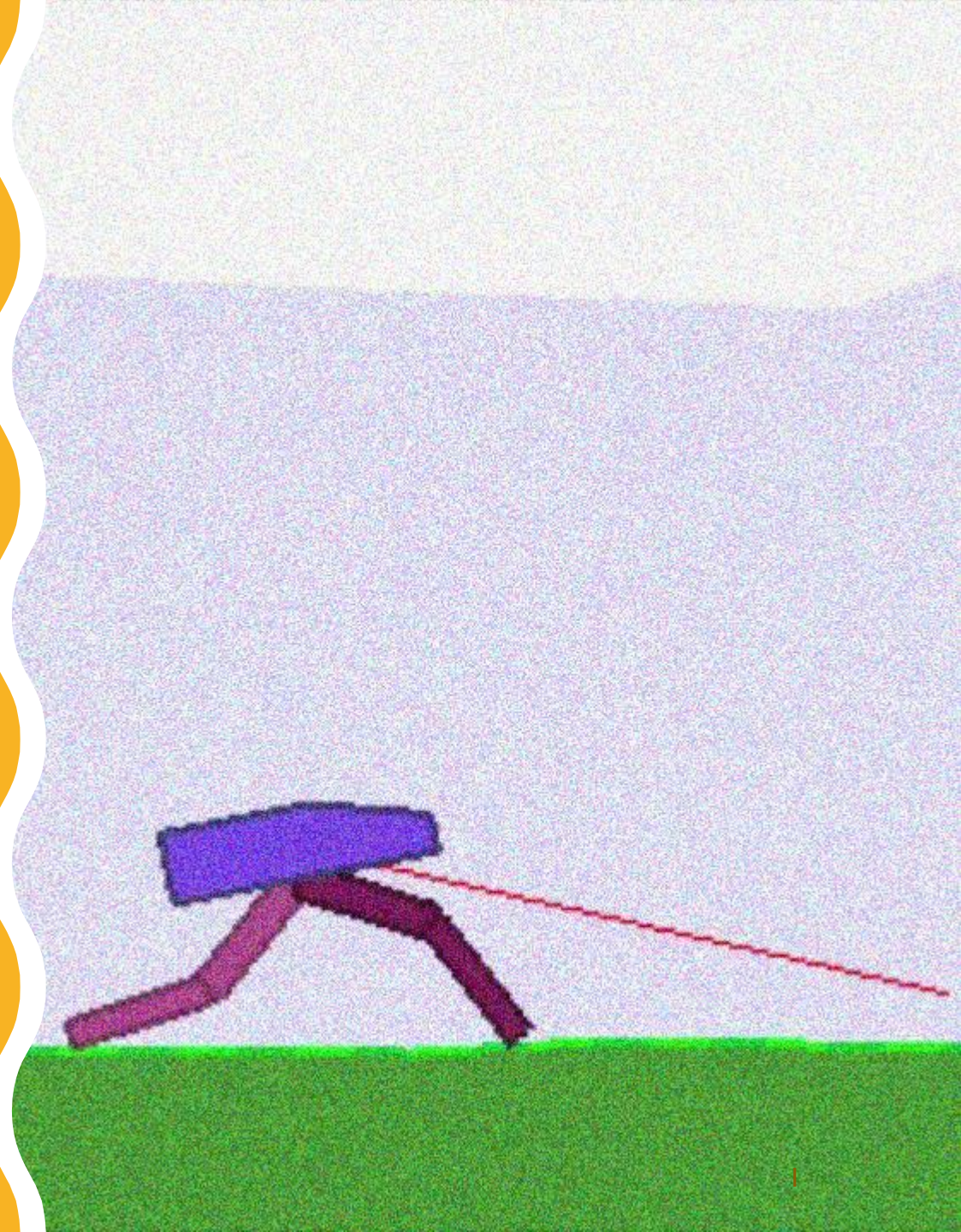


# AI

## BIPEDAL WALKER V3

Iliya Mirzaei - AmirMahdi Shadman





**ENV**

## **BOX2D ENVIRONMENTS**

Based on Gymnasium documentation

# BOX2D ENVIRONMENTS

## BIPEDAL WALKER V3

Action Space	Box(-1.0, 1.0, (4,), float32)
Observation Shape	(24,)
Observation High	[3.14 5. 5. 5. 3.14 5. 3.14 5. 5. 3.14 5. 3.14 5. 5. 1. 1. 1. 1. 1. 1. 1. 1.]
Observation Low	[-3.14 -5. -5. -5. -3.14 -5. -3.14 -5. -0. -3.14 -5. -3.14 -5. -0. -1. -1. -1. -1. -1. -1. -1. -1.]
Import	<code>gymnasium.make("BipedalWalker-v3")</code>



## BOX2D ENVIRONMENTS EXPLANATION

### Description

This is a simple 4-joint walker robot environment. There are two versions:

Normal, with slightly uneven terrain.

To solve the normal version, you need to get 300 points in 1600 time steps.

