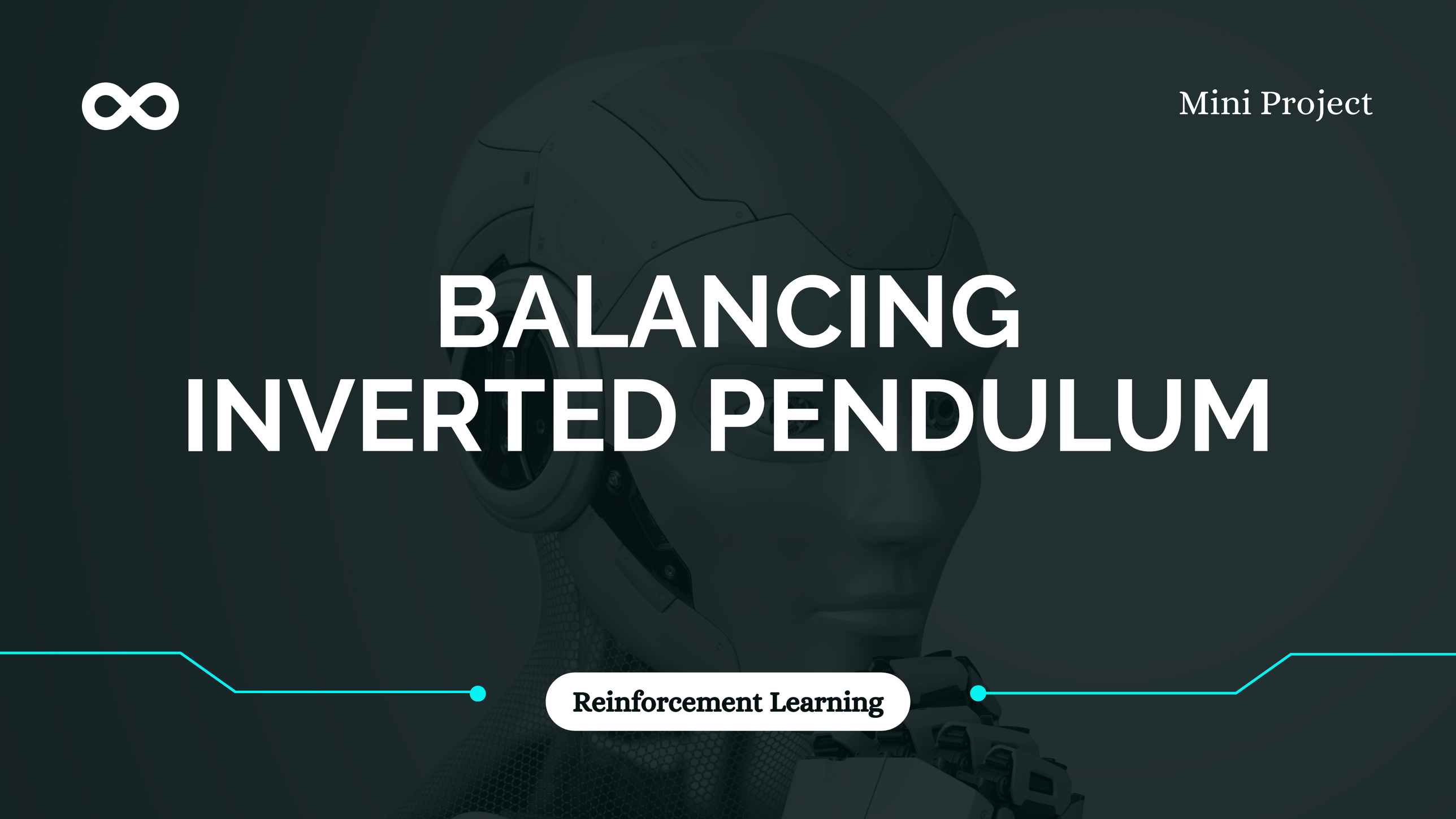
Mini Project

BALANCING INVERTED PENDULUM



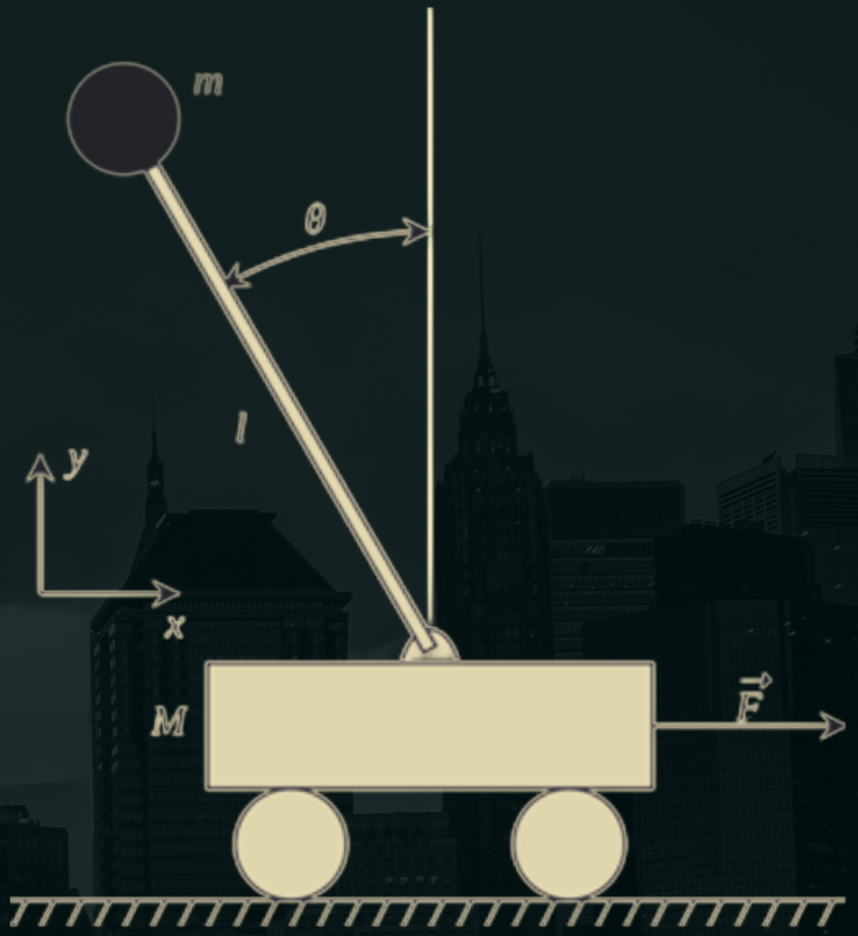
Reinforcement Learning

About the System

Cart Pole, also known as an inverted pendulum, is a game that seeks to keep the pole in balance for as long as possible.

