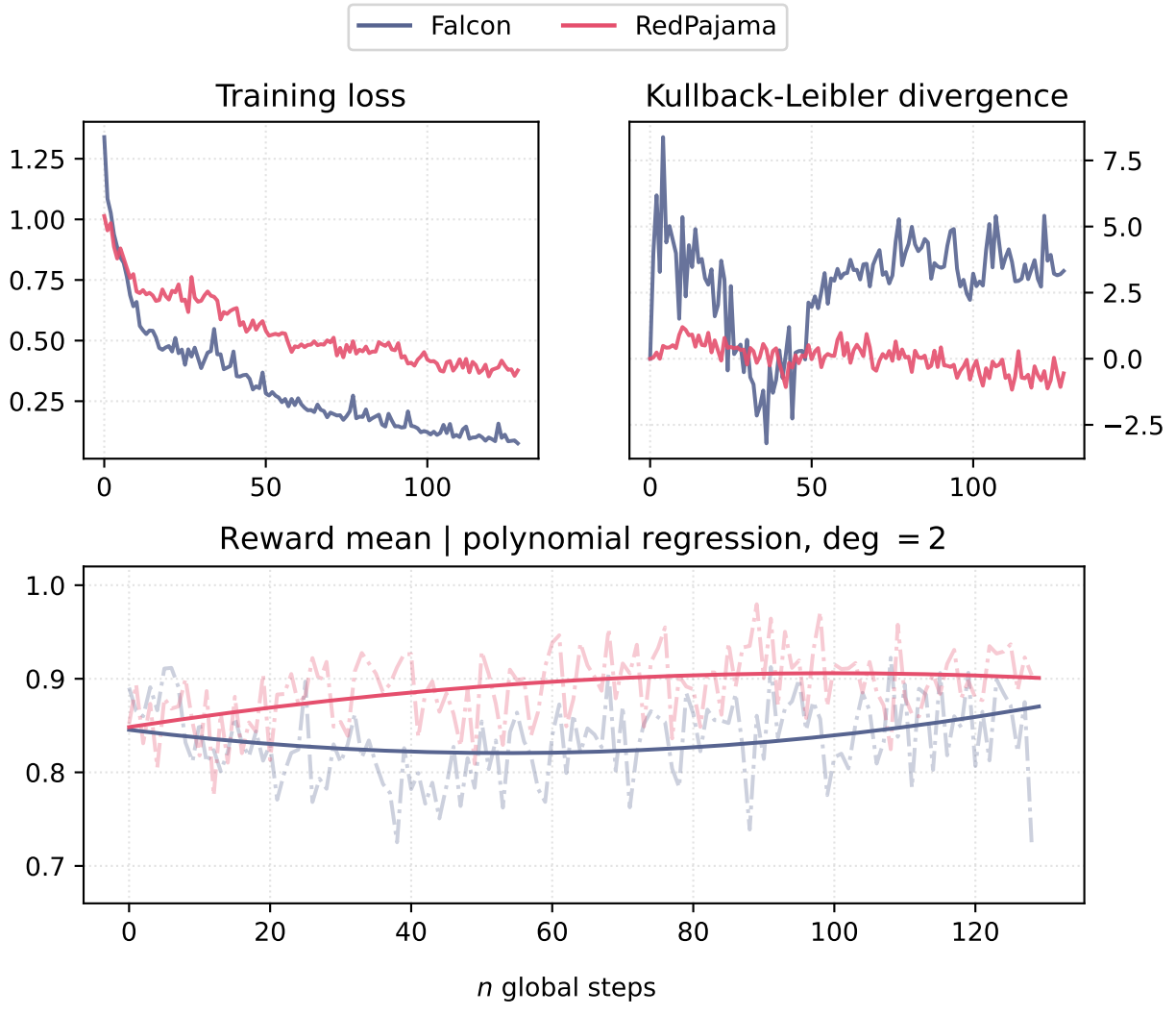
Falcon
RedPajama
Training loss
Kullback-Leibler divergence
1.25
1.00
0.75
0.50
0.25
0
50
100
7.5
5.0
2.5
0.0
−2.5
Reward mean | polynomial regression, deg = 2
1.0
0.9
0.8
0.7
0
20
40
60
80
100
120
n global steps