

CliffWalking

Episode reward

0

-100

-200

-300

-400

-500

0

25

50

75

100

125

150

175

200

of episodes

- On-policy. Cum. reward=-37151
- LORO. Cum. reward=-17304
- Qwen-7B. Cum. reward=-241959
- Qwen-32B. Cum. reward=-71680
- Random. Cum. reward=-383500