Agent

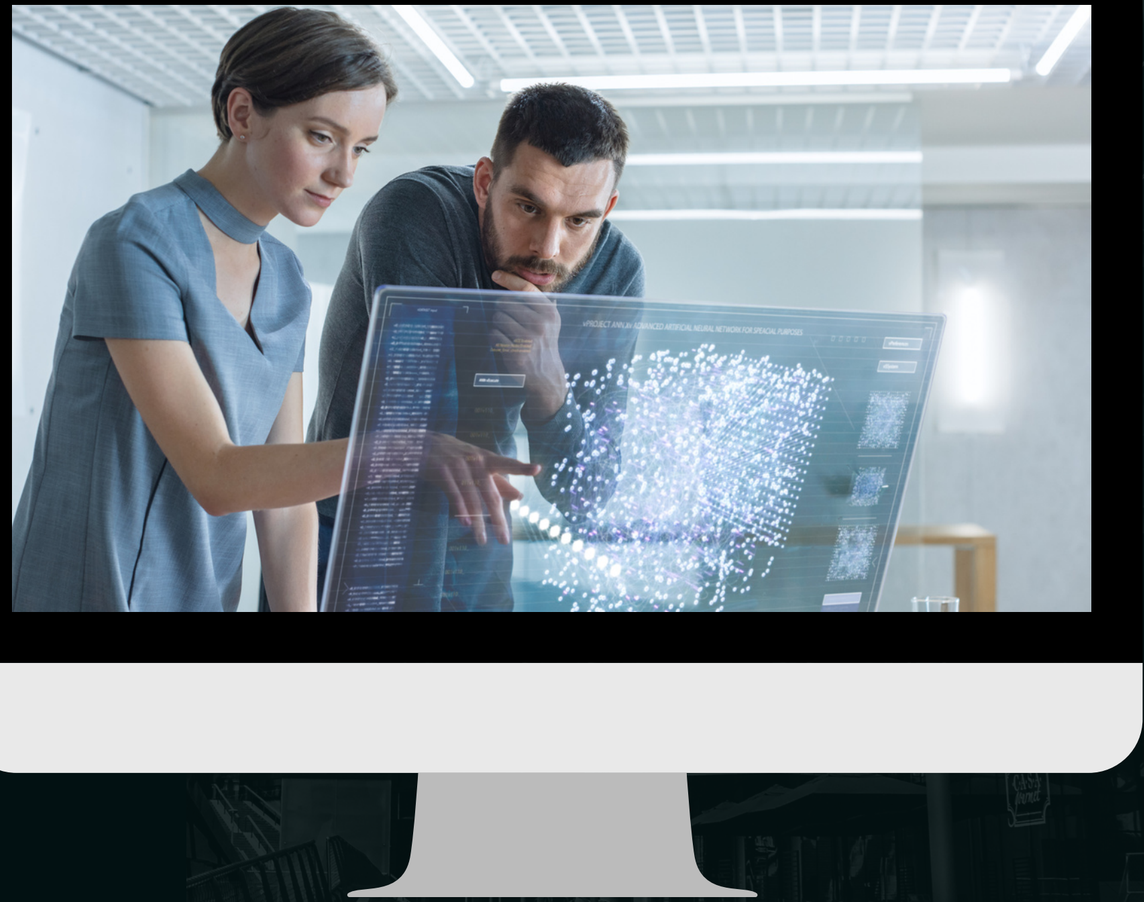
Environment



OBJECTIVE

The Fundamental objective of this project is to keep the Inverted Pendulum balanced using reinforcement learning . AI agent should be able to decide the best action to select, based on its current state at time t.

