**Reward Function Comparison**
**(RL training with lr=2e-4)**