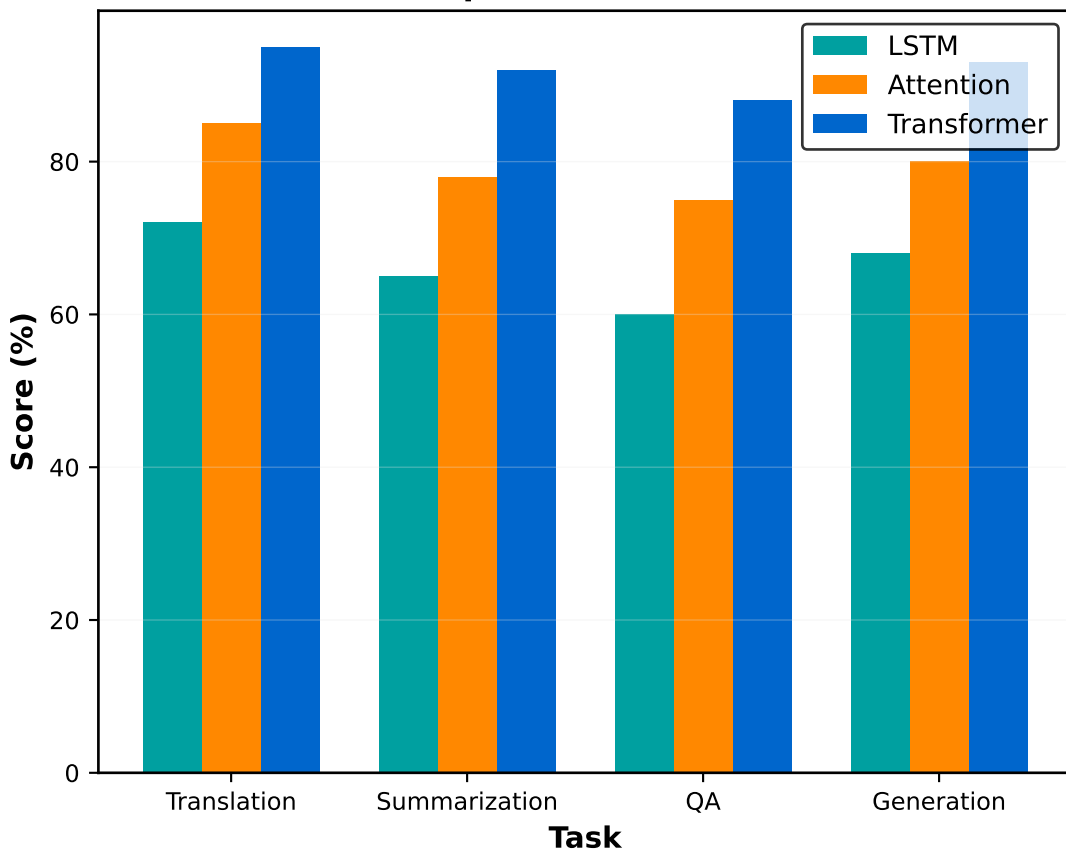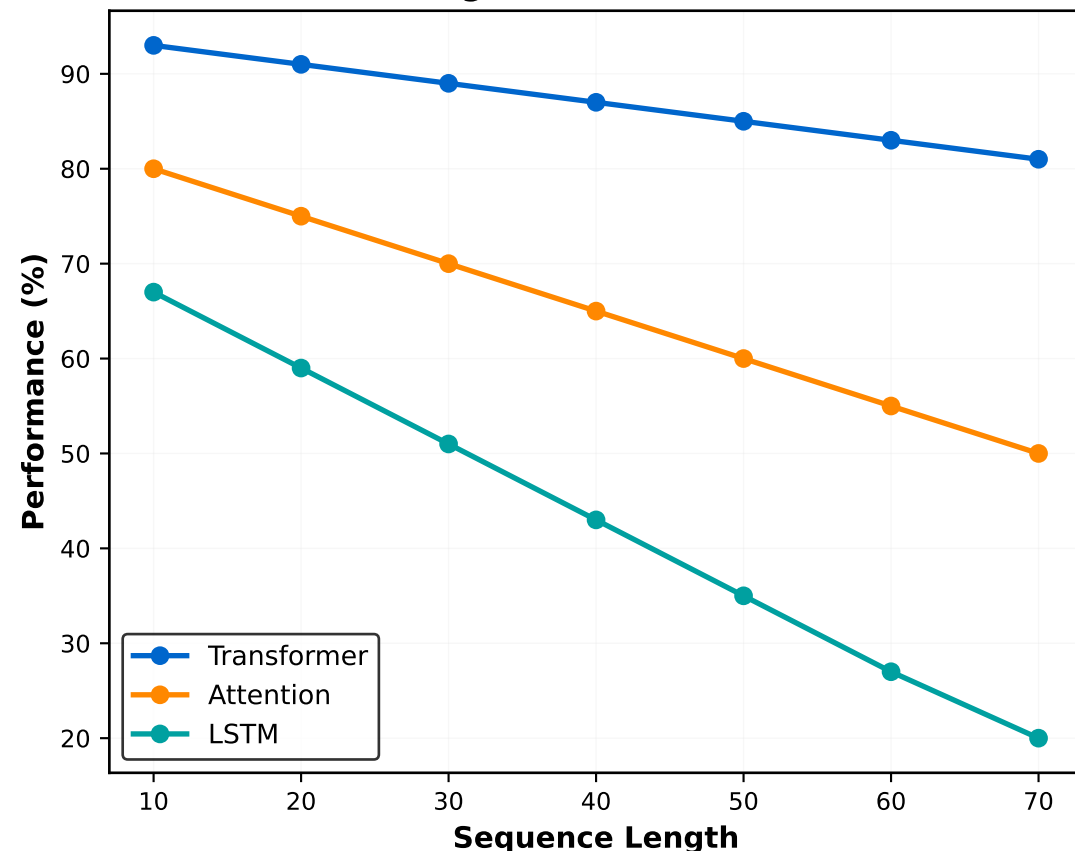**Comprehensive Performance Analysis**
