# Reinforcement Learning for Game Environment

Proximal Policy Optimization

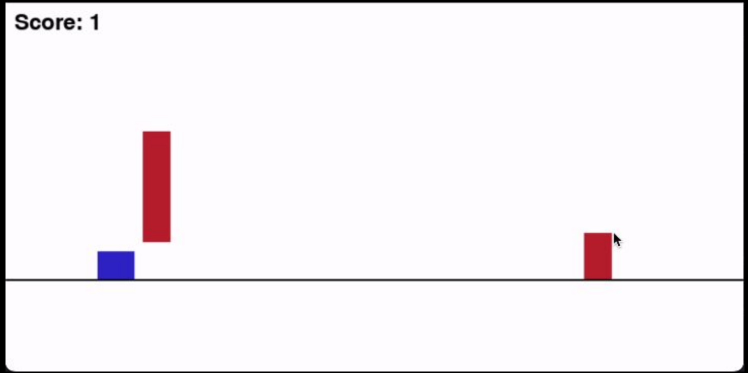
Team GAIL RL-2025

# Environment Overview

The **Environment** simulates an obstacle-navigation task inspired by the Chrome dinosaur game. The agent *must jump or squat to overcome obstacles to maximize total reward while progressing forward.*

## Key Features:

- **Agent:** Moves at a fixed speed along the x-axis.
- **Obstacles:** Appear at varying distances.
- **Actions:** jump or squat.
- **Termination:** Collision or goal completion.



# Markov Decision Process (MDP)

## State Space

A **6-dimensional** observation vector:

1. Player height
2. Jump/squat state (ternary, 1 - is jumping, 2 - is squatting)
3. Distance to next obstacle
4. Distance to second obstacle
5. Type of next obstacle
6. Type of second obstacle

## Action Space

Discrete choices:

- **0:** No jump/squat
- **1:** Jump
- 2: Squat

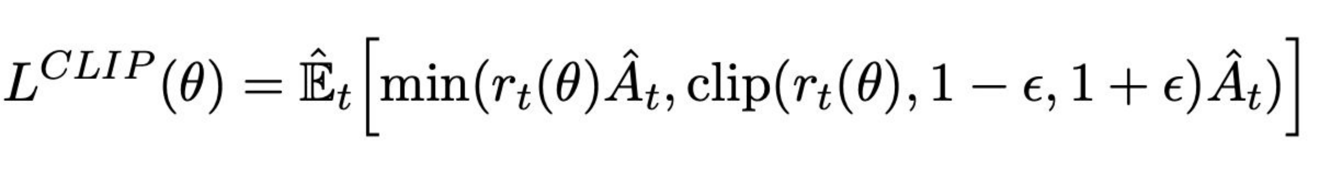
## Reward Function

- **+1** per step survived
- **+50** for passing an obstacle
- **-50** for collis**ion**
- **+100** for reaching the goal
- **-2** per jump or squat (penalizing unnecessary jumps/squats)