### Action Space

Actions are motor speed values in the  $[-1, 1]$  range for each of the 4 joints at both hips and knees.

### Observation Space

State consists of hull angle speed, angular velocity, horizontal speed, vertical speed, position of joints and joints angular speed, legs contact with ground, and 10 lidar rangefinder measurements. There are no coordinates in the state vector.

### Rewards

Reward is given for moving forward, totaling 300+ points up to the far end. If the robot falls, it gets -100. Applying motor torque costs a small amount of points. A more optimal agent will get a better score.

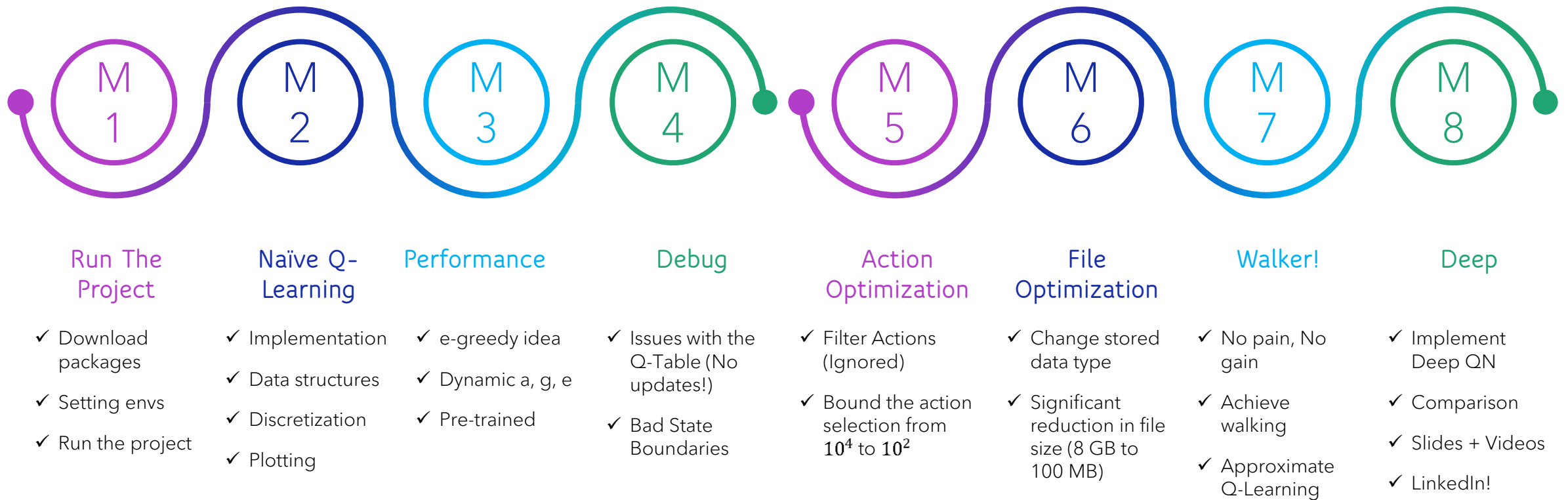
### Starting State

The walker starts standing at the left end of the terrain with the hull horizontal, and both legs in the same position with a slight knee angle.

