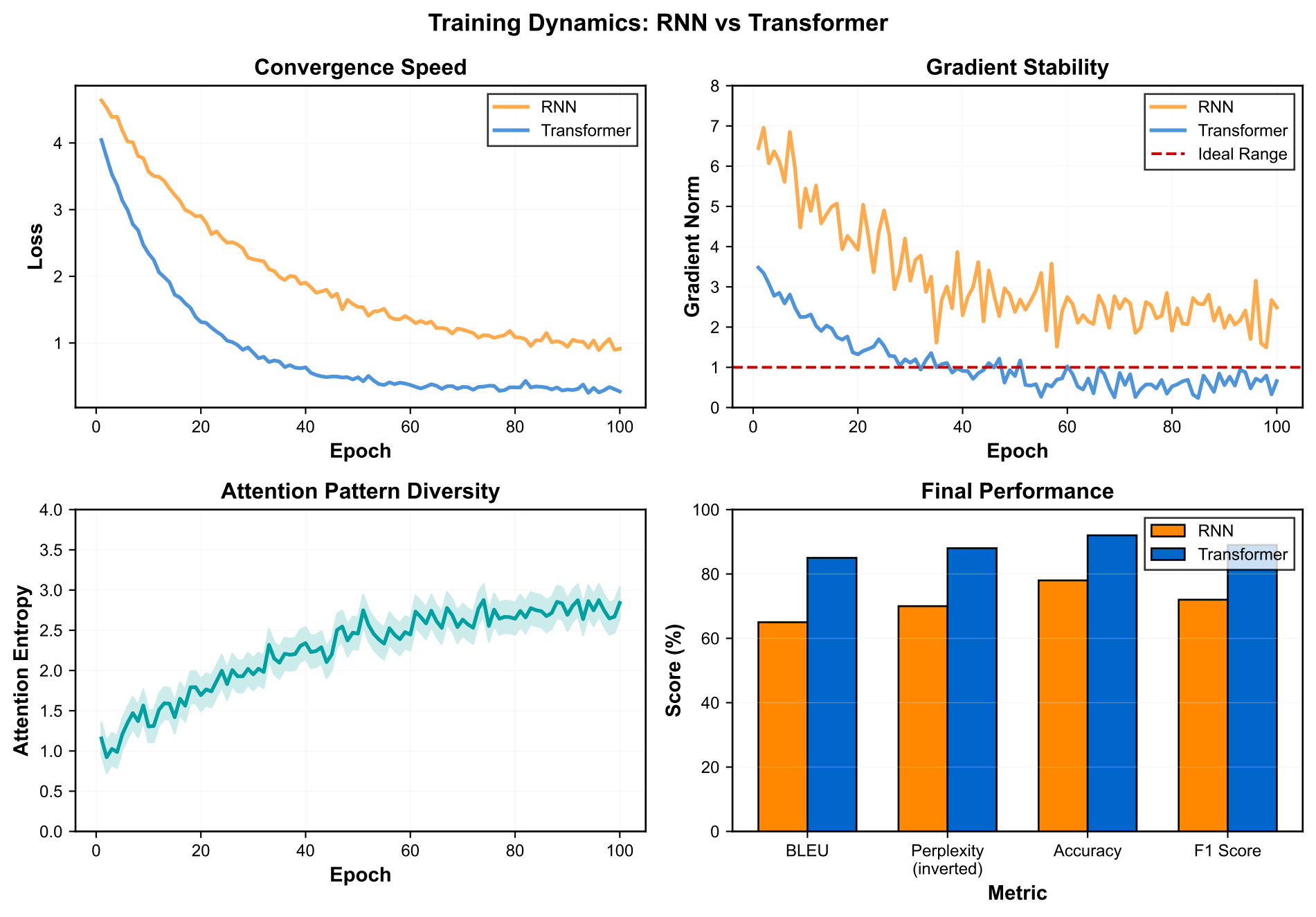
Training Dynamics: RNN vs Transformer
Convergence Speed
Loss
Epoch
RNN
Transformer
Gradient Stability
Gradient Norm
Epoch
RNN
Transformer
Ideal Range
Attention Pattern Diversity
Attention Entropy
Epoch
Final Performance
Score (%)
RNN
Transformer
BLEU
Perplexity (inverted)
Accuracy
F1 Score
Metric