

## 1. Behaviour Cloning

--ep len is 1000  
-- eval batch size is 5000  
-- training steps: 1000  
-- Size of hidden layers: 64  
-- Learning rate: 5e-3

Env	Train_Mean	Eval_Mean	Eval_Std	performance
Ant	4713.65	4813.08	94.44	97.89%
Humanoid	10344.51	316.49	49.47	3.05%

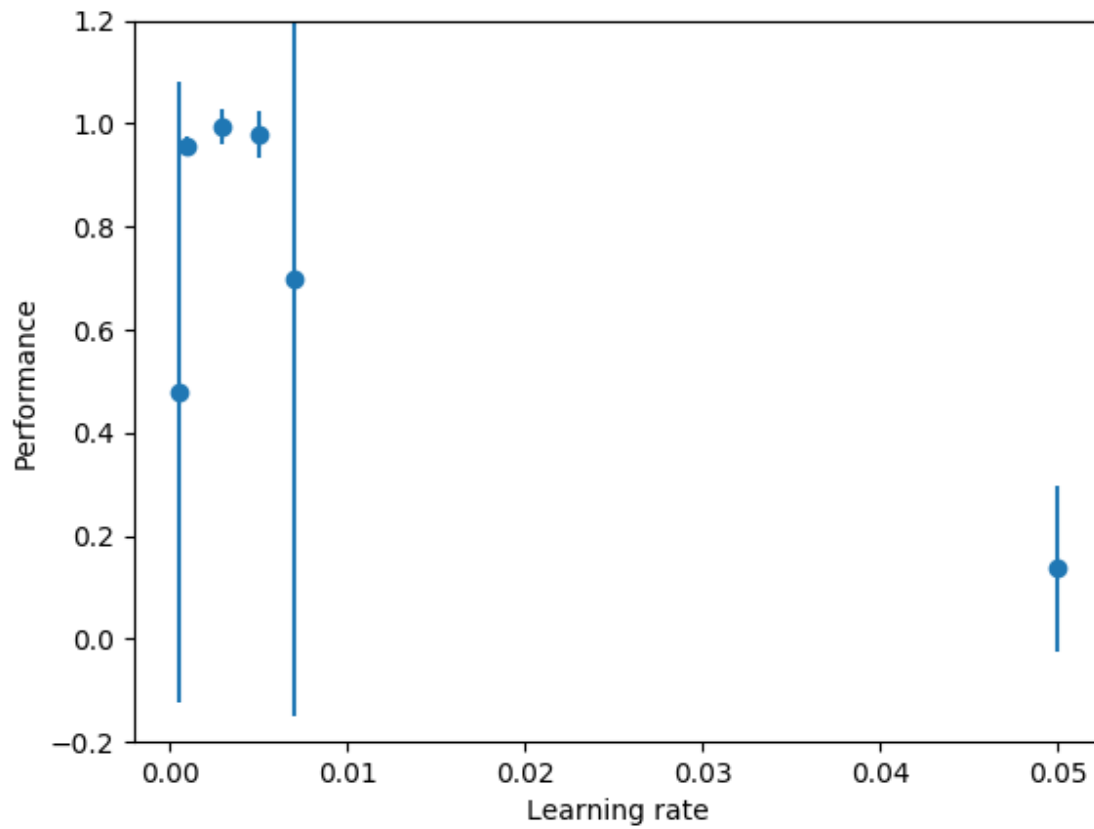


Fig. 1: The best rl for the Ant env is  $3e-3$  with 99.44% performance and 66.98 variance