### Episode Termination

The episode will terminate if the hull gets in contact with the ground or if the walker exceeds the right end of the terrain length.

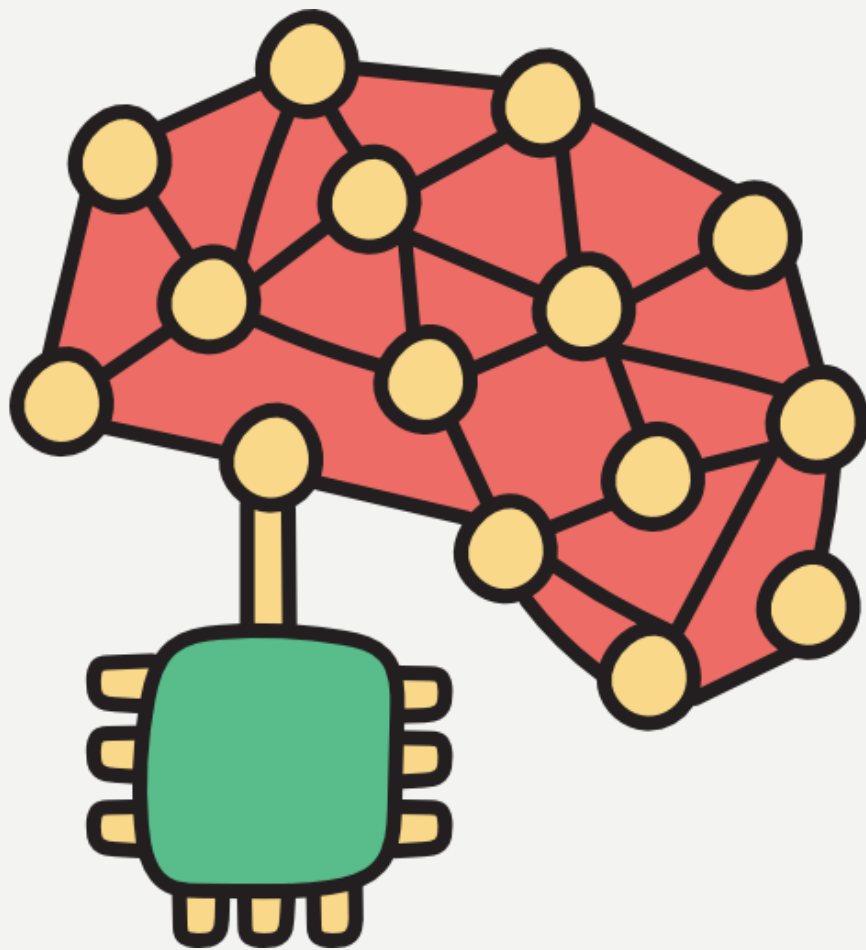
# ROADMAP MILESTONES





## MEMORY OPTIMIZATION

- ✓ Discretize states and actions
- ✓ Bound action selection
- ✓ Reduce the file size



