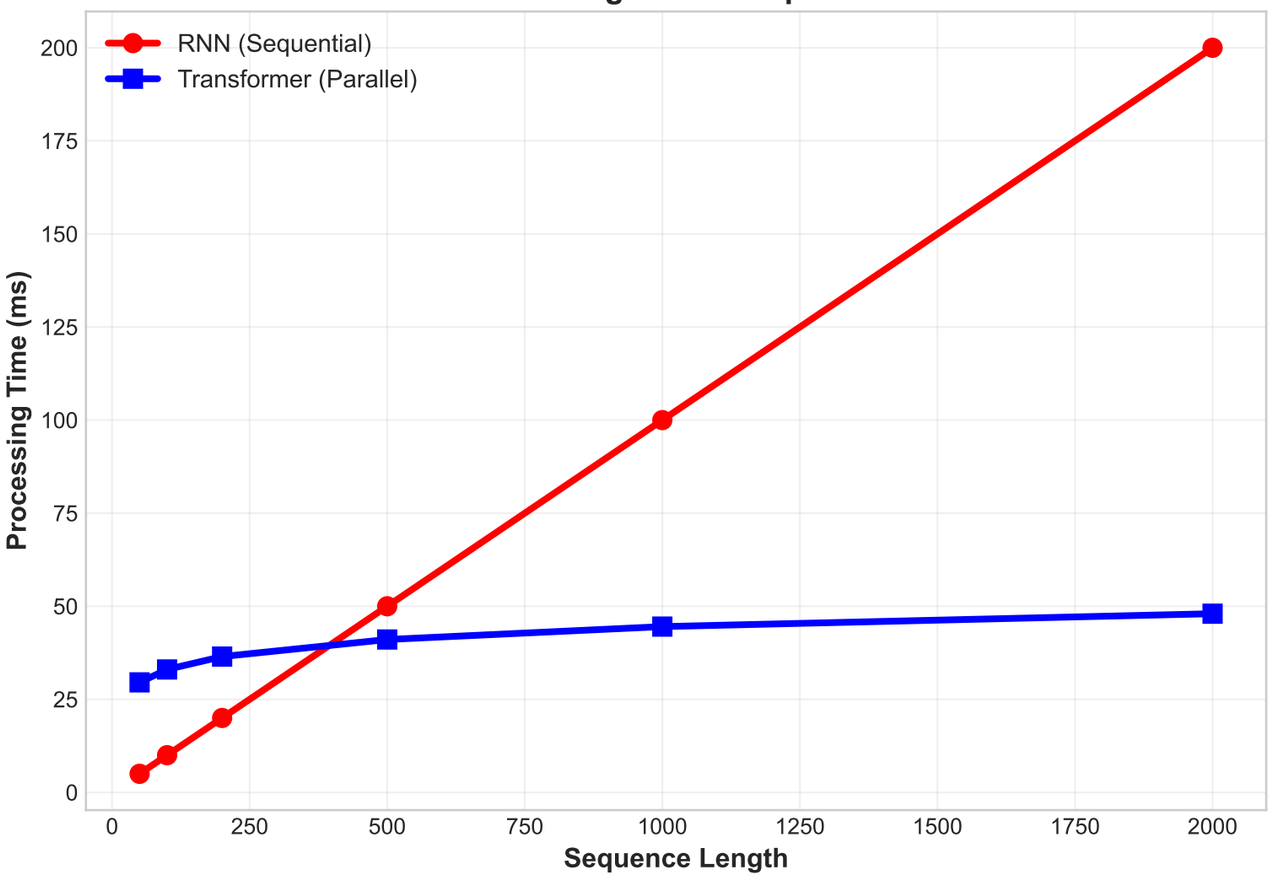
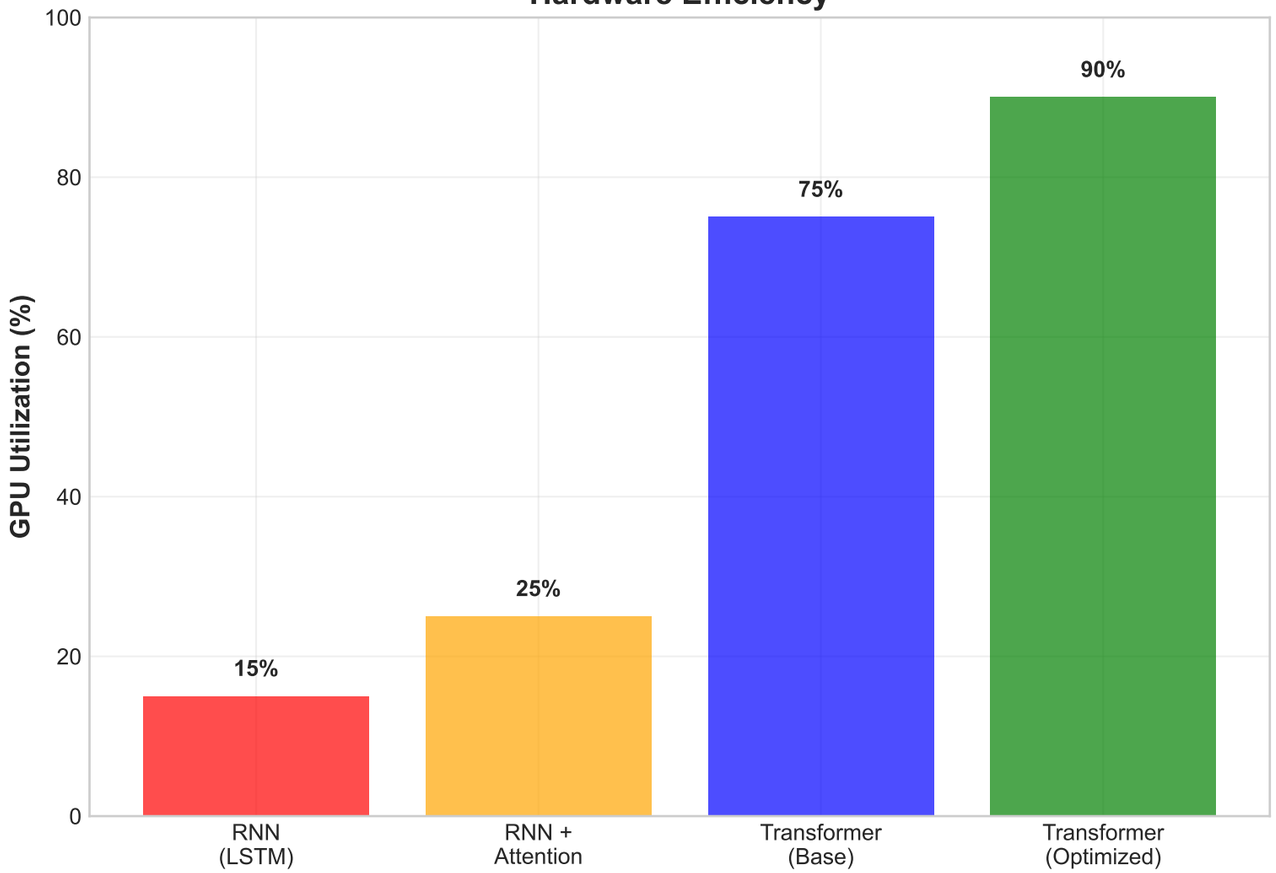
Transformer vs RNN: Training Speed Analysis

