

Here, I document some simple runs for Gaussian behavioral clones for a PPO Agent (the "expert" or "oracle") trained in the easiest Safety Gym PointGoal environment.

Tyna +

## The Expert

The expert in question is a simple Gaussian MLP Actor-Critic architecture with 4 [128x128] layers. Marigold is trained for 1000 epochs, running 20 episodes with a maximum of 1,000 steps in each epoch.

Furthermore, we train this agent with a time discounting parameter  **$\gamma = 0.99$**  and **GAE  $\Lambda = 0.98$** .

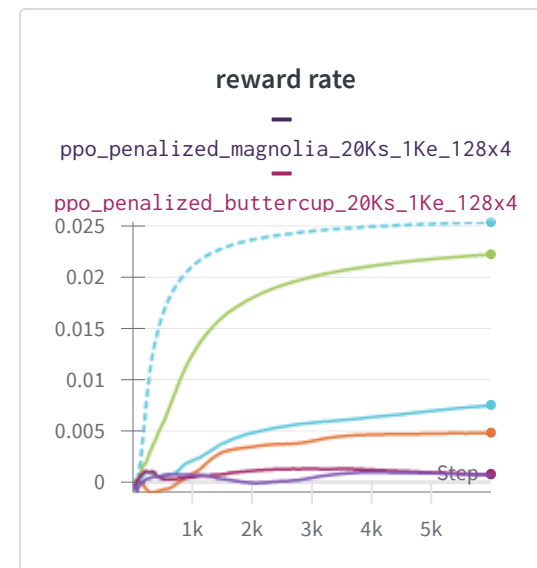
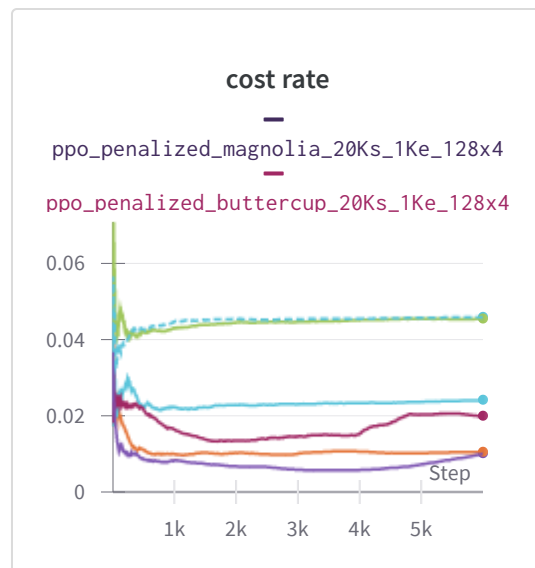
Lastly, as we aim to limit Marigold's average total episode costs to some parameter, which we call the *cost limit*. In this case, we set this parameter to 50, and set a learning rate for this parameter to 0.025. Here is an example of cost and return outcomes for some agents trained at different cost limits.

Note that there are no explicitly set reward parameters, only implicitly learned. We notice a meaningful **exploitation/exploration** trade-off in the sense that highly-constrained agents tend to



be more conservative in their exploration and therefore less successful than their less constrained counterparts. This is in spite of attempting to tease out reward-seeking behavior by increasing **gamma**. For these risk-averse agents, the penalty for making mistakes is simply too great to overcome. This may be further explored by capturing some metric of **occupancy measure** in future reports.

Also note that, as usual, performance can strongly depend on hyper-parameters. Observe how the *Lilly* and *Marigold* agents observe the same cost limit, but achieve very different reward performance. Some of these discoveries were aided by Weights&Biases parameter sweeps (in my profile), but many more by trial and error.



