RepresentedPong

On-policy. Cum. reward=-2088
LORO. Cum. reward=-1764
Pretrain w/ On-policy data. Cum. reward=-2123
Pretrain w/ random policy data. Cum. reward=-1934