# Task 1 Report

*Fuzail Hamid Banday | 20408698 | Mechatronics Engineering | University of Nottingham*

## The Prototypes

Two distinct prototypes were developed using Q-Learning algorithms. The first prototype was tailored to navigate the CartPole environment, while the second prototype was designed for the MountainCar environment (non-continuous state space).

Both prototypes utilized Q-Learning as the foundational algorithm, albeit applied in different contexts, demonstrating adaptability to diverse environments and reinforcing the fundamental principles of reinforcement learning.

## The Paper

**Patrick Sunden's Research on CartPole:** Sunden's study in "Q-Learning and Deep Q-Learning In OpenAI Gym CartPole Classic Control Environment" serves as a foundation, comparing Q-Learning and DQN in a basic setting. His work establishes benchmarks in simple environments like CartPole but highlights gaps in understanding their performance in more complex scenarios like Acrobot. This gap forms the basis for extending our investigation.

**Insights from 'A Deeper Look at Experience Replay':** Zhang and Sutton's paper offers crucial insights into the role of replay buffers in deep reinforcement learning. Their empirical findings on the impact of replay buffer configurations on learning outcomes underscore the complexities and potential of integrating such techniques in Q-Learning, relevant to our study's focus on replay buffer implementation.

## Gaps (and Justifications)

Sunden's research concluded that for a simple environment like cartpole, Q-Learing is just as good at solving as DQN while being faster. The gap identified lies in the need for testing these algorithms in environments with larger state spaces, such as Acrobot, which presents greater complexity than CartPole. This extension is justified to ascertain if the conclusion drawn for CartPole—Q-Learning being as effective as DQN—holds true in more intricate environments. Examining Acrobot can reveal insights into the algorithms' adaptability and performance when faced with increased complexity beyond what CartPole offers.

Additionally, integrating findings from "A Deeper Look at Experience Replay" into our study adds depth to our exploration, particularly regarding the role of replay buffers in Q-Learning. This integrated approach allows us to assess not just algorithmic performance in varied environments but also how specific enhancements like replay buffers can influence learning dynamics in more complex scenarios.

# Research Questions

1. **Impact of Complexity on Algorithm Performance:** Does the effectiveness of Q-Learning in comparison to DQN, as observed in the CartPole environment, remain consistent when transitioning to more complex environments like Acrobot? Specifically, does an increased state space size directly influence the algorithms' comparative performance?
2. **How does Q-Learning fare against DQN in environments with a higher degree of complexity than CartPole?** This question aims to uncover whether the efficiency and computational advantages of Q-Learning over DQN observed in simpler environments persist or change in more challenging settings.
3. **Investigating the Impact of Replay Buffers in Q-Learning:** This question delves into the effects of implementing replay buffers within the Q-Learning framework. Specifically, it seeks to determine whether the enhancements in learning efficiency and stability observed when replay buffers are used in DQN can be replicated in Q-Learning. Does the integration of replay buffers in Q-Learning confer similar benefits in terms of learning efficiency, particularly in more complex environments like Acrobot, or does it present unique challenges?