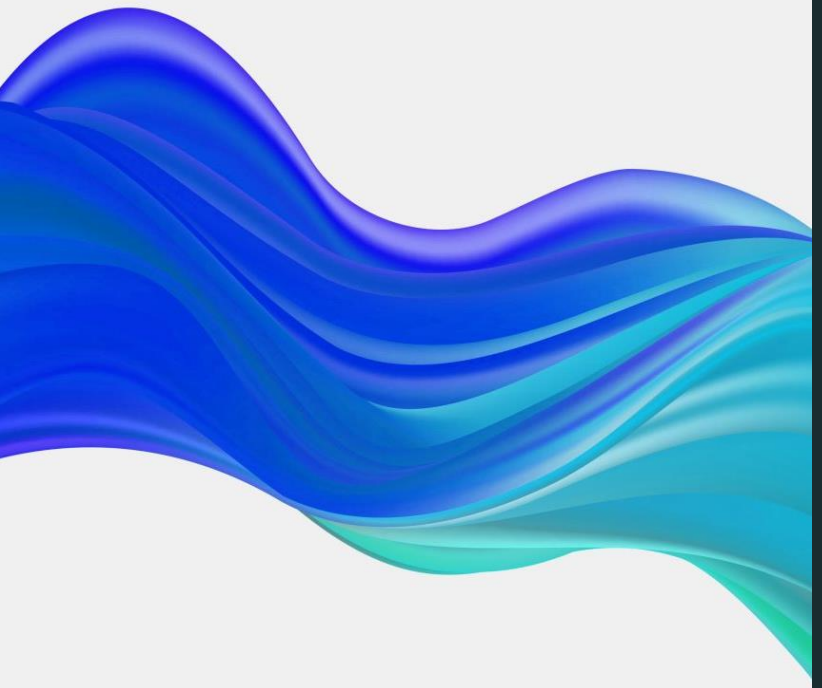


COMPONENT-BASED DESIGN: REINFORCEMENT LEARNING ALGORITHMS TAILORED FOR THE LUNAR LANDER GYM ENVIRONMENT

CS 6376 Class Project

-Agrima Khanna

THE LUNAR LANDER ENVIRONMENT

Action Space:
- Discrete or Continuous

Observation Space:
-The state is an 8-dimensional vector

Rewards
- An episode's reward is a measure of the success of the lander.

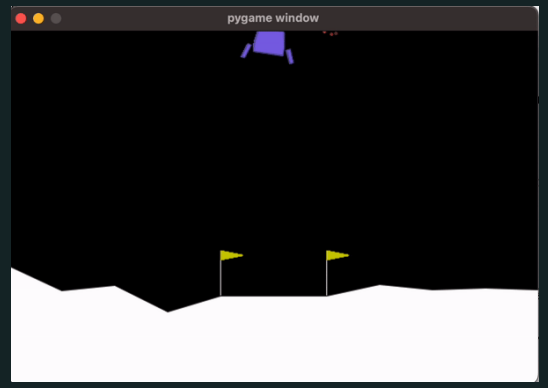
Starting State
- The lander starts at the top centre of the viewport with a random initial force applied to its centre of mass.

Episode Termination
- The lander is supposed to land between two flags at (0,0)