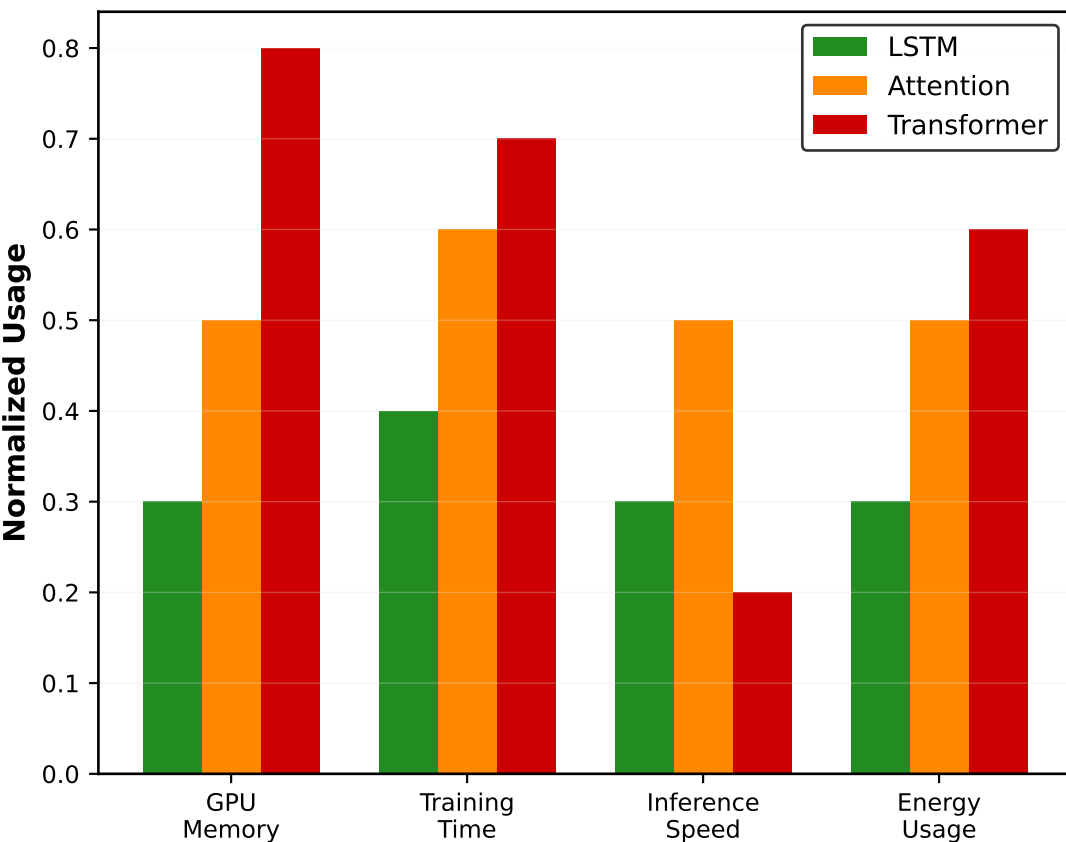
Task-Specific Performance

Length Generalization

Resource Utilization (Lower is Better)

Error Analysis by Type