To analyse various reinforcement learning models such as Q learning, Deep Q learning, Double Deep Q learning using experience replay and target network and choosing the best model to get maximum cumulative reward.

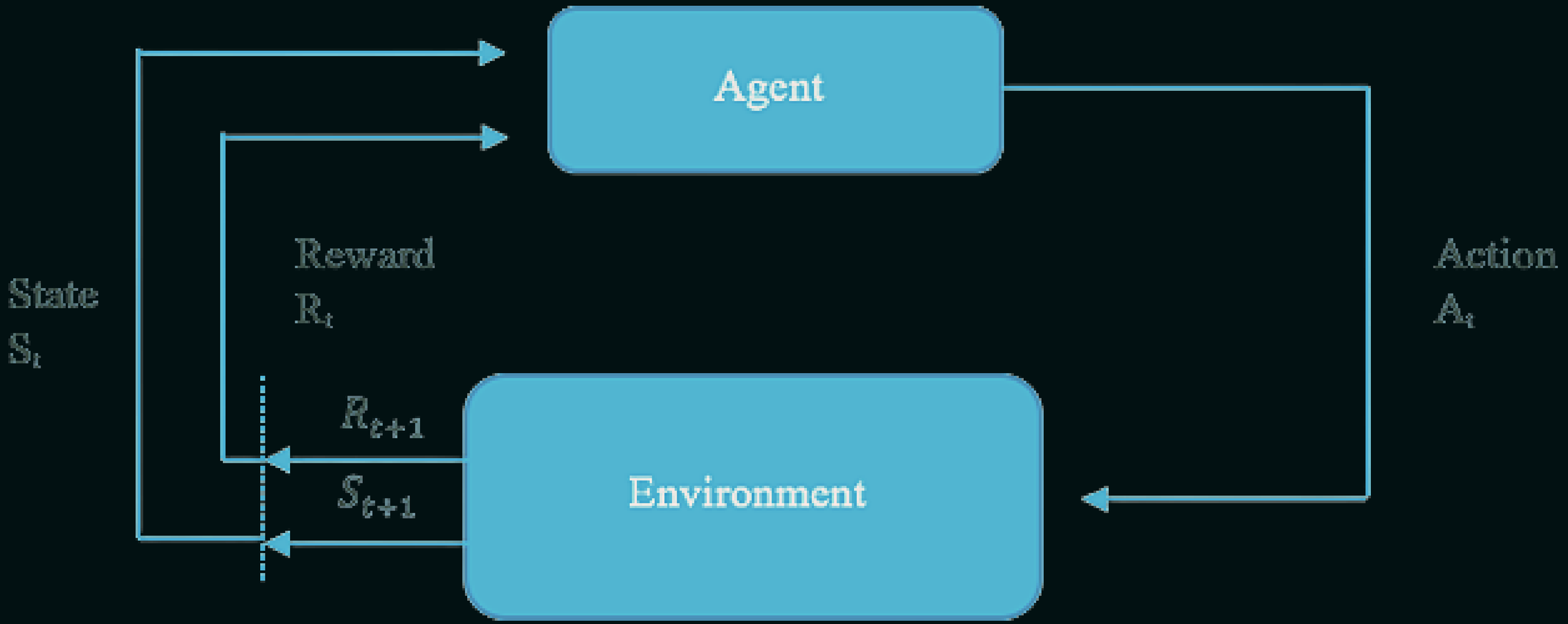


Reinforcement Learning

Reinforcement learning is the training of machine learning models to make a sequence of decisions. The agent learns to achieve a goal in an uncertain, potentially complex environment.

Markov Decision Process (MDP) is a mathematical framework to describe an environment in reinforcement learning.

