Training Perplexity Comparison

Legend: Dynamic (gamma=0.5), MoD (gamma=0.5)

X-axis: Training Step

Y-axis: Training Perplexity (Log Scale)