TD3 policy\_delay Sweep Training Performance: Mean  $\pm$  Std across Seeds (3 seeds) policy\_delay 1 2 -21 5 50 -22 -23 Return **-**24 -25 -26 20 70 Ó 10 30 40 50 60 Episode