Markov Decision Process



Q Learning

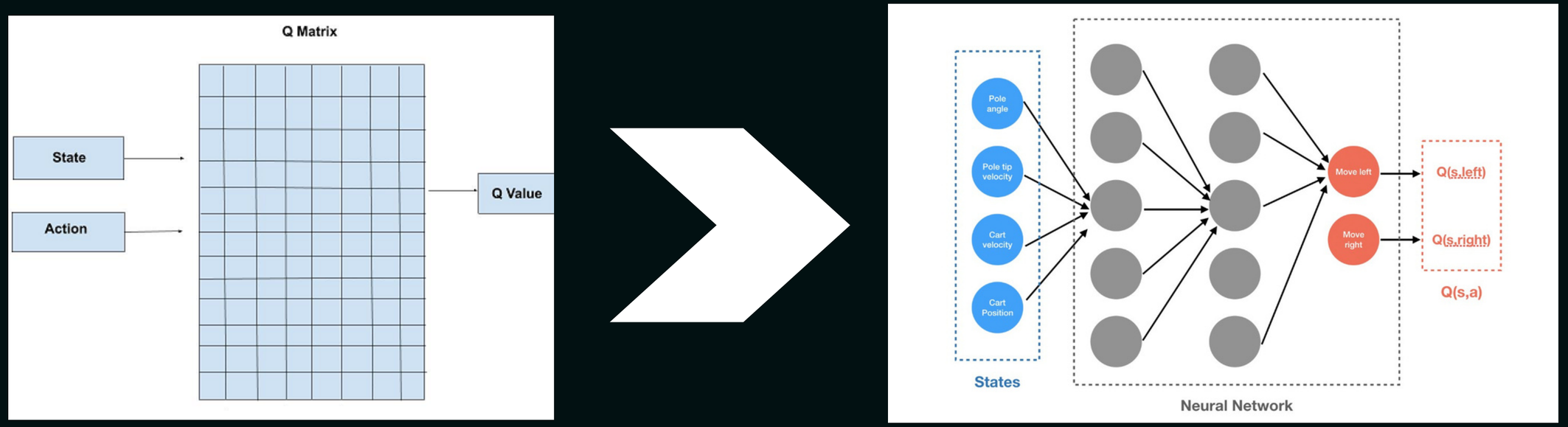
Model-free reinforcement learning algorithm to learn the value of an action in a particular state

Bellman's Equation

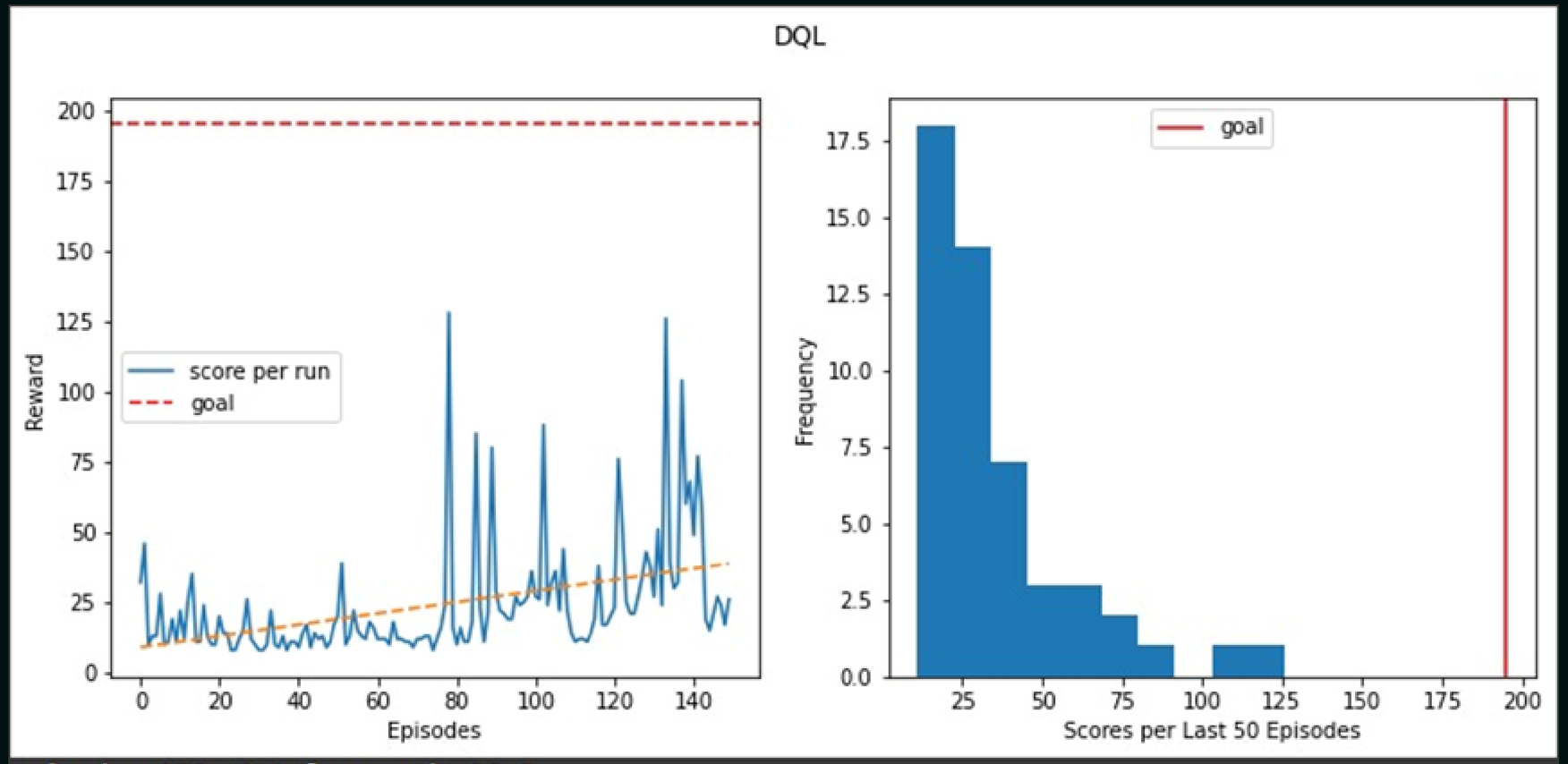
$$q^{new}(s,a) = (1-\alpha)\underbrace{q(s,a)}_{\text{old value}} + \alpha\overbrace{\left(R_{t+1} + \gamma \max_{a'} q(s',a')\right)}^{\text{learned value}}$$

