Waleed Chaudhry – 261014122 COMP 579

# Project Proposal: Investigating the Performance of Policy Gradient Methods in High-Dimensional Continuous Control Tasks

## Problem Statement & Motivation

**Reinforcement Learning (RL) has demonstrated remarkable success in complex decision-making problems, particularly in high-dimensional continuous control environments such as robotics and autonomous systems. Among various RL approaches, policy gradient methods—including REINFORCE, Actor-Critic (A2C, A3C), Deep Deterministic Policy Gradient (DDPG), and Proximal Policy Optimization (PPO)—have become standard techniques due to their ability to optimize policies directly. However, these methods are often sensitive to hyperparameter tuning, function approximation, and exploration-exploitation trade-offs, making their stability and efficiency highly dependent on implementation details. This project aims to systematically compare different policy gradient methods in continuous control tasks, analyzing key factors that influence their performance, stability, and sample efficiency. By doing so, we hope to provide practical insights for improving training robustness and making RL algorithms more viable for real-world applications.**

## Methodology & Experimental Setup

**Our study will focus on high-dimensional continuous control tasks using the MuJoCo physics simulator and OpenAI Gym environments, specifically Hopper-v2, HalfCheetah-v2, and Walker2d-v2. These tasks require fine-grained control over robotic motion, making them ideal benchmarks for testing policy optimization techniques. We will implement and compare REINFORCE, A2C, A3C, DDPG, and PPO, assessing their performance based on three key metrics: total reward over episodes, convergence speed, and variance across multiple runs. To further investigate training dynamics, we will conduct ablation studies on reward shaping, entropy regularization, and neural network architectures to identify their impact on learning stability.**

## Expected Contribution

**This project will provide a detailed comparison of policy gradient methods in continuous control environments, highlighting their strengths and limitations. By analyzing the trade-offs between sample efficiency, stability, and hyperparameter sensitivity, we aim to offer recommendations for improving policy optimization in robotics and real-world applications. Our findings will contribute to the broader understanding of reinforcement learning algorithms and their practical deployment.**