SAFETY AND LIVENESS REQUIREMENTS

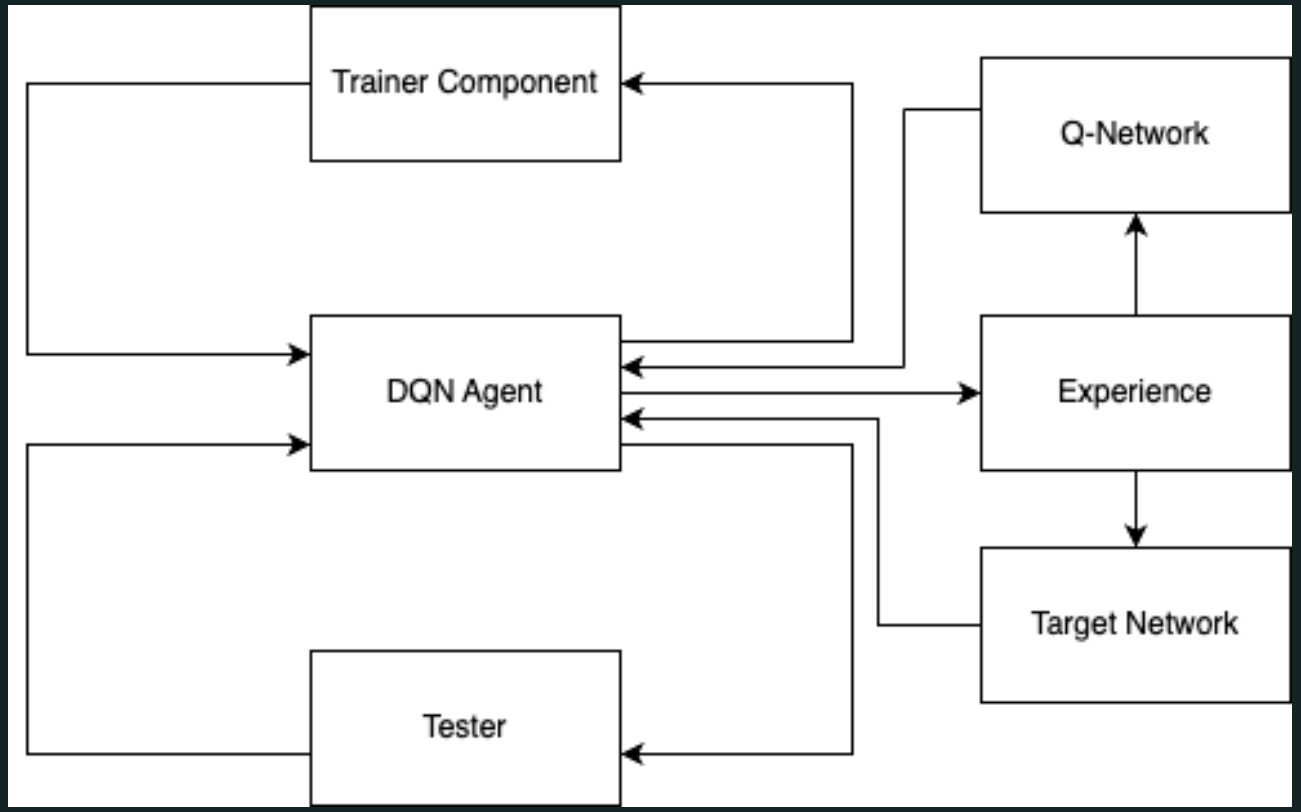
Safety Requirements

Liveness Requirements



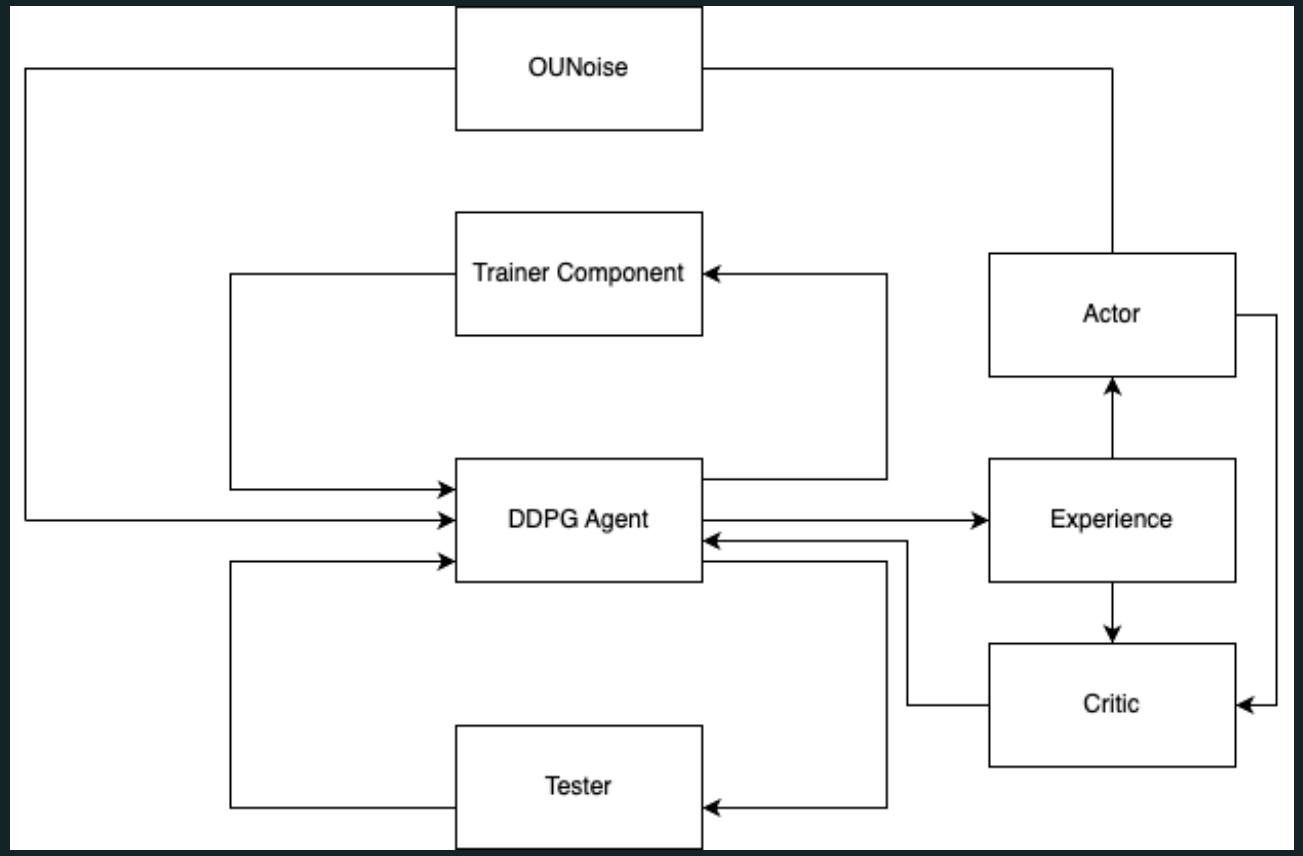
DEEP Q-NETWORK

- Trained on the Lunar Lander environment with discrete action space
- Components the DQN Agent interacts with:
  1. *Q-Network*
  2. *Target Network*
  3. *Experience Replay Buffer*
- The Agent component interacts with the three components and the trainer component.



