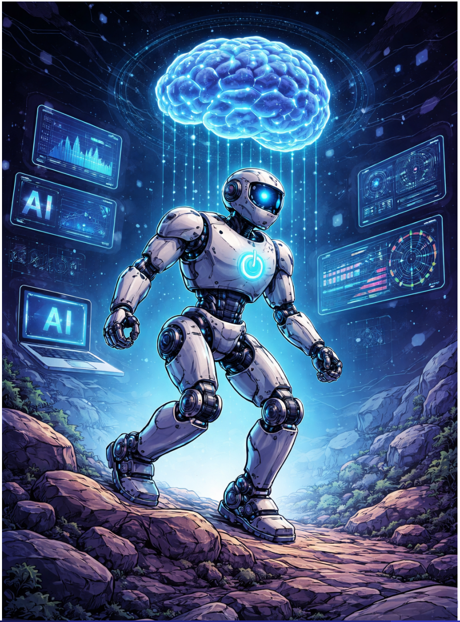
Soft Actor–Critic Based Bipedal Walker Control

Arkaprava Jas Manas Mondal

Department of Computer Science
RKMVERI

December 17, 2025

Picture



Project Overview

- Problem: Learning stable bipedal locomotion
- Environment: *BipedalWalker-v3* (Gymnasium + Box2D)
- Approach: Reinforcement Learning as an MDP
- Algorithm: Soft Actor–Critic (SAC)
- Goal: Stable, efficient forward walking

State Space

- Continuous, fully observable state vector
- Includes:
  - Body orientation and angular velocity
  - Linear velocities
  - Joint angles and angular velocities
  - Ground contact indicators
  - Terrain sensor readings
- State taken directly from environment

Action Space and Constraints

Action Space

- Continuous 4-dimensional action vector

$$a = [a_1, a_2, a_3, a_4]$$

- Joint torques for hips and knees
- Bounded using `tanh`

Constraints

- Torque limits
- Robot must remain upright
- Episode ends if robot falls

Soft Actor–Critic: Core Idea

- Off-policy actor–critic algorithm
- Designed for continuous control
- Maximizes reward and entropy

$$J(\pi) = \mathbb{E}\left[\sum_t r(s_t, a_t) + \alpha \mathcal{H}(\pi(\cdot|s_t))\right]$$

- Entropy encourages exploration
- Improves robustness and stability

SAC Architecture and Training

- Actor network: stochastic policy
- Two critic networks: reduce overestimation
- Replay buffer: sample-efficient learning
- Target networks: stable updates

- Train by interacting with environment
- Store transitions and update networks

Results and Key Takeaways

- Agent learns stable bipedal walking
- Smooth and coordinated leg motion
- Handles continuous, high-dimensional actions
- No explicit dynamics modeling required
- SAC proves effective for locomotion tasks