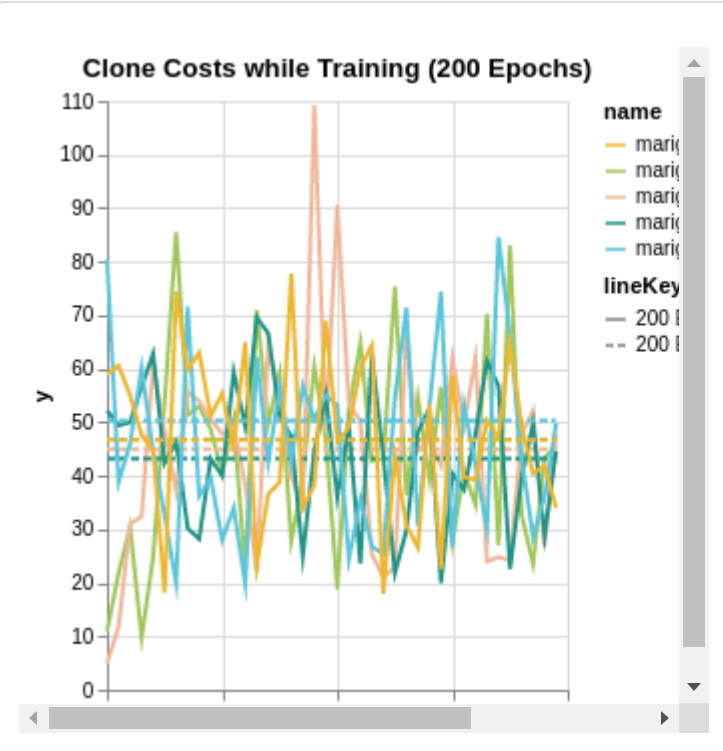
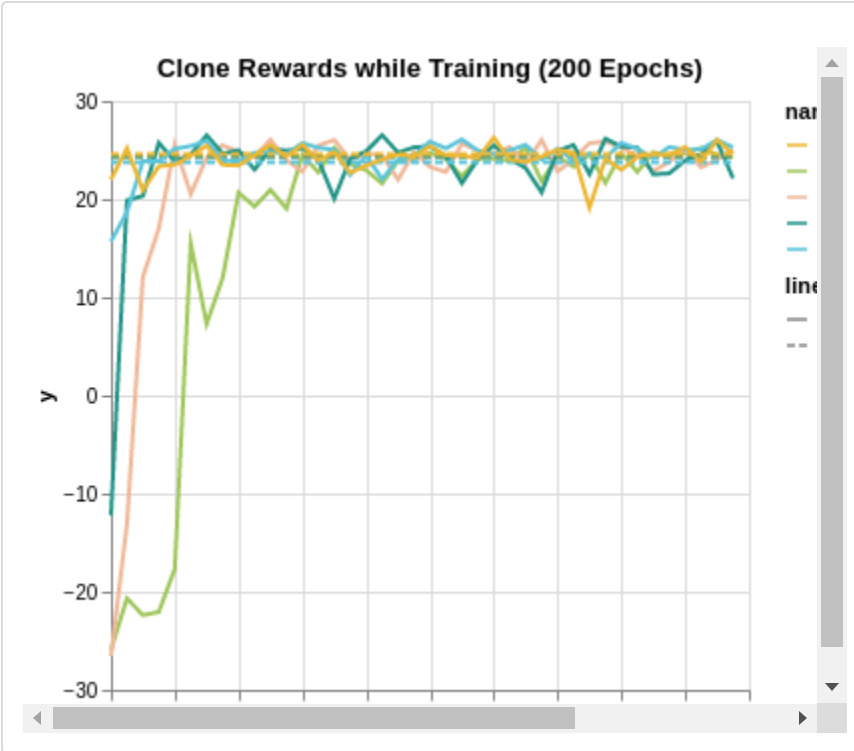
Add panel



# The Clones

Clones in this case are simply Gaussian MLPs learning actions in a supervised manner from the given expert's experiences. This is *behavioral cloning*. The pictures below summarize some of the in-training performance of the supervised clones under different training regimens; I vary the number of epochs and number of training sample per epoch. Note that the clones in question do not have access to the environment, and observe only the states and actions that their designated expert observed in the course of play. The clones each have access to expert trajectories for **1000** episodes, and do not observed costs, reward or any other auxiliary environmental signal.