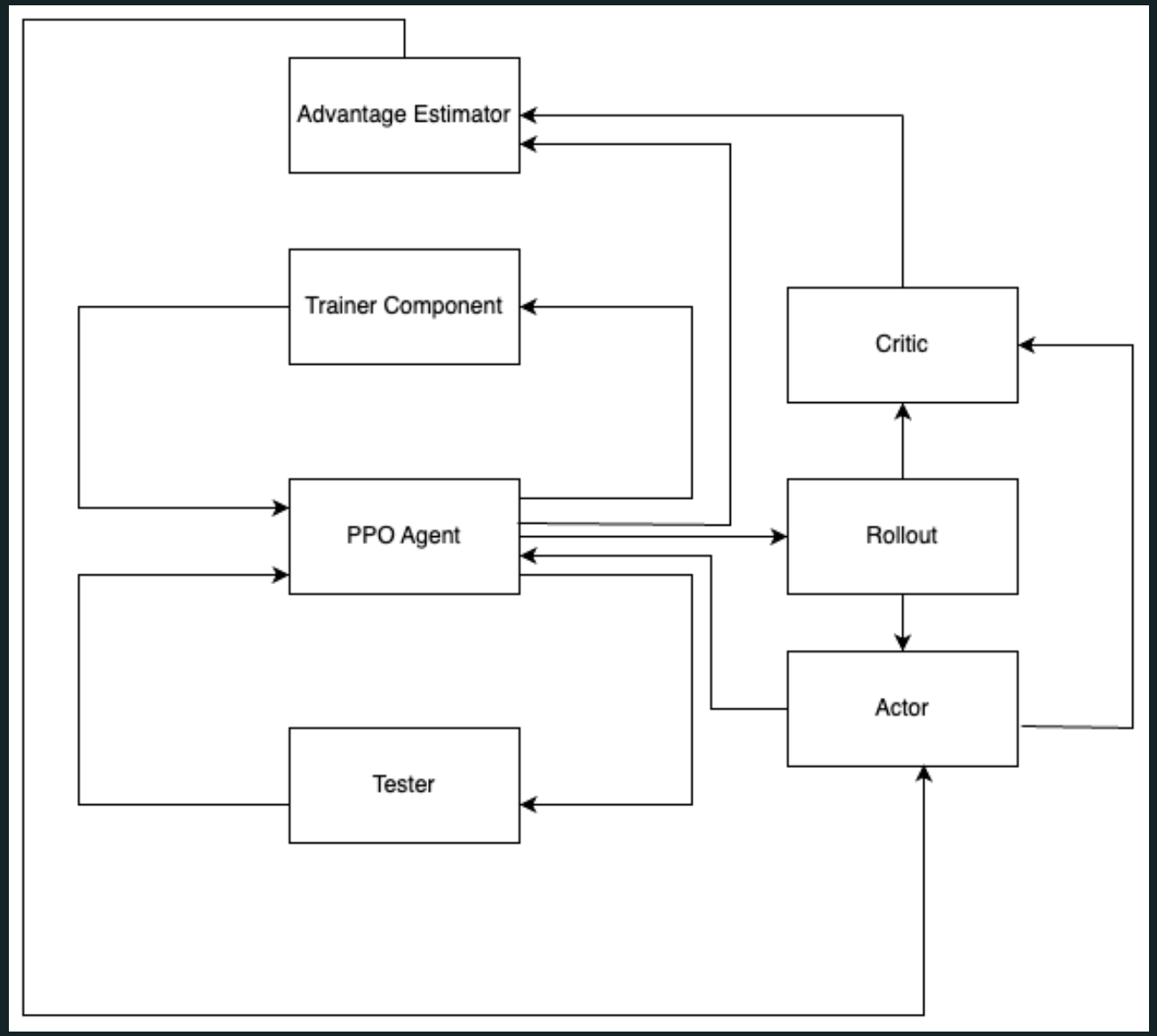
DEEP DETERMINISTIC POLICY GRADIENT

- Trained on the Lunar Lander with a continuous action space.
- Components the DDPG Agent interacts with:
  1. *Actor Network (and the Target Actor Network)*
  2. *Critic Network (and the Target Critic Network)*
  3. *Ornstein-Uhlenbeck Noise*
  4. *Replay Buffer*



