



Figure S1: Training loss and perplexity curves across all 10 epochs, demonstrating smooth and consistent convergence without plateau or divergence. The gradual decrease in training loss from 0.883 at epoch 1 to 0.663 at epoch 10, alongside perplexity reduction from 2.42 to 1.94, shows stable model learning.