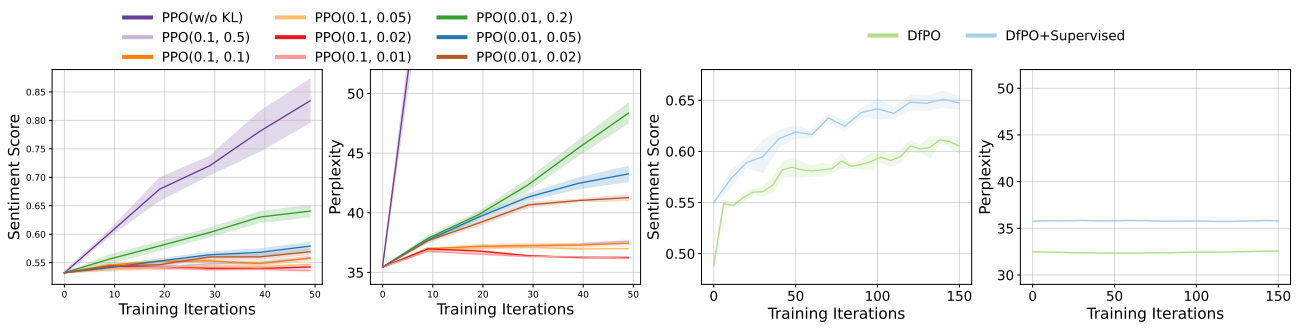

PPO(w/o KL)
PPO(0.1, 0.5)
PPO(0.1, 0.1)
PPO(0.1, 0.05)
PPO(0.1, 0.02)
PPO(0.1, 0.01)
PPO(0.01, 0.2)
PPO(0.01, 0.05)
PPO(0.01, 0.02)
DfPO
DfPO+Supervised
Sentiment Score
Perplexity
Training Iterations