CartPole

LORO. Cum. reward=26035
Online RL. Cum. reward=24640
Qwen-7B. Cum. reward=4299
Qwen-32B. Cum. reward=7675
Random. Cum. reward=3104