Deep Q Learning

Replaces the regular Q-table with a neural network.

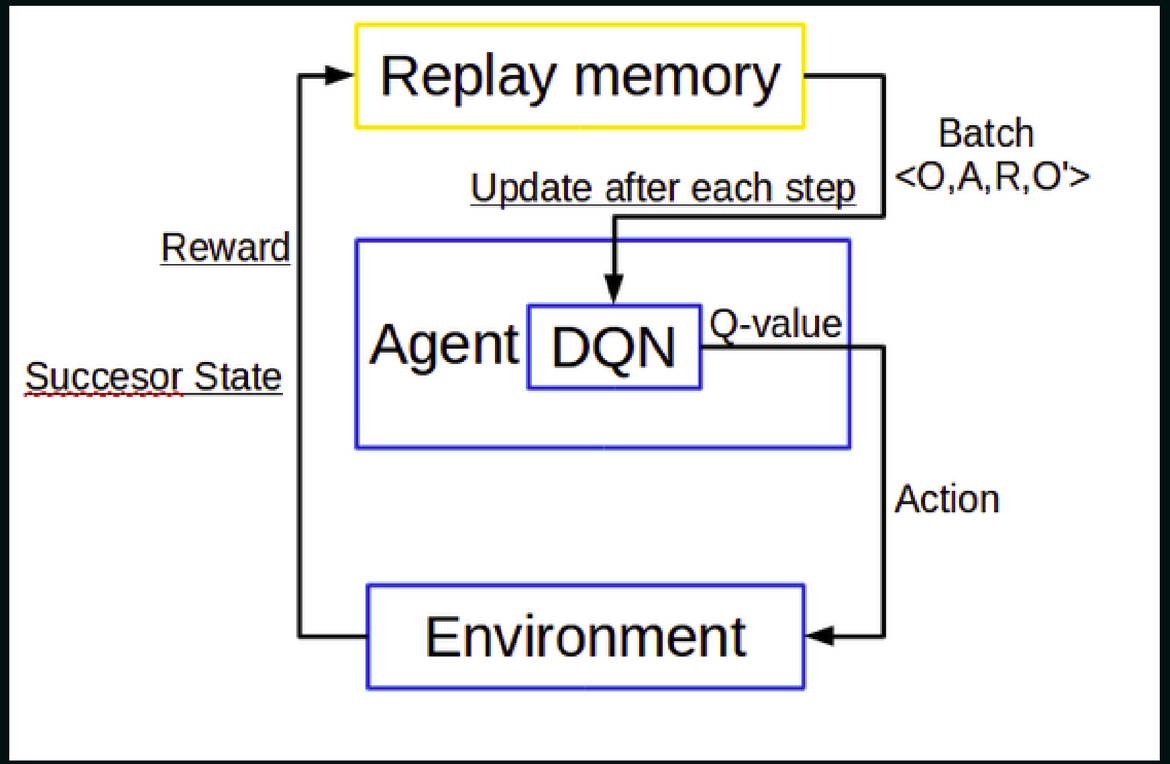


Deep Q Learning Observed Results

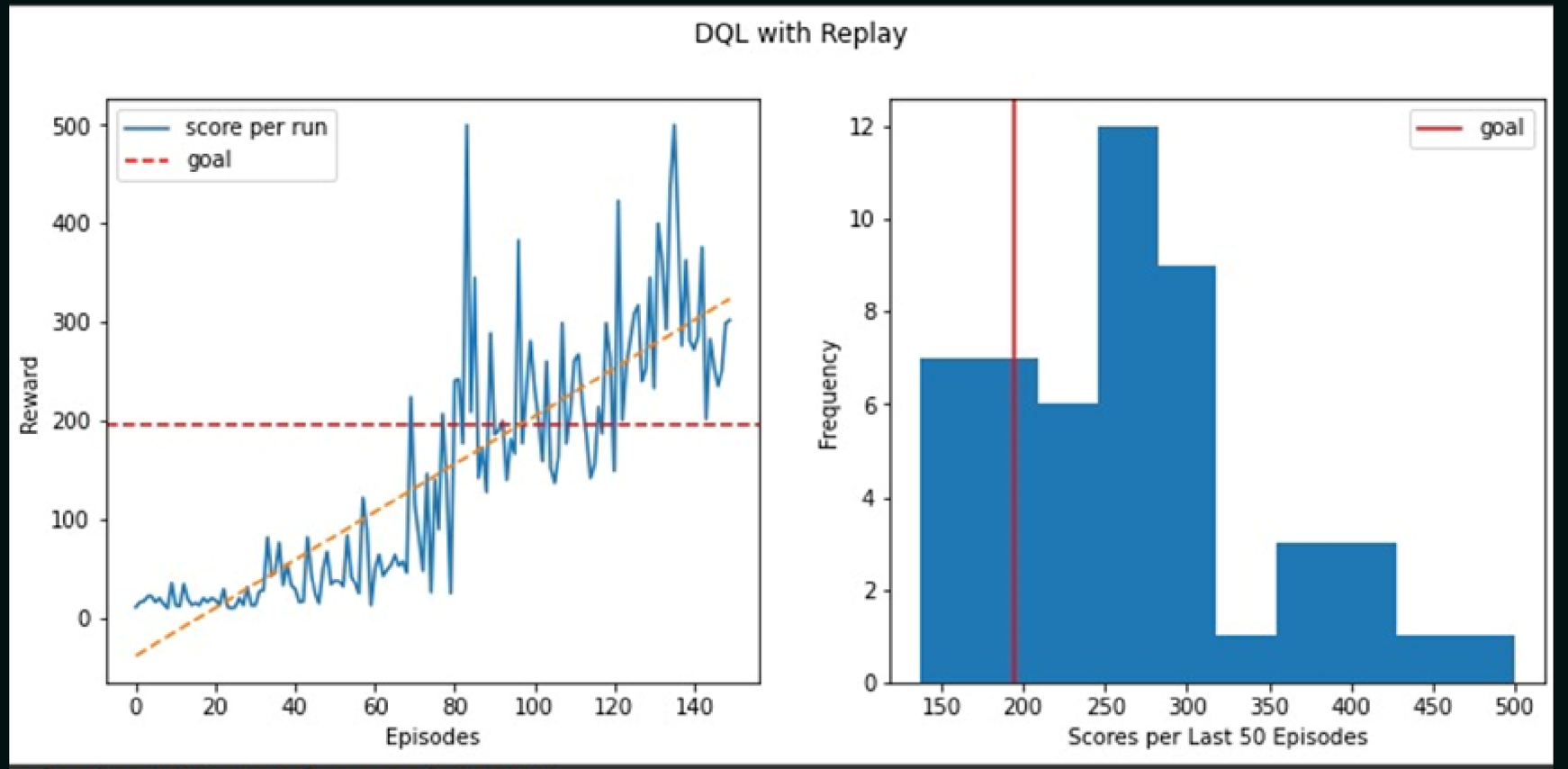


Deep Q Learning with Replay

Stores past experiences and then using a random subset of these experiences to update the Q-network, rather than using just the single most recent experience.