## Terminal Conditions

- Collision with an obstacle
- Goal reached

# Proximal Policy Optimization

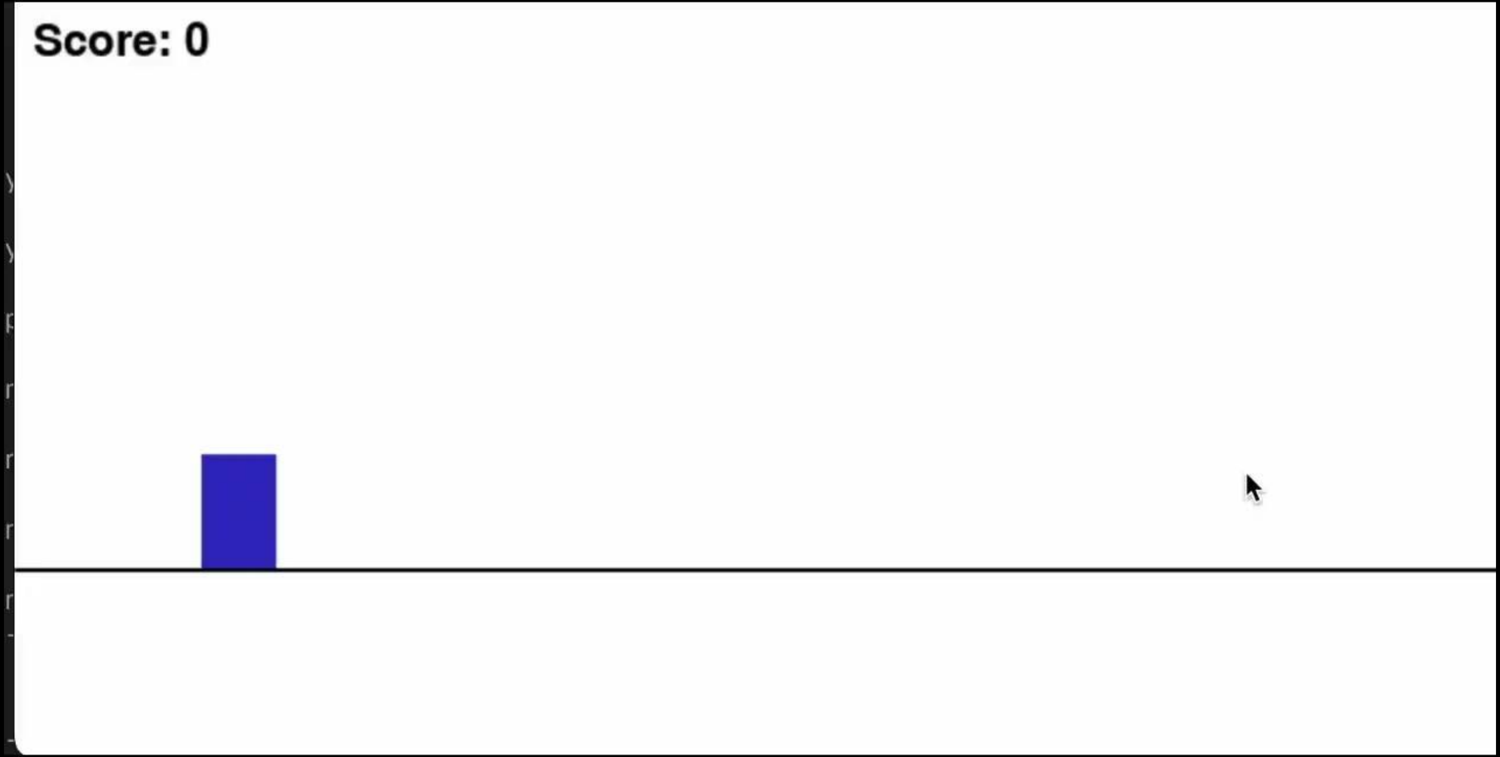


$$L^{CLIP}(\theta) = \hat{\mathbb{E}}_t\left[\min(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)\right]$$

