intercept-n7-r1 cumulative reward without goal (higher = better) -20-40 **Cumulative reward** -60-80 · ppo dqn -100ddqn random policy 5 10 15 20 25 30 35 40 i-th hundred iteration