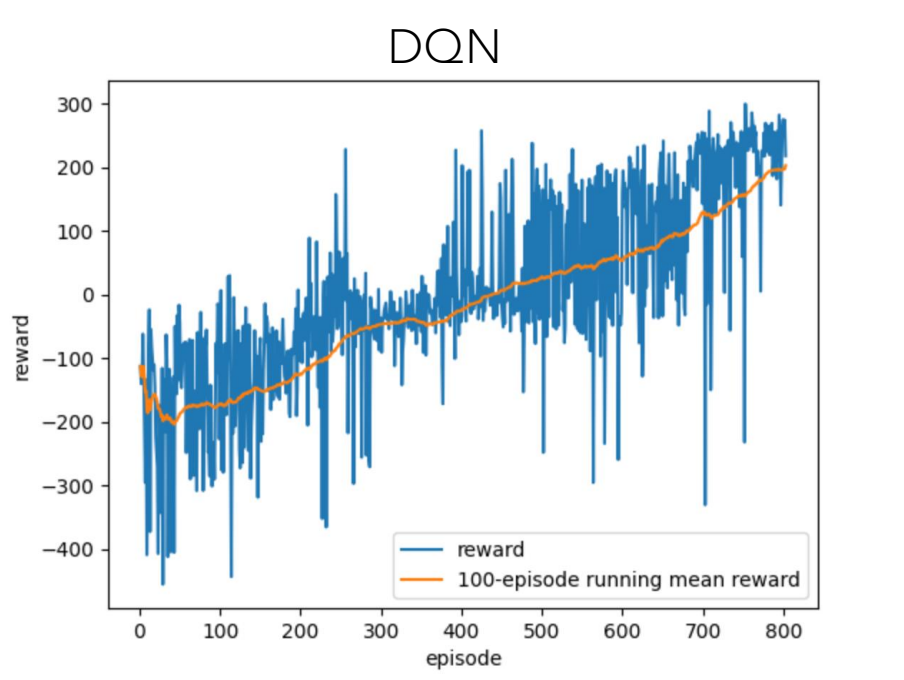
PROXIMAL POLICY OPTIMIZATION

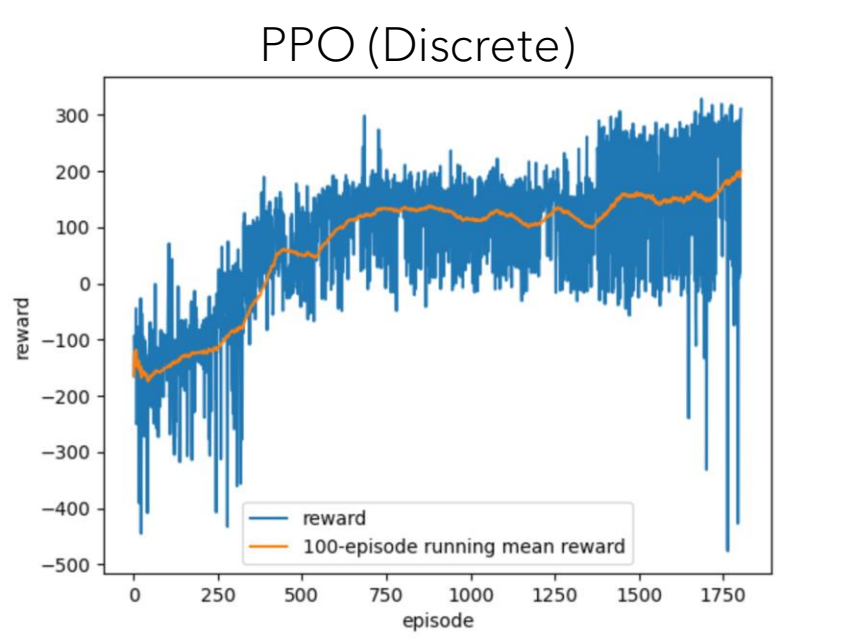
- Trained for the both types of action spaces.
- The PPO Agent interacts with:
  1. *Actor*
  2. *Critic*
  3. *Advantage Estimator*
  4. *Rollout Buffer*

