Regularization Loss During Training
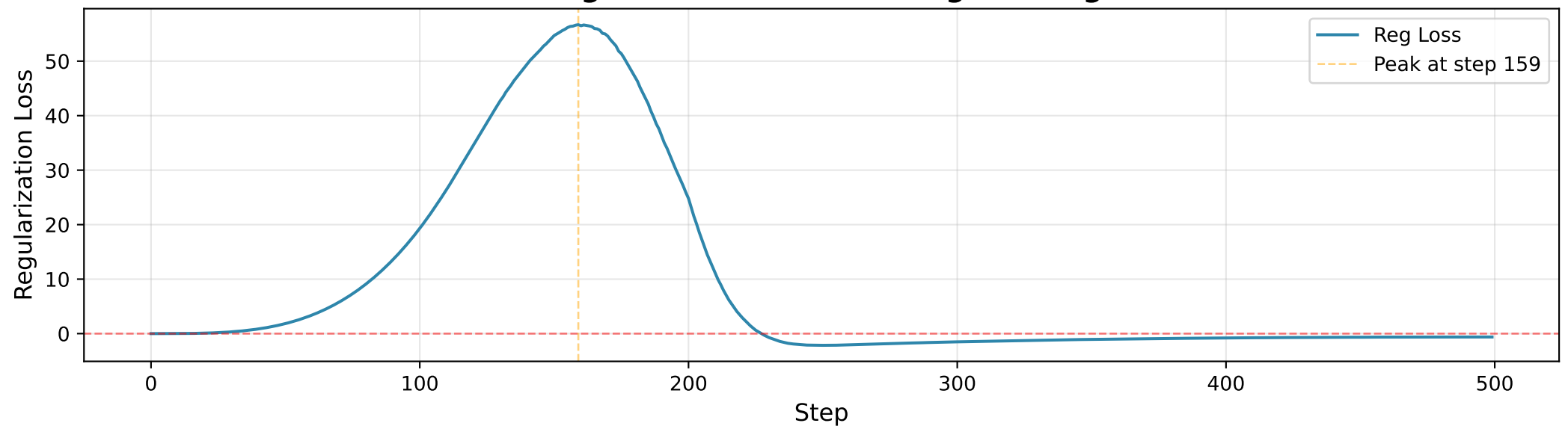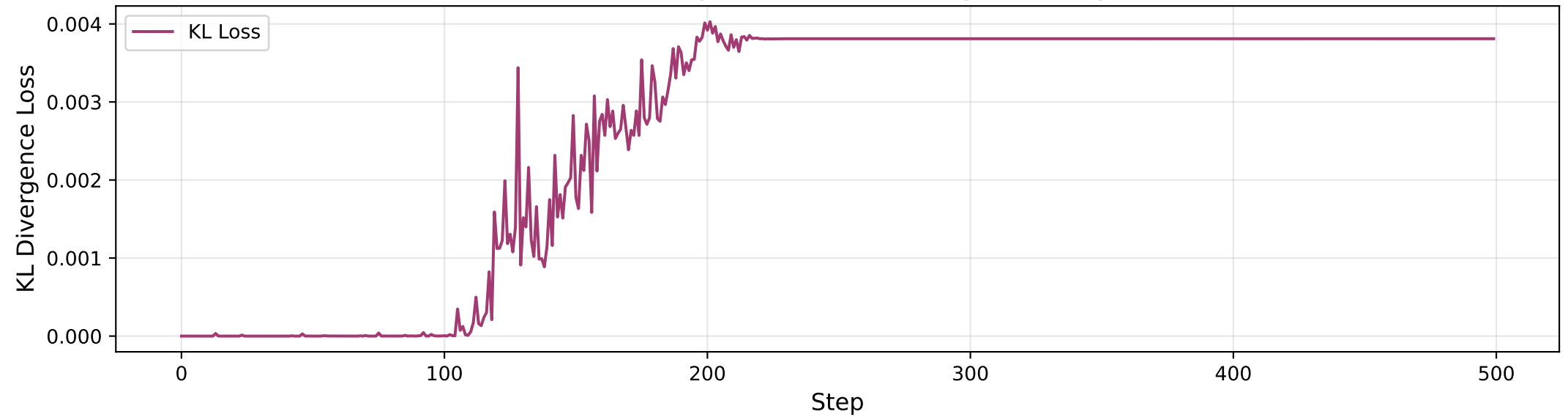
KL Divergence Loss During Training

Both Losses (Steps 250-500)