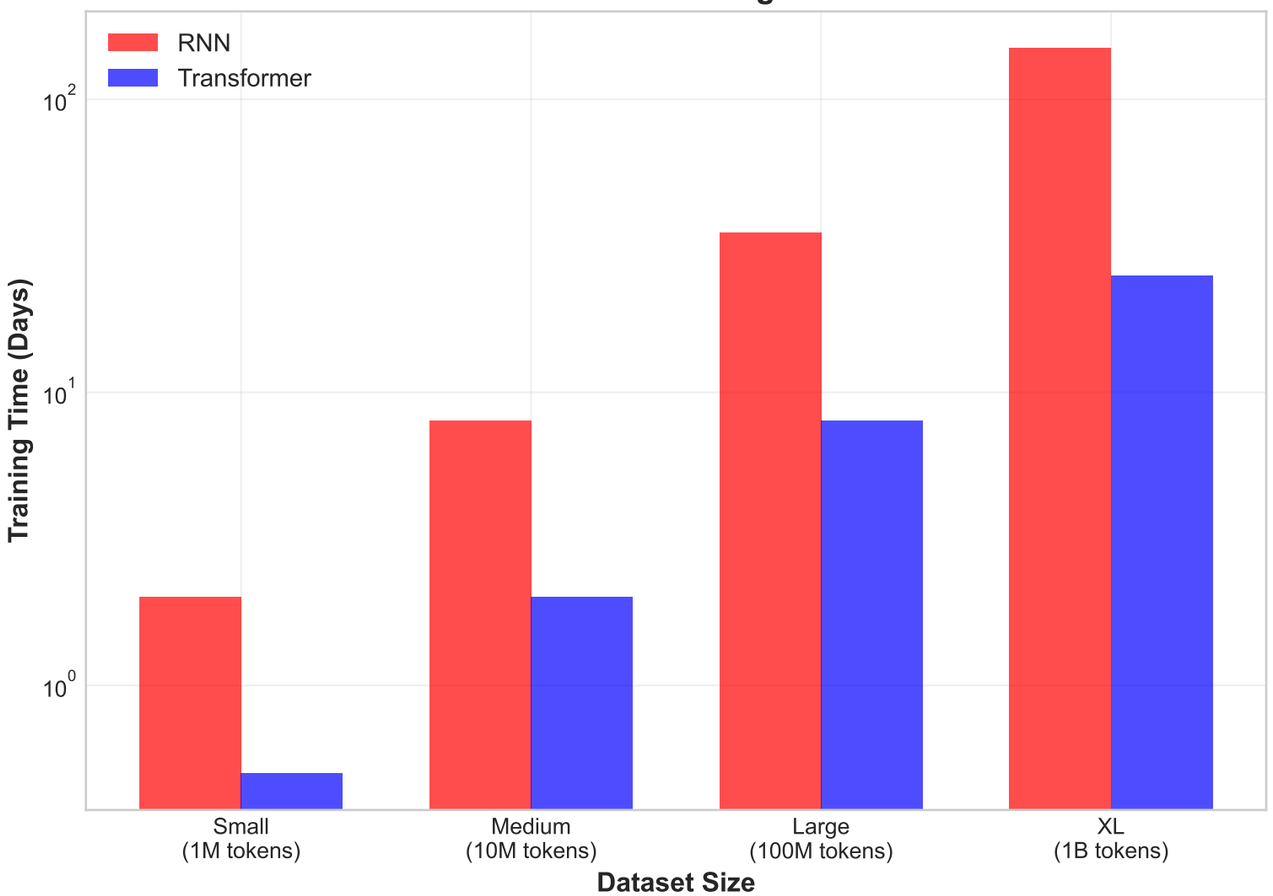
Processing Time Comparison



Hardware Efficiency



Real-World Training Times



Memory Scaling