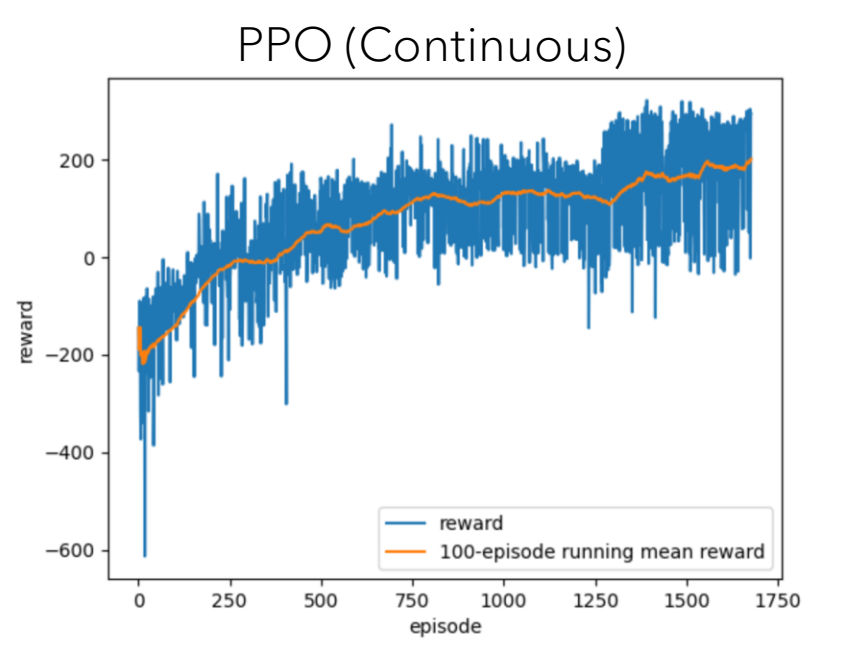


RESULTS

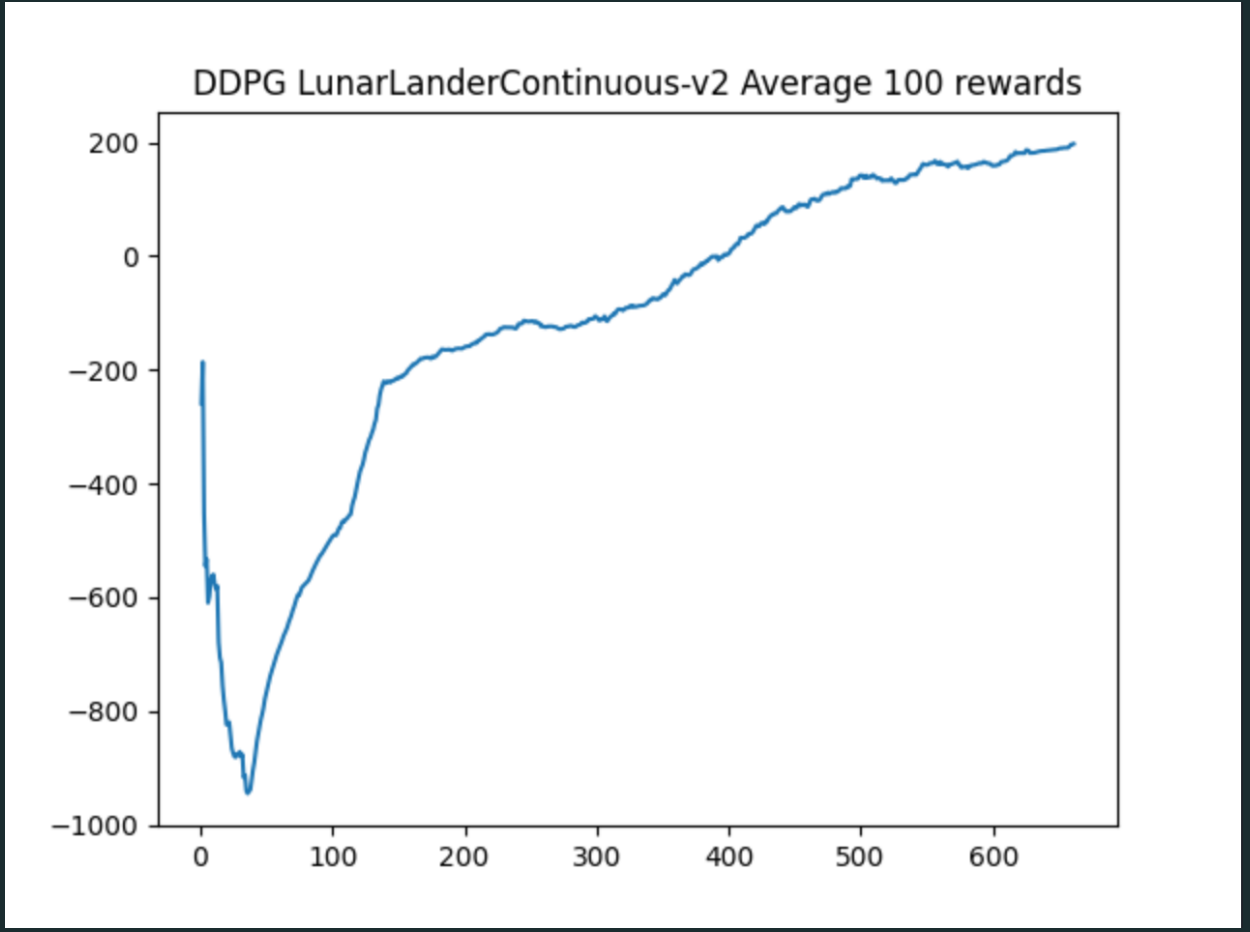
Training results





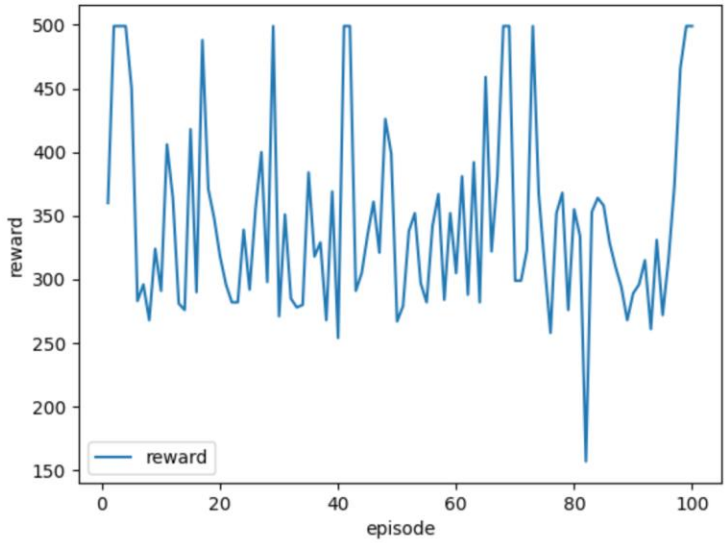


DDPG

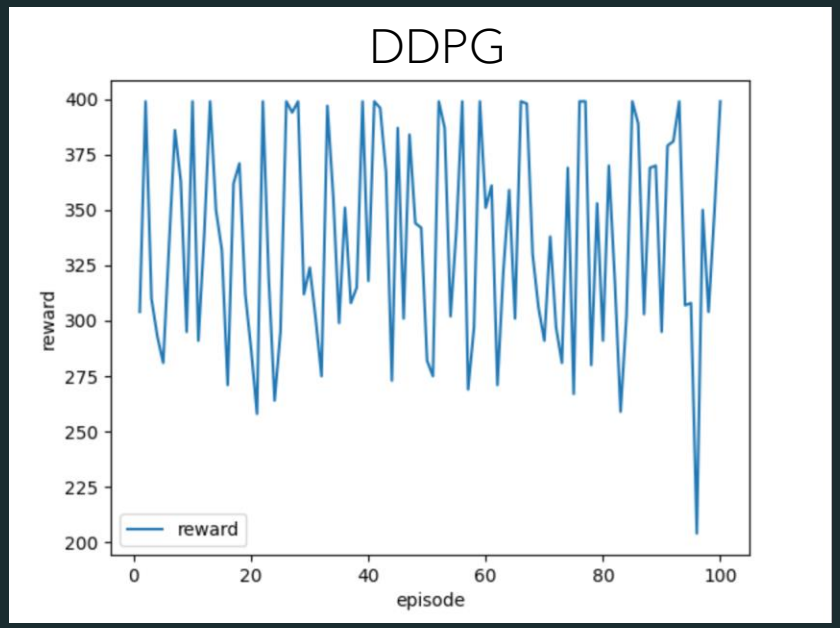

