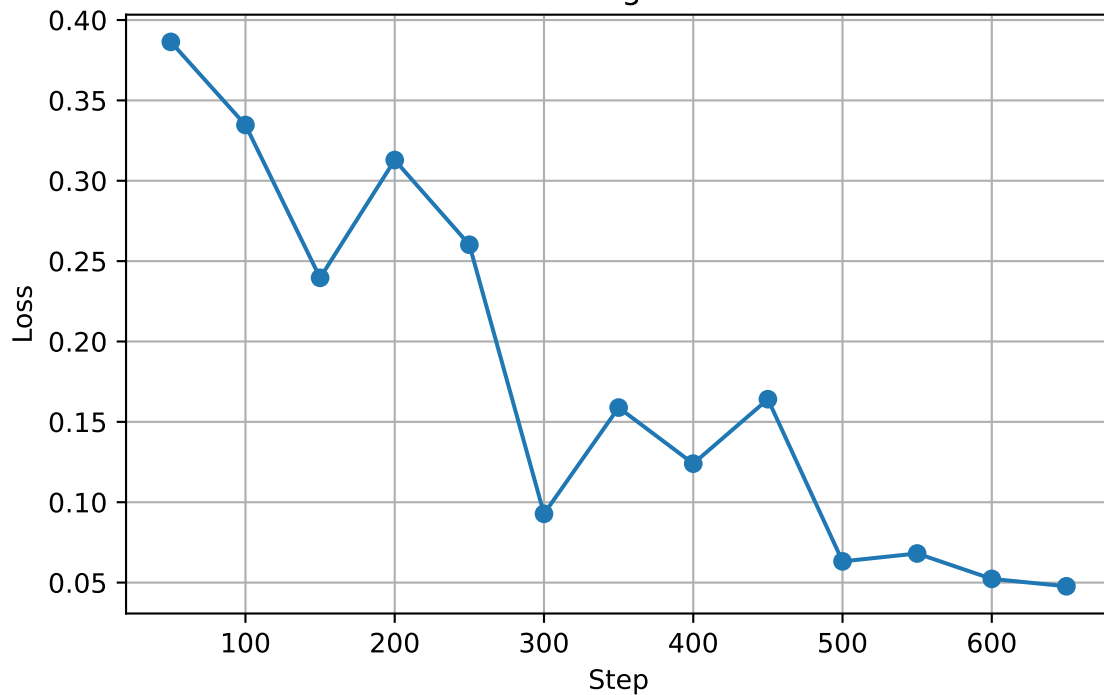
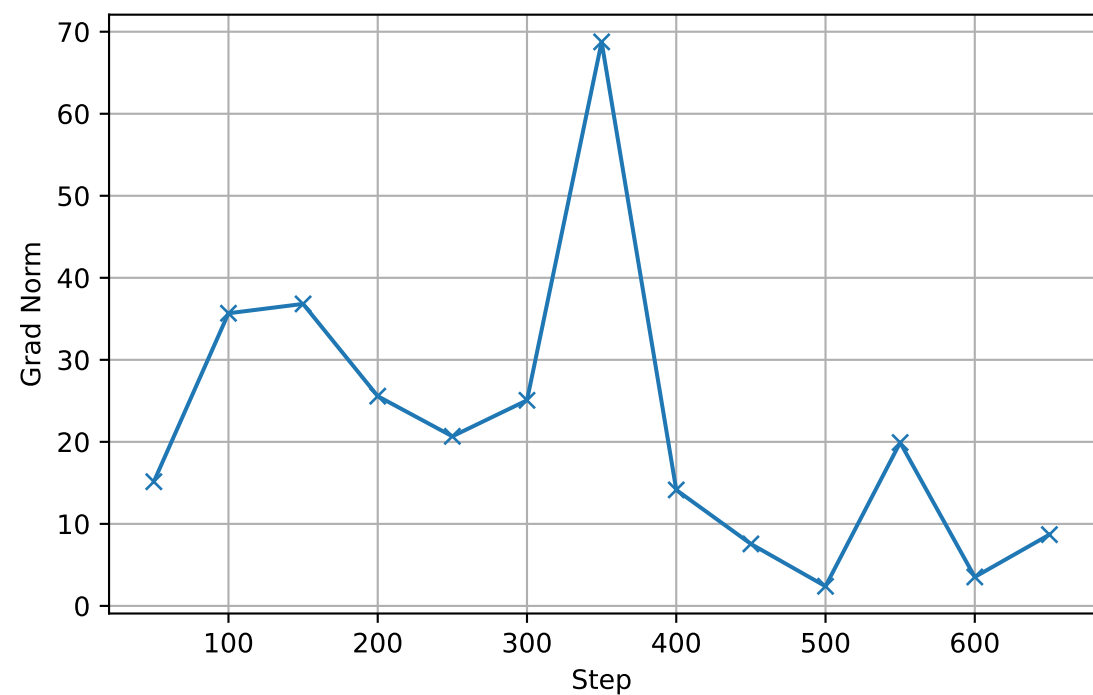


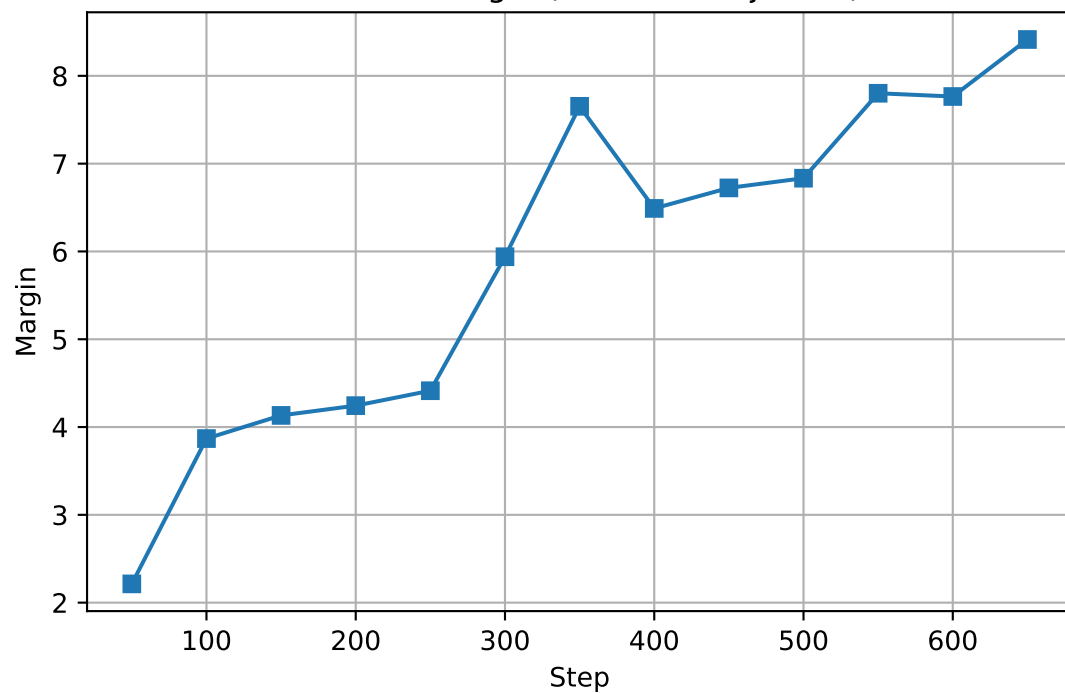
Training Loss



Gradient Norm



Reward Margin (Chosen - Rejected)



Reward Accuracy

