Learning rate: 0.00025, batch size: 32, target update: 500

ε-greedy, $\varepsilon = 0.01$

Reward

Episode