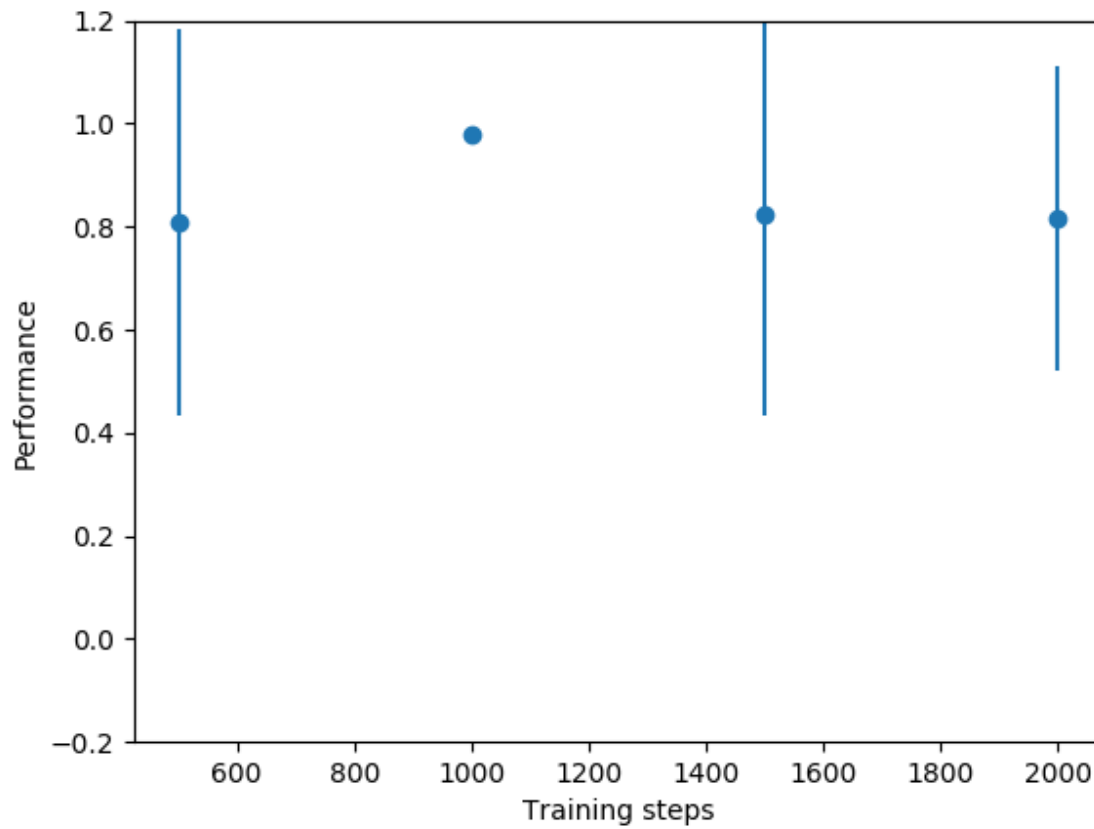


Fig. 1: The best Training steps for the Ant env is 1000 with 97.89% performance and 94.89 variance

## 2. DAgger

--ep\_len is 1000  
-- eval batch size is 5000  
-- training steps: 1000  
-- Size of hidden layers: 64  
-- Learning rate: 5e-3

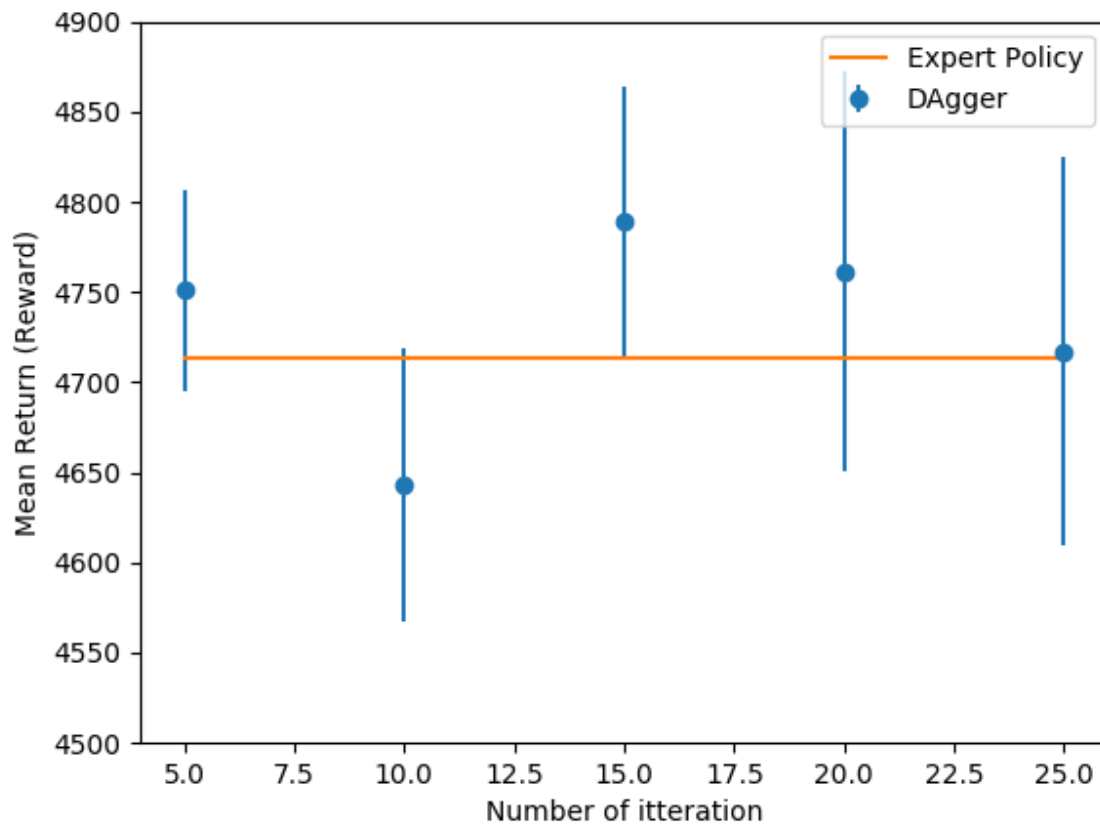
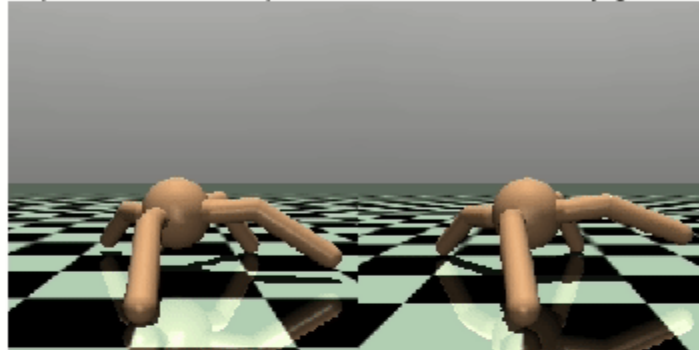


