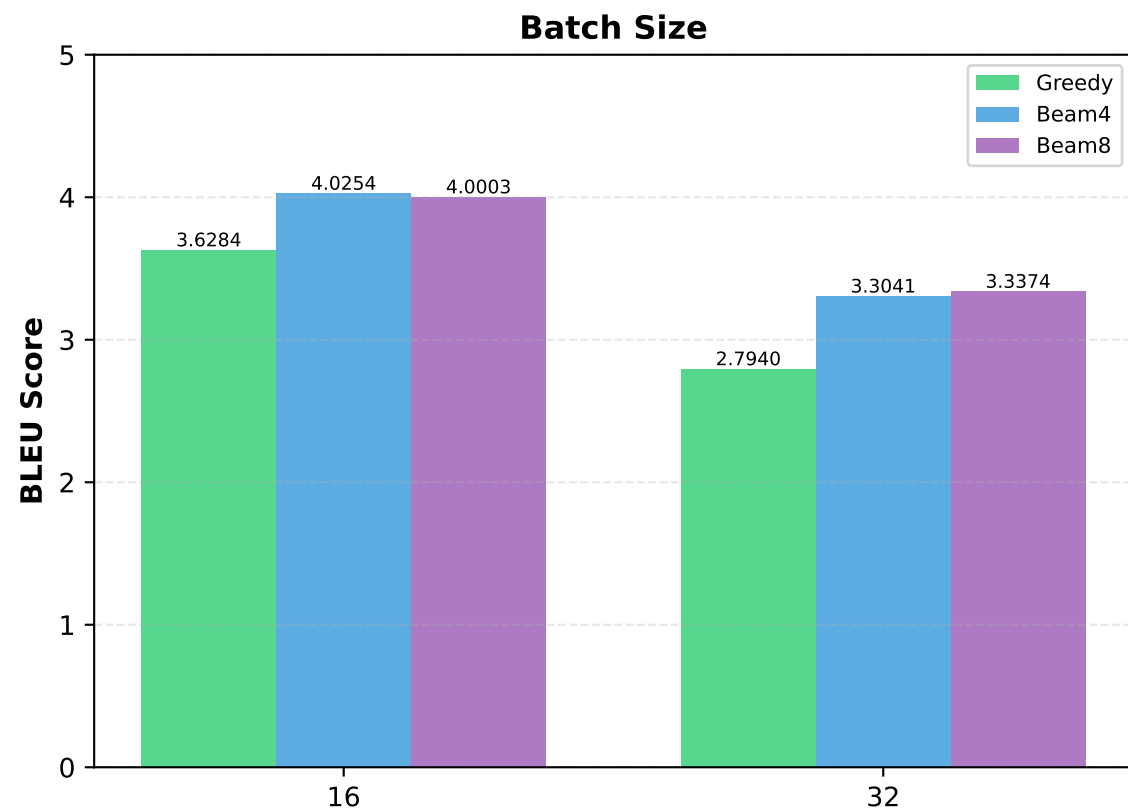
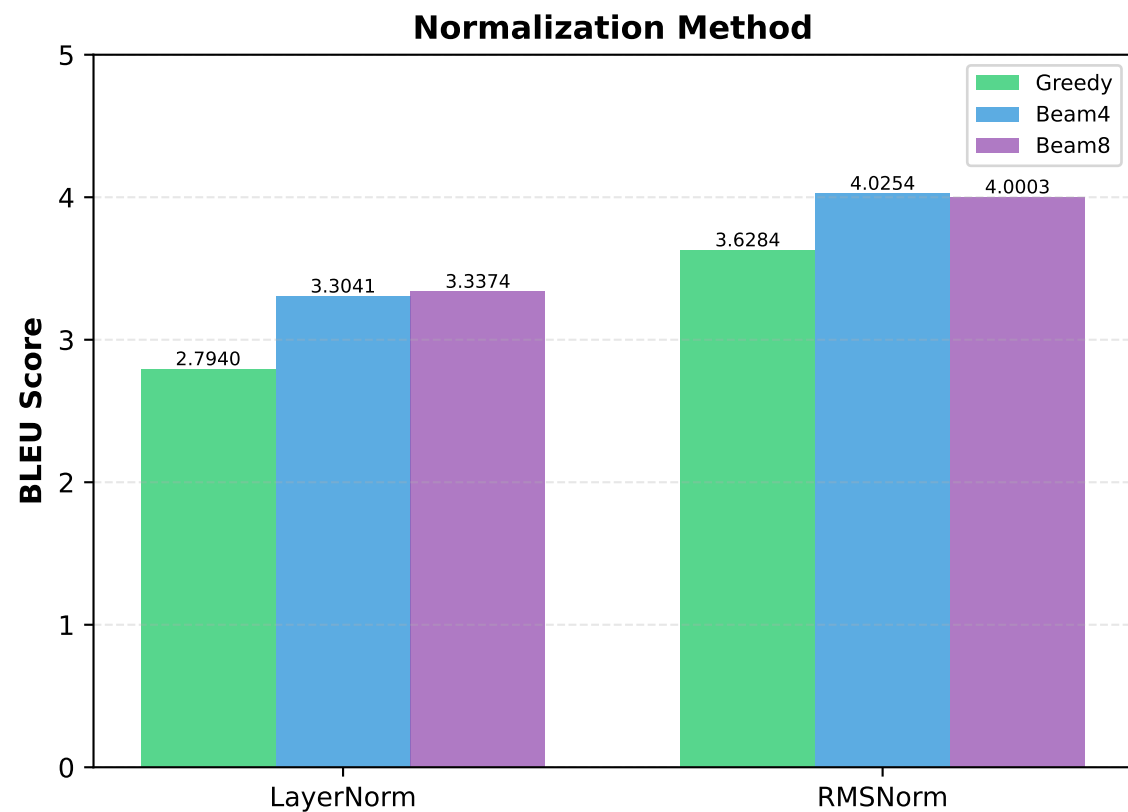
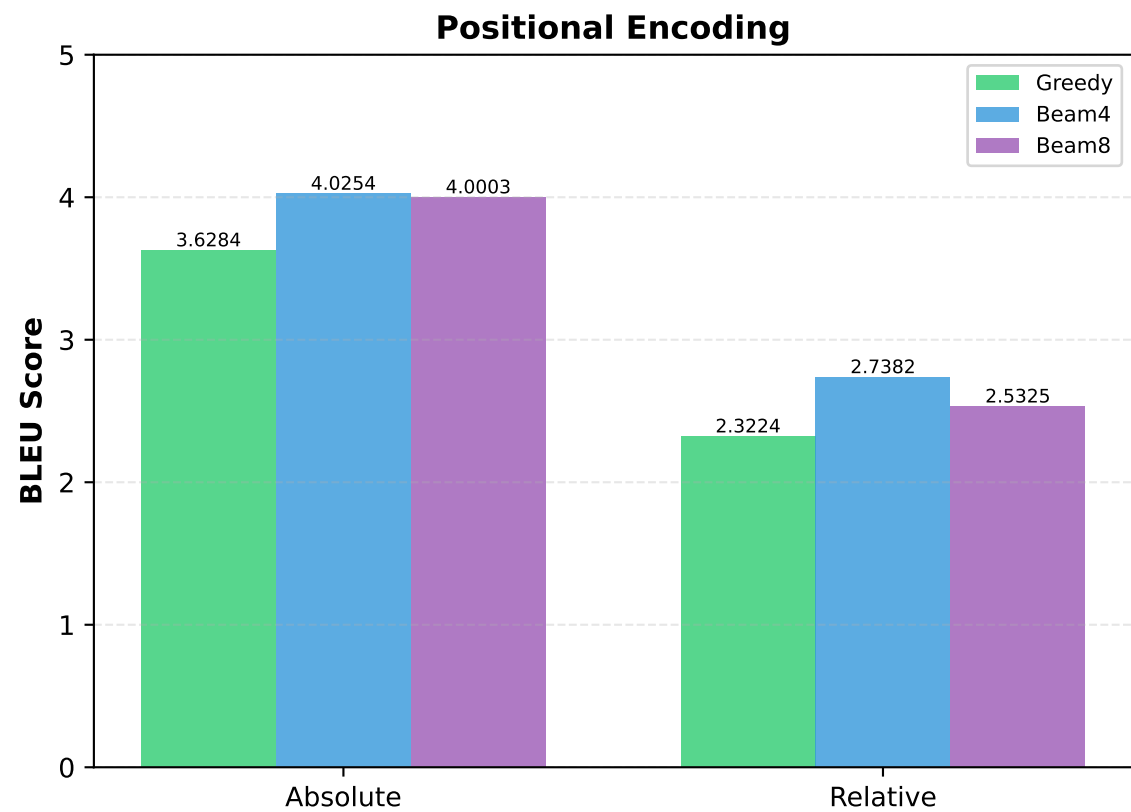


Transformer Architecture Ablation Study



Key Findings:

Positional Encoding:

- Absolute PE: 4.0254 (Best)
- Relative PE: 2.7382

Normalization:

- RMSNorm: 4.0254 (Best)
- LayerNorm: 3.3374

Batch Size:

- BS=16: 4.0254 (Best)
- BS=32: 3.3374

Learning Rate:

- LR=0.0001: 4.0254 (Best)
- LR=0.0005: 3.3374

Model Size:

- Hidden=256: 4.0254 (Best)
- Hidden=512: 3.3374

Optimal Config:

Absolute PE + RMSNorm
+ BS=16 + LR=0.0001
+ Hidden=256