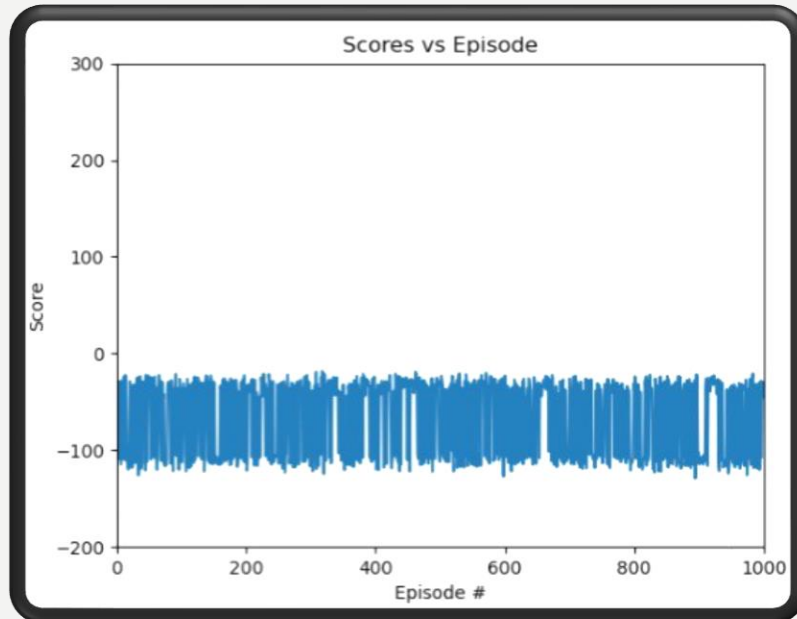
## BETTER AGENT

- ✓ Pre-trained agent
- ✓ Episode steps number
- ✓ Deep QN

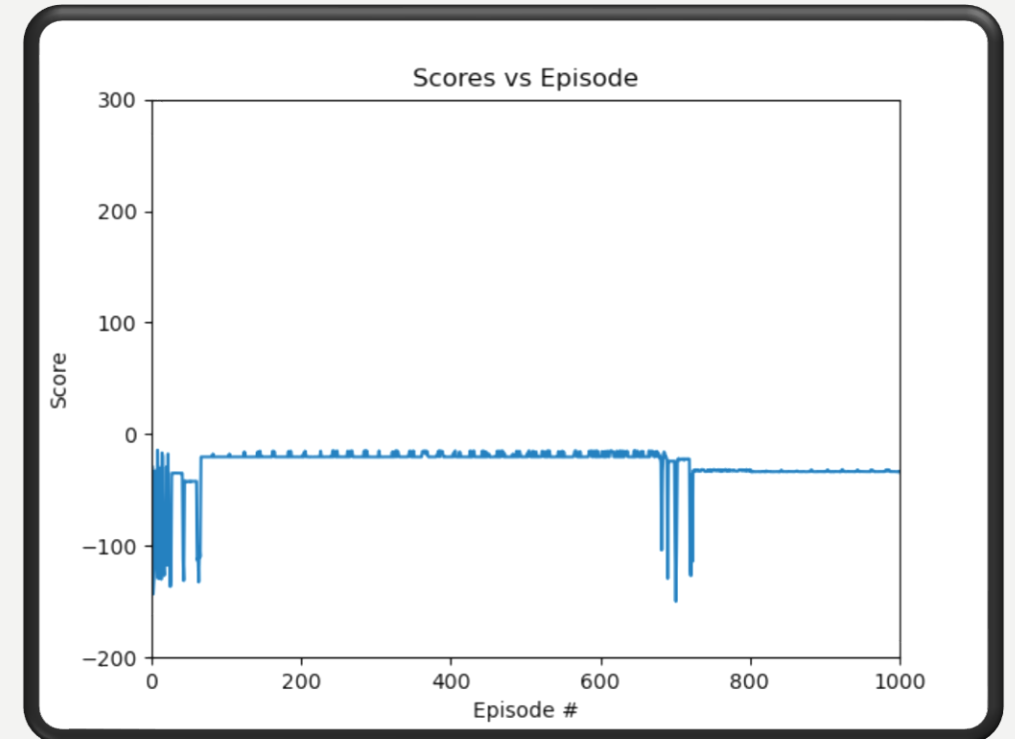


# Q-LEARNING IMPROVEMENTS

- Before

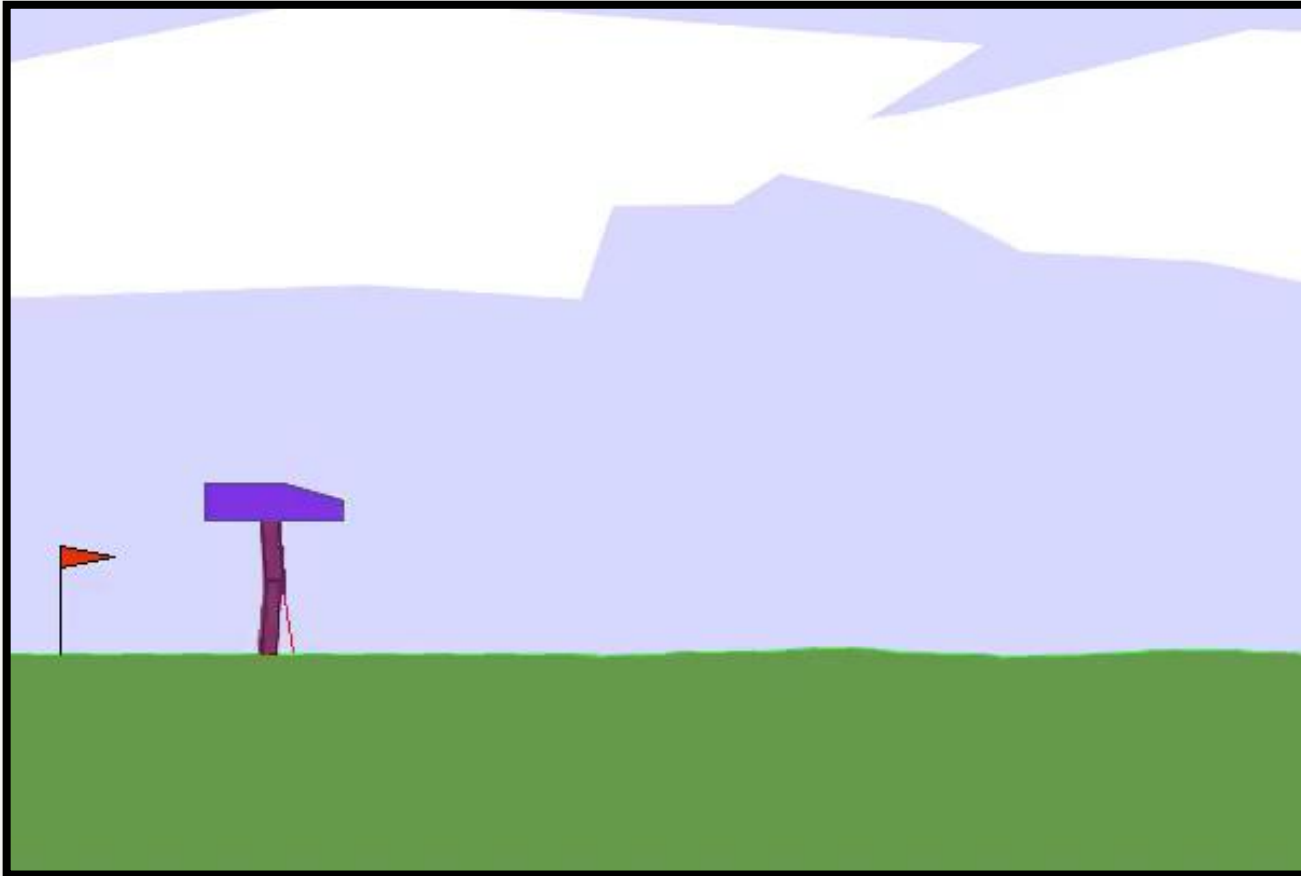