Fig 2. Policy's mean return vs the number of iterations with expert policy in the Ant environment with the above-mentioned configuration

### eval\_rollouts

eval\_rollouts  
step 5

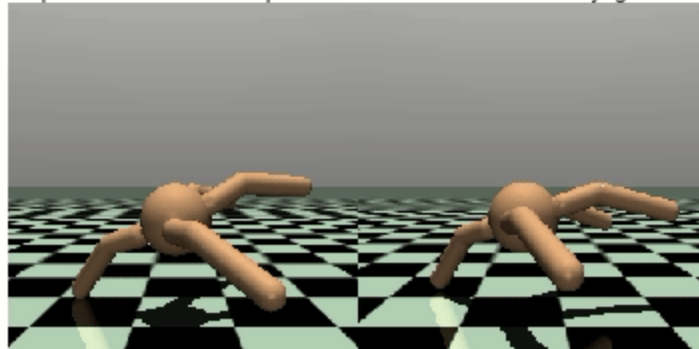
Wed Apr 27 2022 14:07:29 Pacific Daylight Time



### train\_rollouts

train\_rollouts  
step 5

Wed Apr 27 2022 14:07:27 Pacific Daylight Time



--ep\_len is 1000  
-- eval batch size is 5000  
-- training steps: 1000  
-- Size of hidden layers: 64  
-- Learning rate: 5e-3

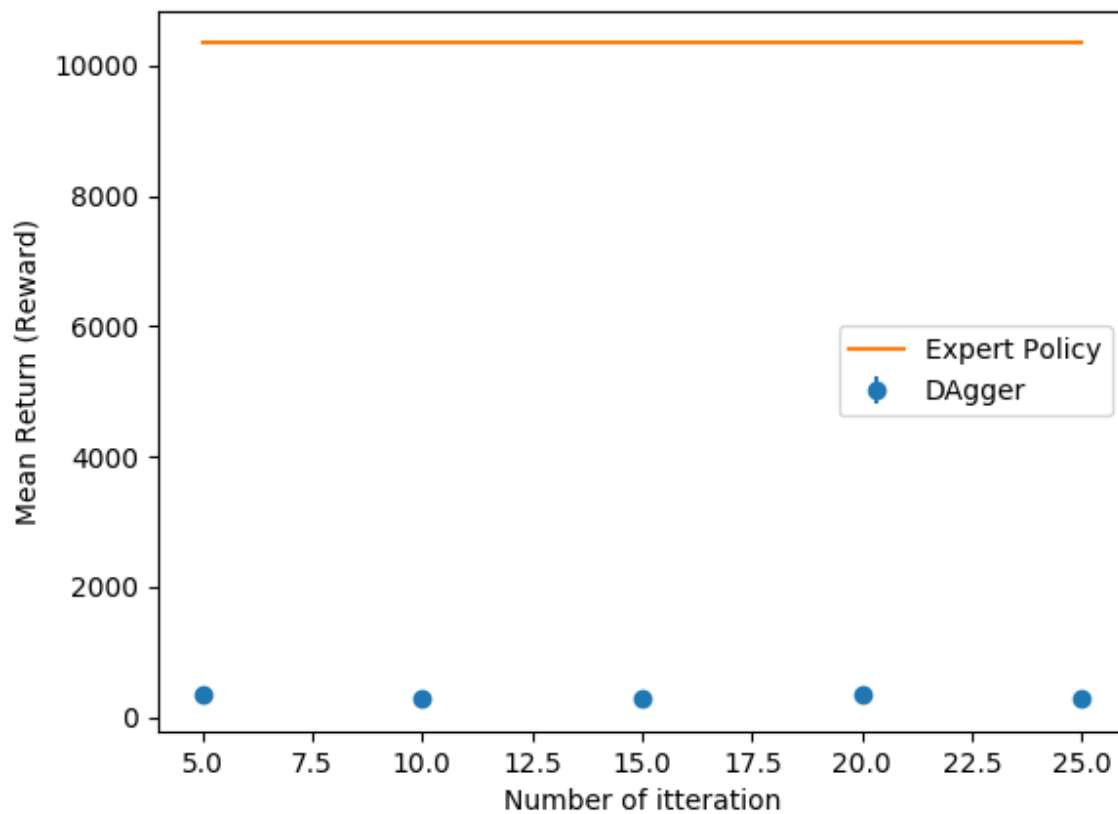


Fig 2. Policy's mean return vs number of iterations with expert policy in the Humanoid environment with the above-mentioned configuration