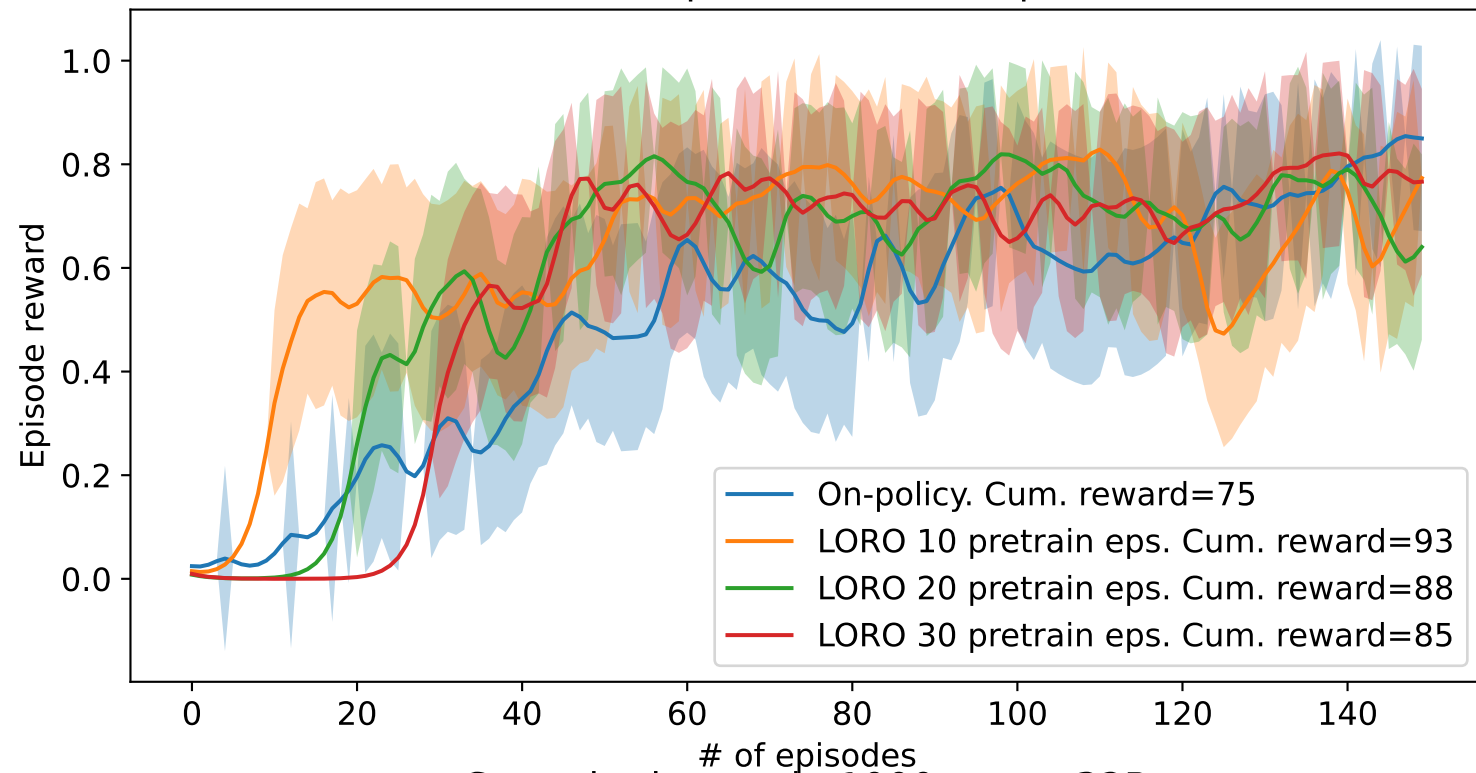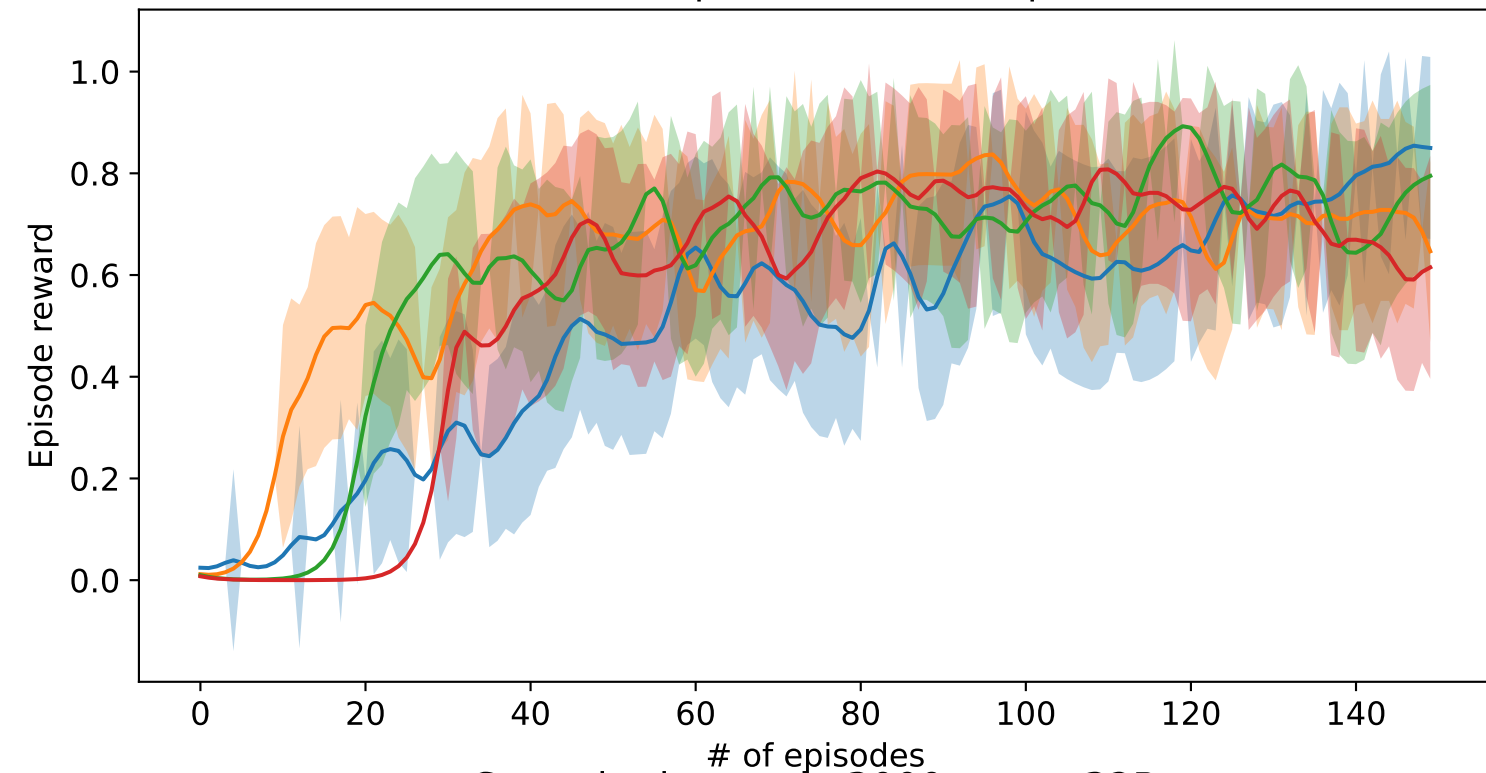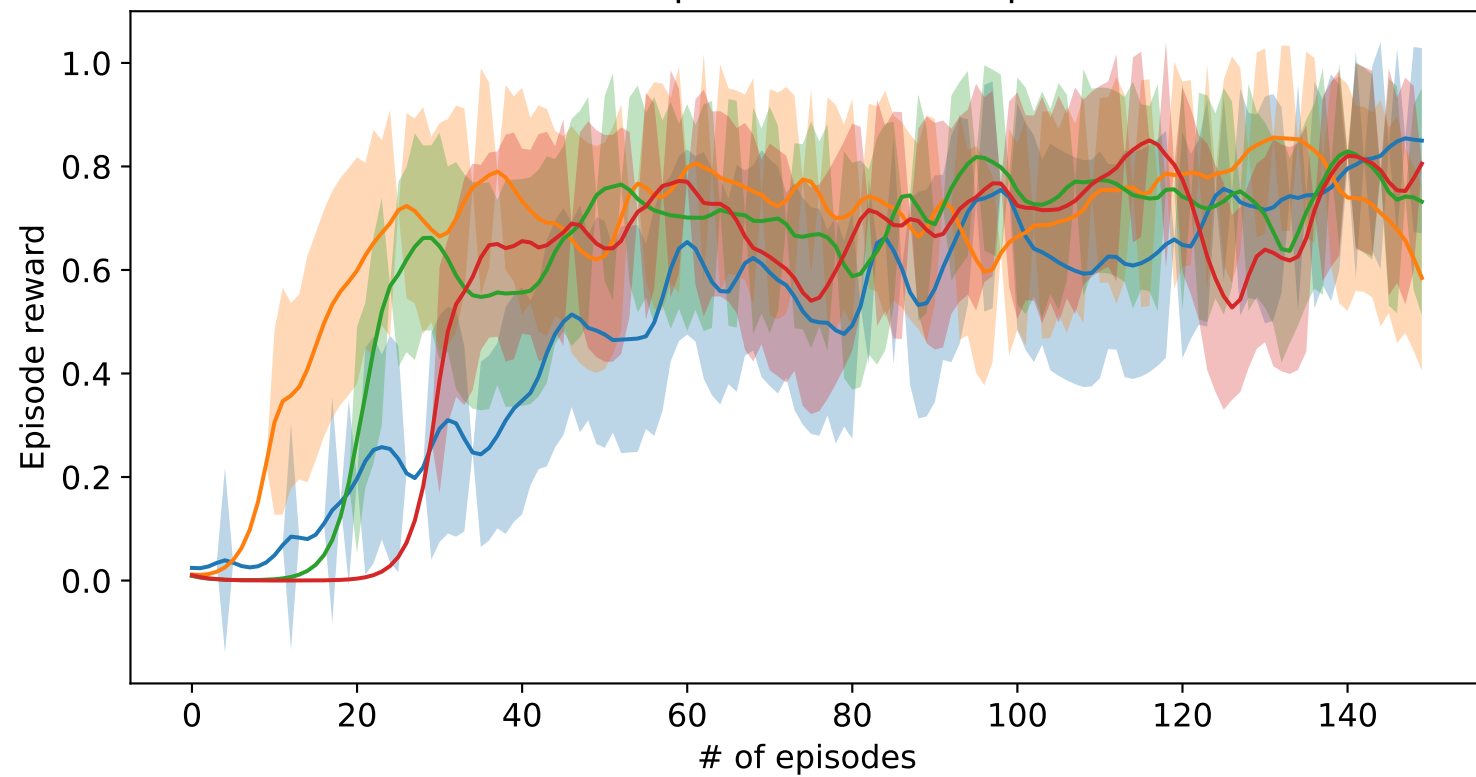Smoothed pretrain 1000 steps, 7B

Smoothed pretrain 3000 steps, 7B

Smoothed pretrain 1000 steps, 32B

Smoothed pretrain 3000 steps, 32B

Legend:
- On-policy. Cum. reward=75
- LORO 10 pretrain eps. Cum. reward=93
- LORO 20 pretrain eps. Cum. reward=88
- LORO 30 pretrain eps. Cum. reward=85

X-axis: # of episodes
Y-axis: Episode reward