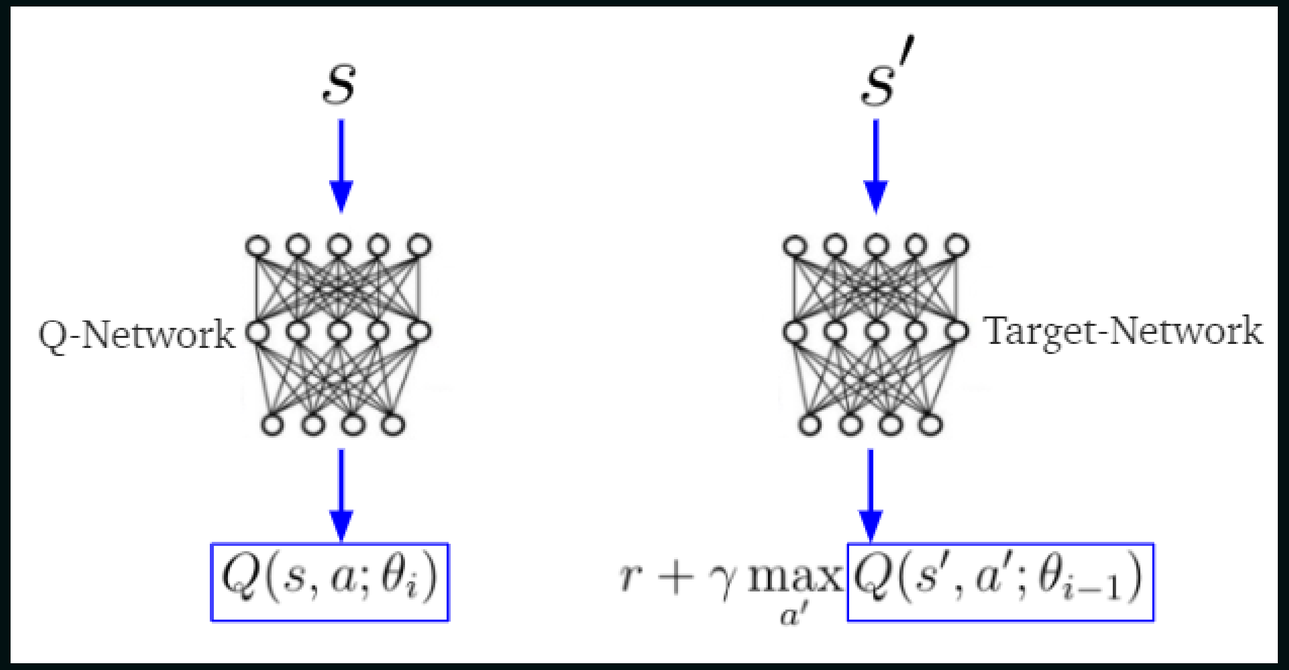


DQL with Replay Observed Results

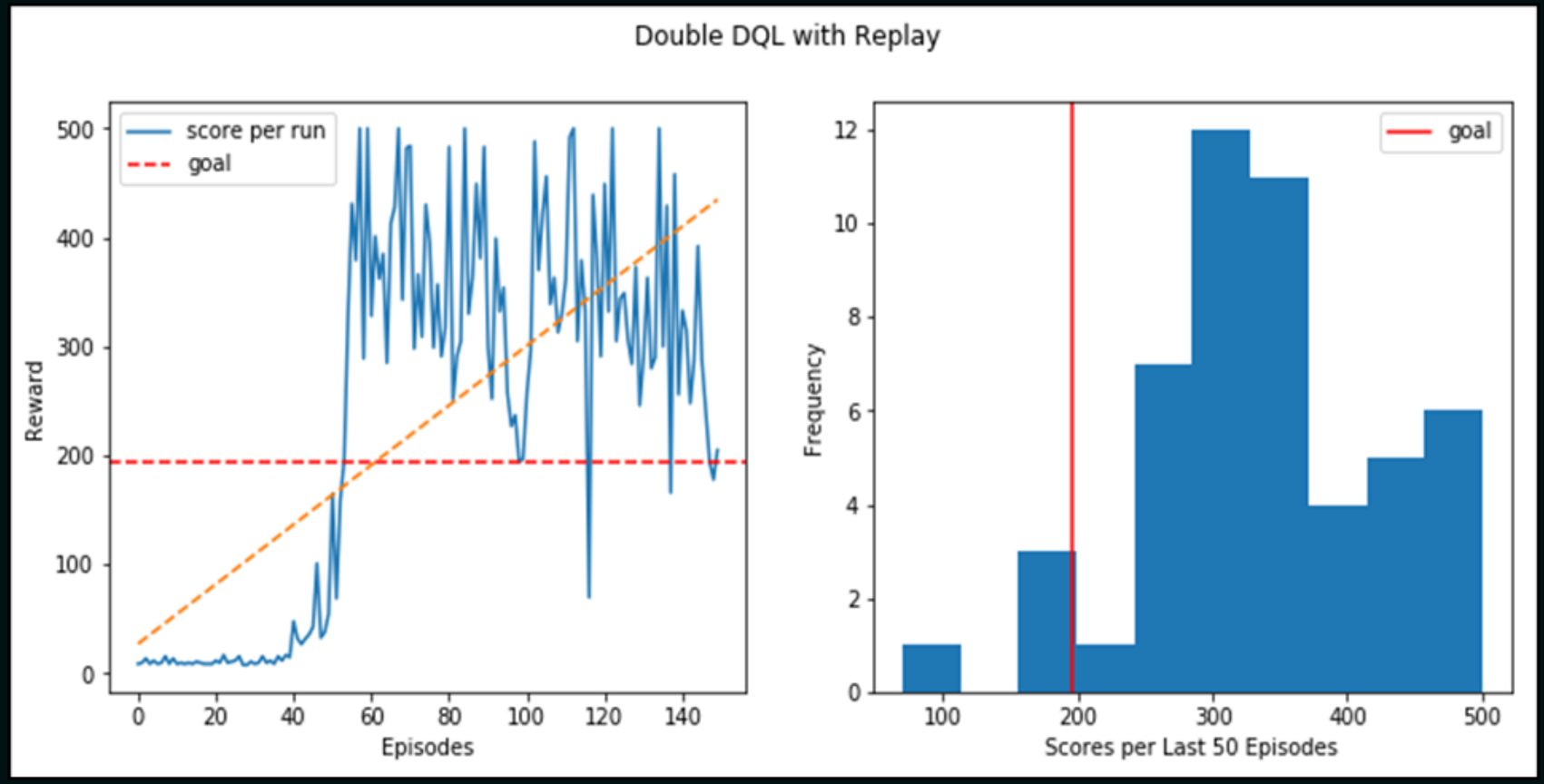


Double Deep Q Learning

Reduces overestimations by decomposing the max operation in the target into action selection and action evaluation.