- After



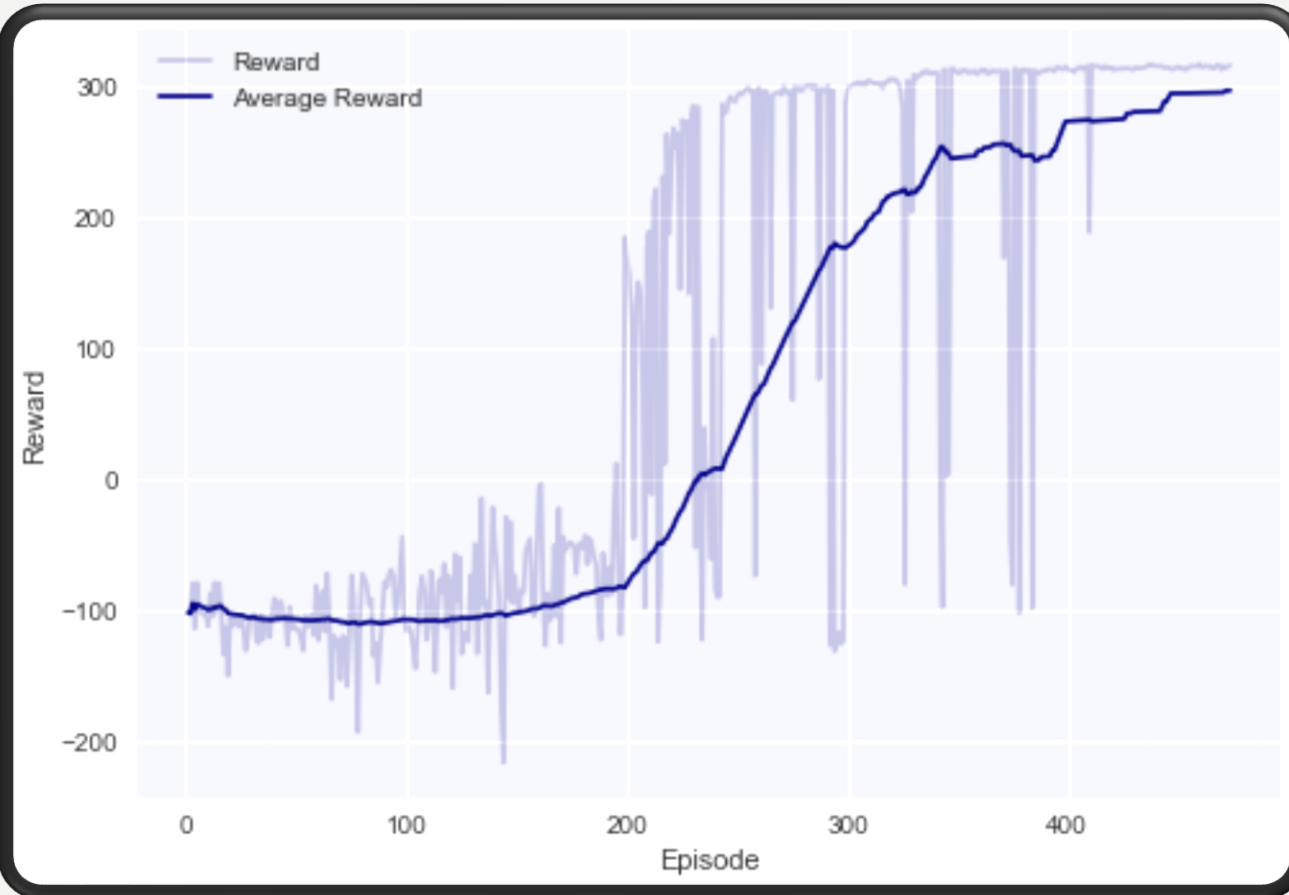


## FINAL Q-LEARNING RESULT VIDEO



- ✓ Memory Optimization
- ✓ Dynamic Alpha and Epsilon
- ✓ Playing with Episode Step Number
- ✓ Pre-trained

## DEEP Q-NETWORK



# Q-LEARNING VS DEEP Q-NETWORK

## ❖ Pure Q-Learning

- Can reach almost maximum average score -50 within 100,000 episodes

## ❖ Deep QN

- Can reach maximum average score +300 within less than 500 episodes