

$\epsilon$ -greedy agent policy - last 500 steps