Double DQL With Replay Observed Results

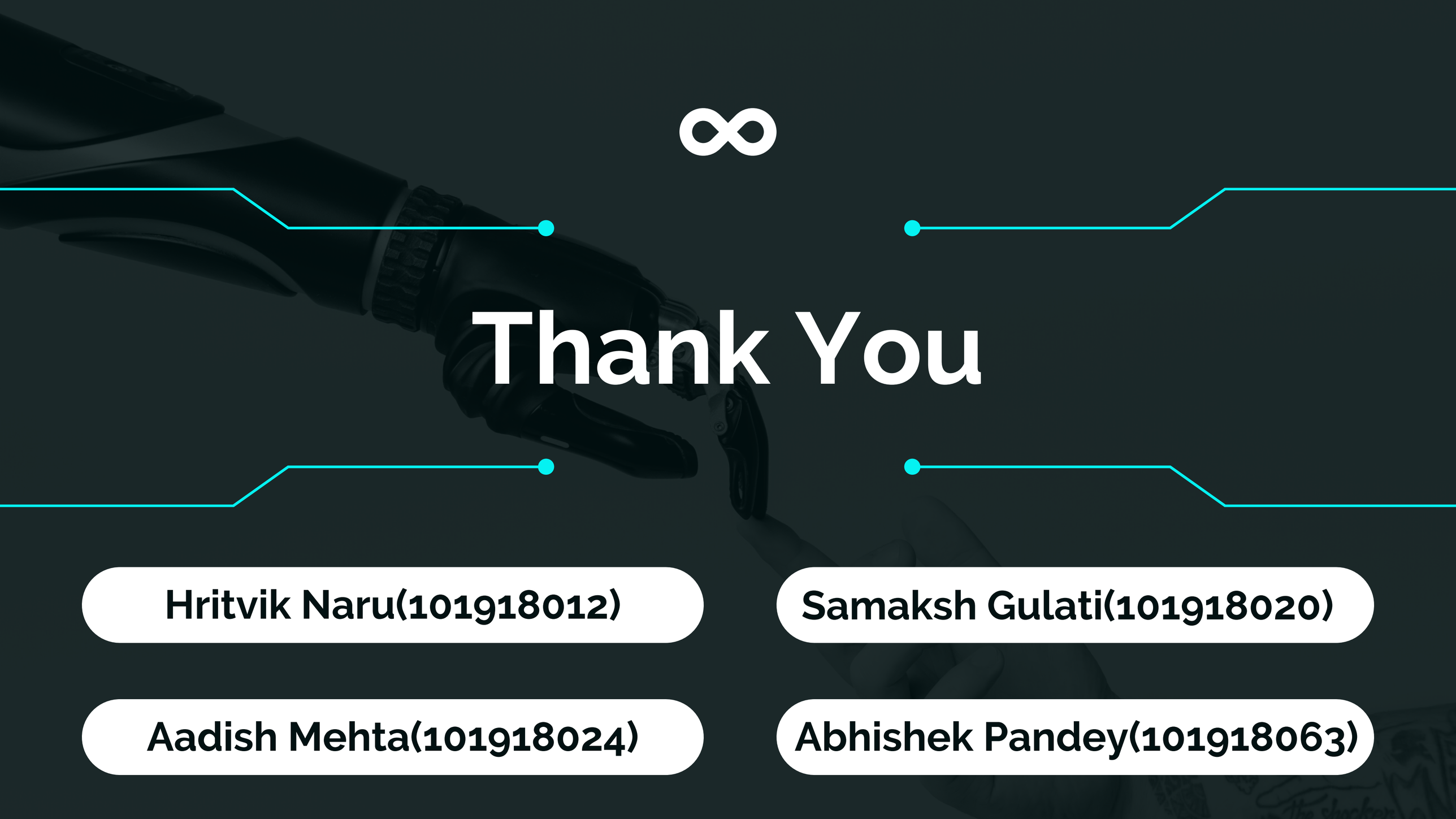


Limitations

- Traditional Q-learning is limited to discrete actions and state space.
- Decreasing these variables leads to ineffective learning due to the curse of dimensionality.
- In DQL, limited action set may not be sufficient to produce the desired behaviour.

Conclusion

We investigated the inverse pendulum problem and used four different RL approaches to solve it. In the Open AI Cart pole scenario, the FCN (Fully Connected Network)-based DQN with experience replay and the target network beats standard RL, which creates a better control strategy in fewer episodes.



Thank You

Hritvik Naru(101918012)

Samaksh Gulati(101918020)

Aadish Mehta(101918024)

Abhishek Pandey(101918063)