DDPG LunarLanderContinuous-v2 Average 100 rewards
200
0
−200
−400
−600
−800
−1000
0
100
200
300
400
500
600

Testing results

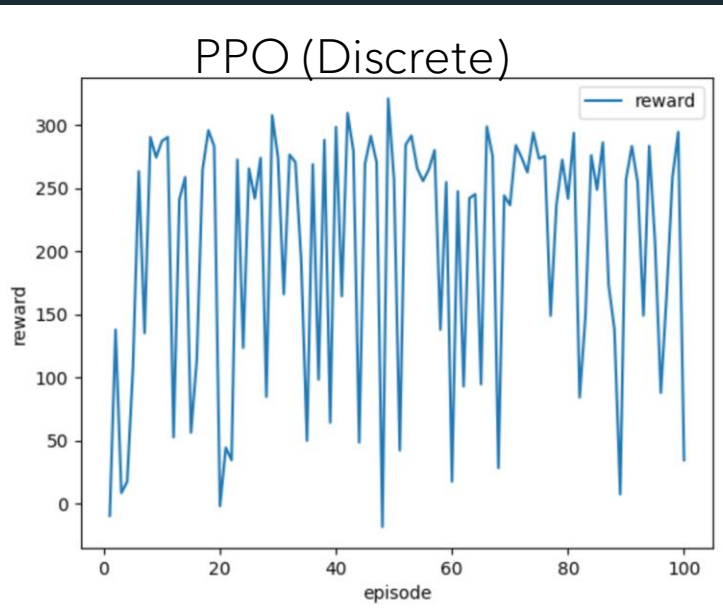
DQN



DDPG



PPO (Discrete)



PPO (Continuous)