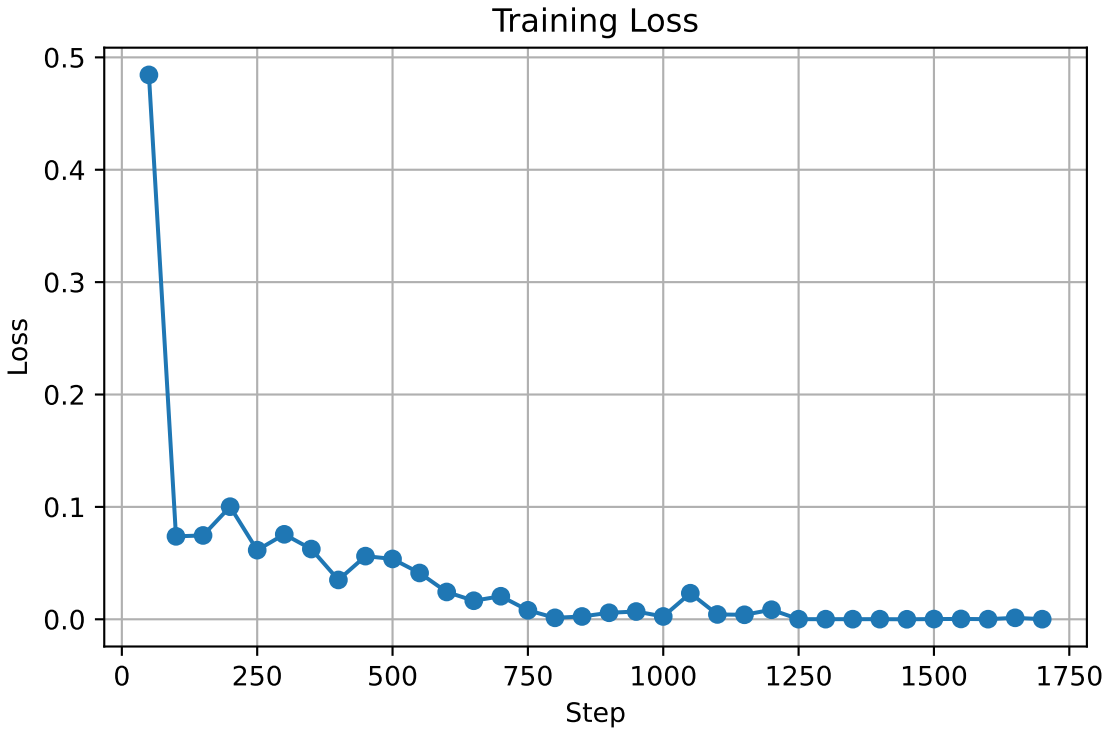
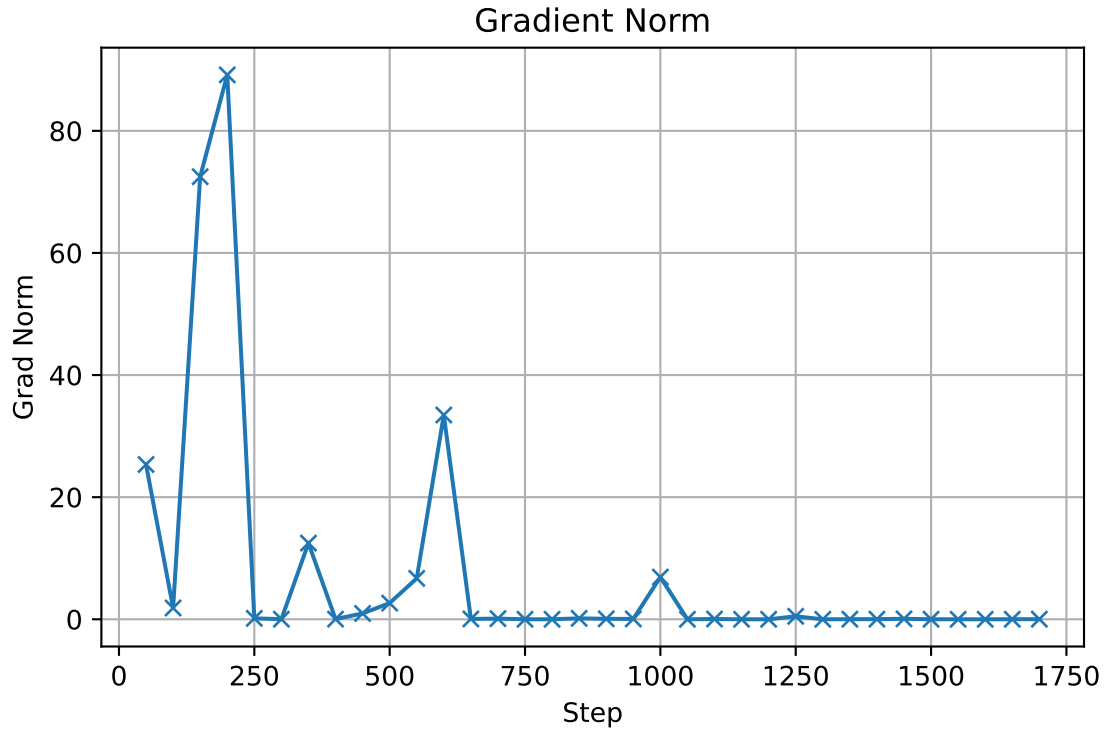
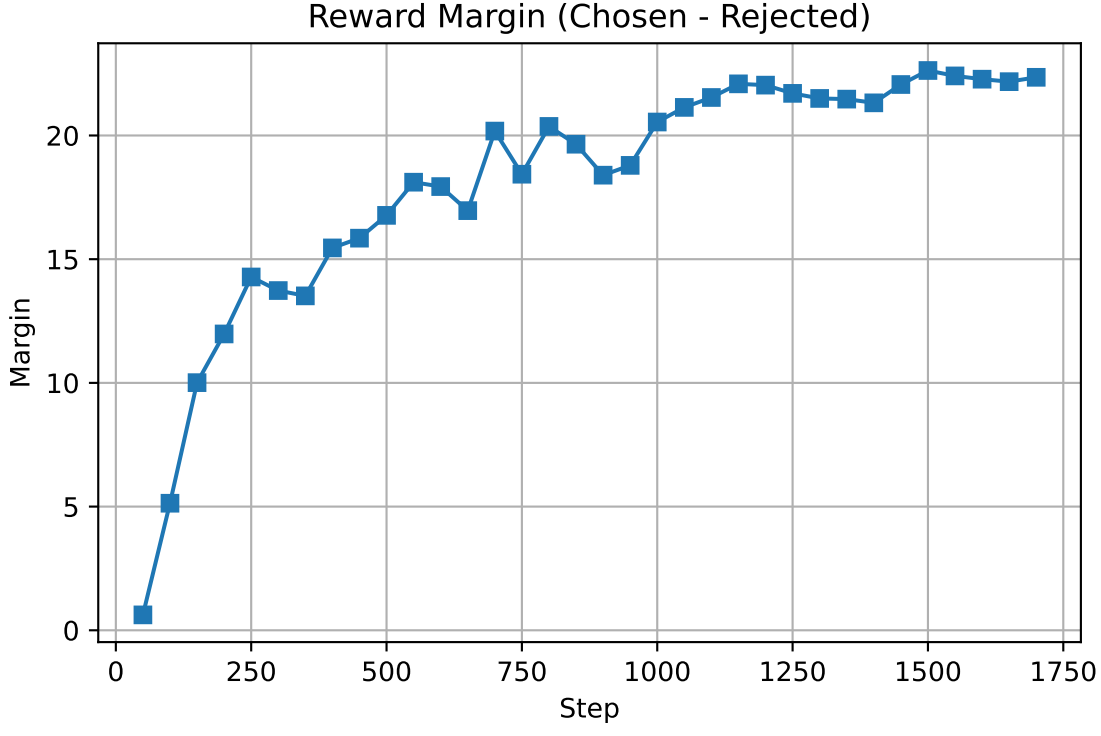
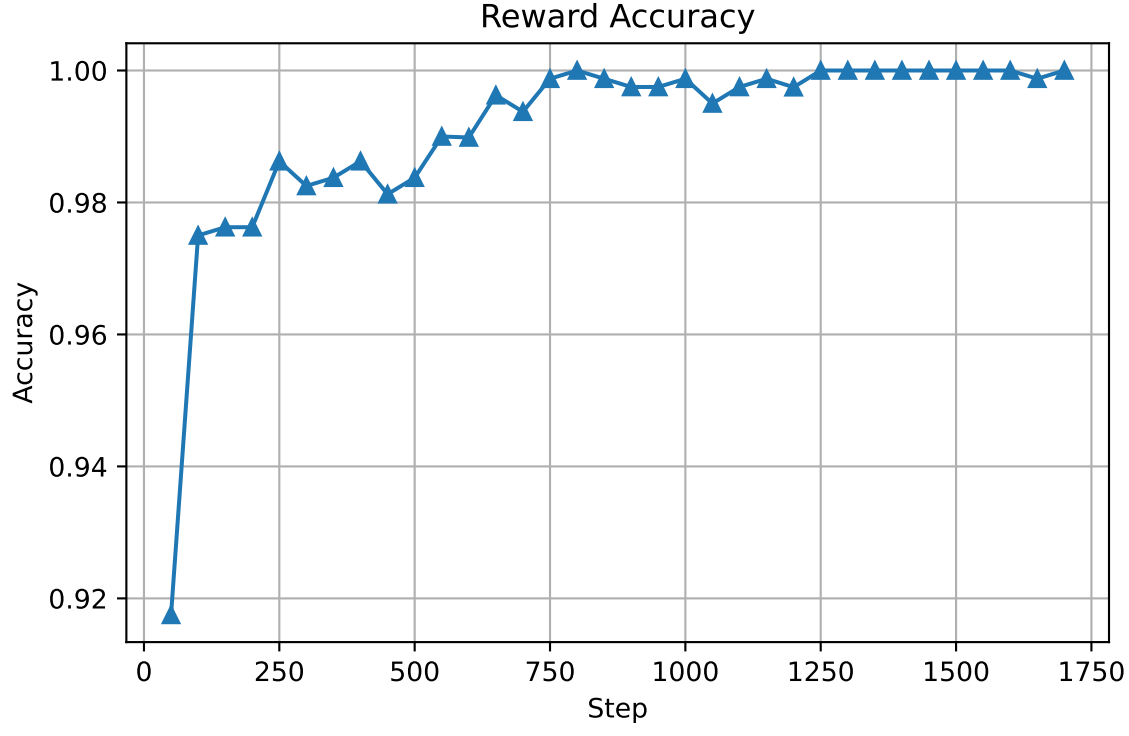

Training Loss
Loss
Step
0.0
0.1
0.2
0.3
0.4
0.5
0
250
500
750
1000
1250
1500
1750


Gradient Norm
Grad Norm
Step
0
20
40
60
80
0
250
500
750
1000
1250
1500
1750


Reward Margin (Chosen - Rejected)
Margin
Step
0
5
10
15
20
0
250
500
750
1000
1250
1500
1750


Reward Accuracy
Accuracy
Step
0.92
0.94
0.96
0.98
1.00
0
250
500
750
1000
1250
1500
1750