Add panel

☒
Run set 2 5

openai-scholars / clone\_benchmarking\_marigold

















Filter

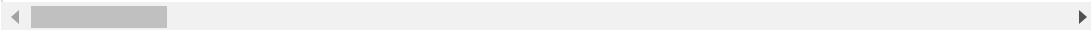
Group

Sort

Tag

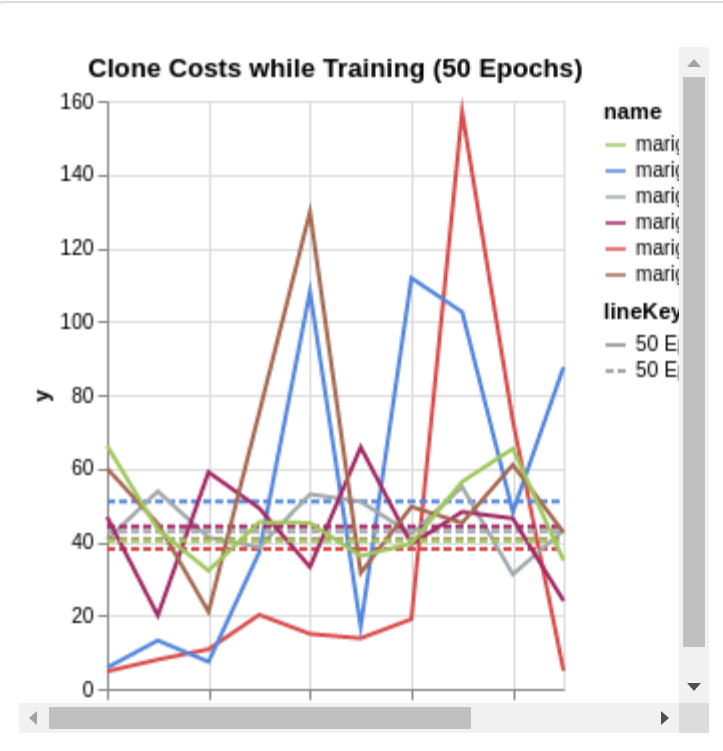
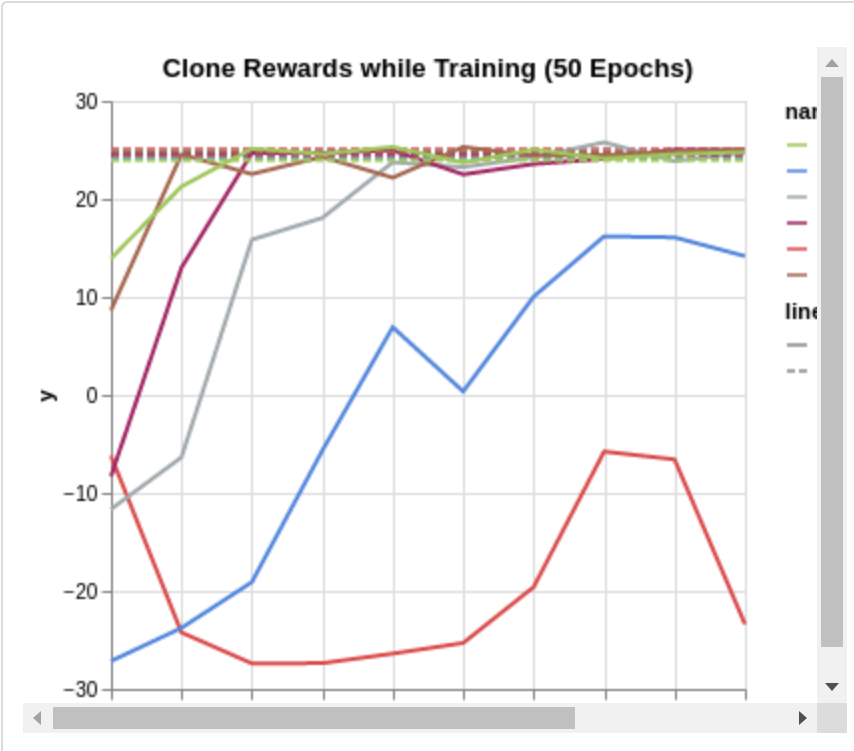
<input type="checkbox"/> Name (5 visualized)	State	Notes	User	Tags	Created ▾	Runtime	Sweep
<div> <div> <div>•</div> <div> </div> <div>marigold_clone_200€</div> </div> </div>	finished	Add notes	felounc		9h ago	6m 58s	-
<div> <div> <div>•</div> <div> </div> <div>marigold_clone_200€</div> </div> </div>	finished	Add notes	felounc		9h ago	7m 23s	-

•   marigold_clone_200€	finished	Add notes	felounc	9h ago	7m 50s	-
•   marigold_clone_200€	finished	Add notes	felounc	9h ago	7m 35s	-
•   marigold_clone_200€	finished	Add notes	felounc	9h ago	7m 39s	-
•   marigold_clone_25ep	finished	Add notes	felounc	10h ago	1m 50s	-
•   marigold_clone_25ep	finished	Add notes	felounc	10h ago	1m 47s	-
•   marigold_clone_25ep	finished	Add notes	felounc	10h ago	1m 37s	-
•   marigold_clone_25ep	finished	Add notes	felounc	10h ago	1m 41s	-
•   marigold_clone_25ep	finished	Add notes	felounc	10h ago	1m 41s	-



1-10▼ of 35 < >



















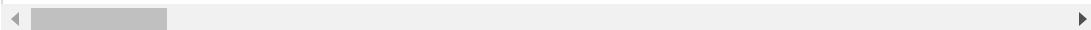


Add panel

☒ Run set 2 6

<input type="checkbox"/> Name (6 visualized)	State	Notes	User	Tags	Created ▾	Runtime	Sweet
marigold_clone_200€	finished	Add notes	felounc		9h ago	6m 58s	-
marigold_clone_200€	finished	Add notes	felounc		9h ago	7m 23s	-

•   marigold_clone_200€	finished	Add notes	felounc	9h ago	7m 50s	-
•   marigold_clone_200€	finished	Add notes	felounc	9h ago	7m 35s	-
•   marigold_clone_200€	finished	Add notes	felounc	9h ago	7m 39s	-
•   marigold_clone_25ep	finished	Add notes	felounc	10h ago	1m 50s	-
•   marigold_clone_25ep	finished	Add notes	felounc	10h ago	1m 47s	-
•   marigold_clone_25ep	finished	Add notes	felounc	10h ago	1m 37s	-
•   marigold_clone_25ep	finished	Add notes	felounc	10h ago	1m 41s	-
•   marigold_clone_25ep	finished	Add notes	felounc	10h ago	1m 41s	-



1-10▼ of 35 < >





Add panel

Add panel grid

Add text block