**Main hyperparameters:**

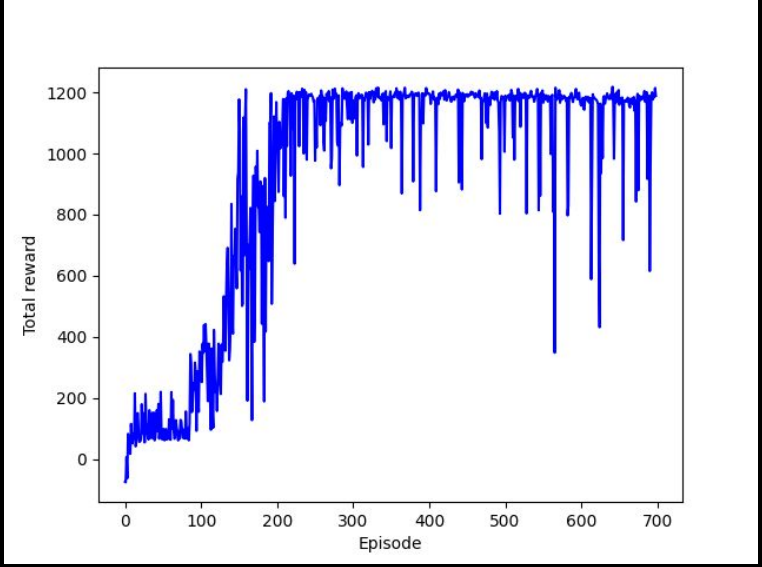
- GAE(0.95)
- Clip Range: 0.2

# Result

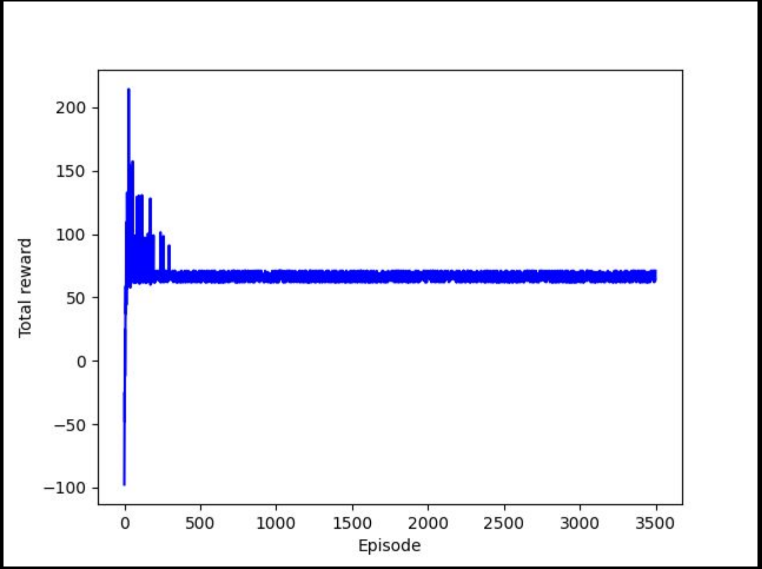


Score: 0

# Training Curves



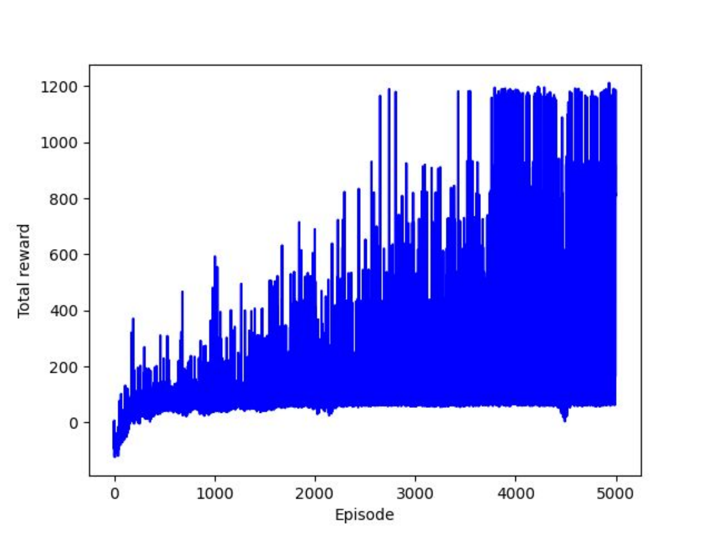
Mainly converge fast



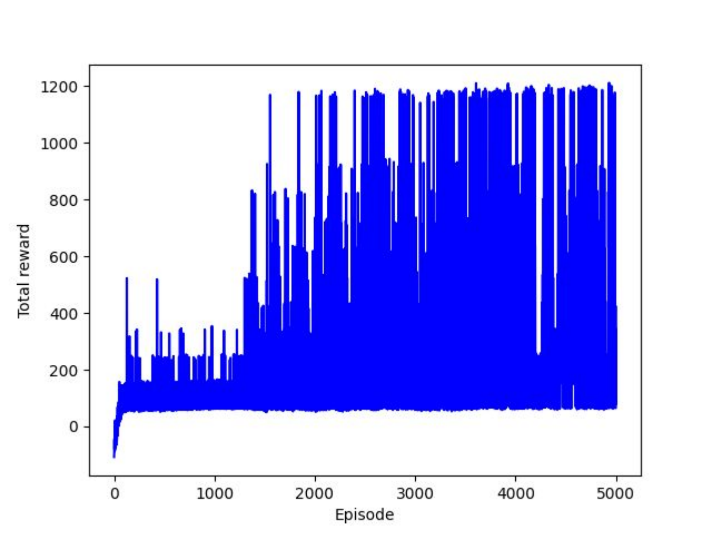
Stuck at some seeds

# Comparison to REINFORCE

Shallow setup: **sample single trajectory per update**



REINFORCE



PPO