FrozenLake-v1 + DQN – Reinforcement Learning Project

- **Environment**: FrozenLake-v1
- **Model**: Deep Q-Network (DQN)
- **Team Members**:
  - Omar Mohamed – ID: 20213661 – Training & Evaluation
  - Abdelraouf Abdelmneam – ID: 20214131– Visualization
  - Abdelrahman Mohamed – ID: 20214037– Documentation
  - Mohamed Hazem –ID: 20213913– Animation

Table of Contents

Environment Details: FrozenLake-v1

Observation Space:

The observation is a single integer from 0 to 15 representing the agent's current position on a 4x4 grid.
Shape: Discrete(16)

Action Space:

There are 4 discrete actions available to the agent:

- 0 = LEFT
- 1 = DOWN
- 2 = RIGHT
- 3 = UP
  Shape: Discrete(4)

State Space:

There are 16 possible states, corresponding to the 4x4 grid positions:

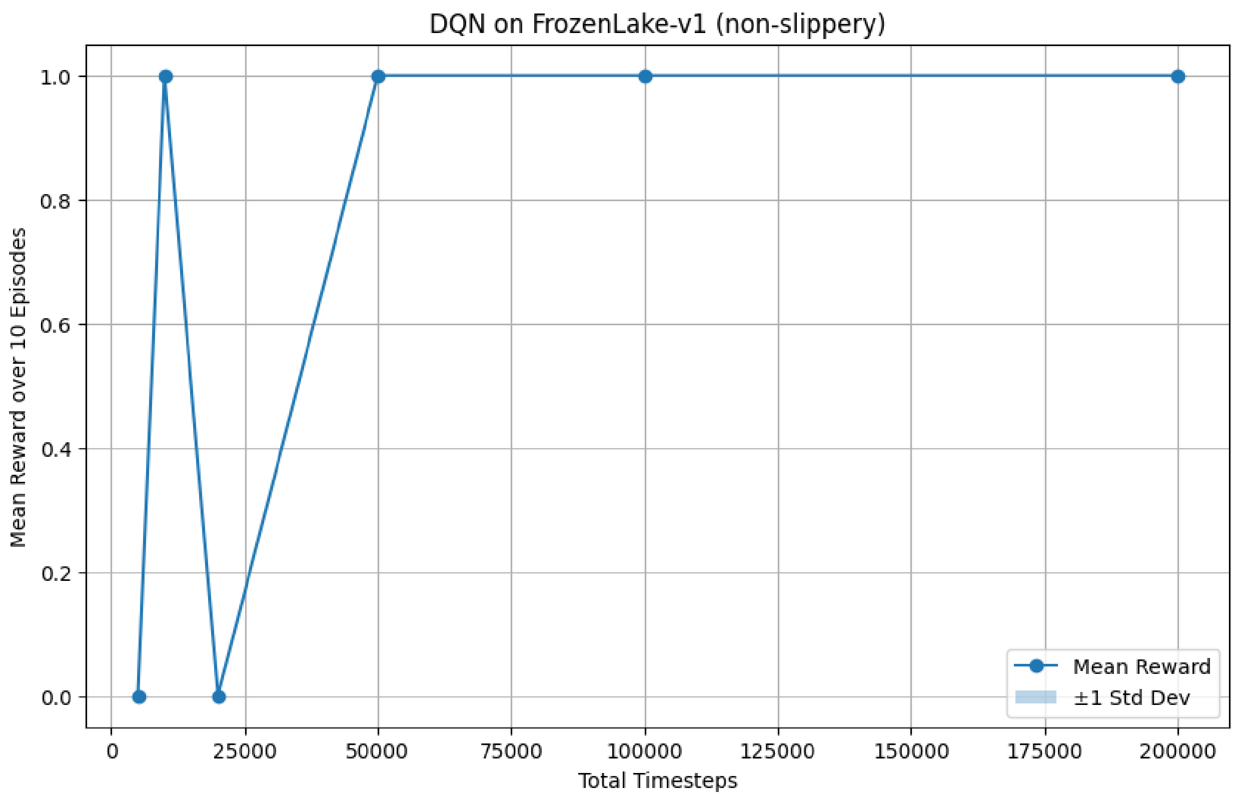
- Start (S)
- Frozen tiles (F)
- Holes (H – falling in ends the episode with zero reward)
- Goal (G – reaching this gives a reward of 1.0)

Model Justification

We chose **Deep Q-Network (DQN)** because it is well-suited for discrete environments like FrozenLake. DQN uses a neural network to estimate Q-values and allows the agent to learn policies in environments with sparse rewards.

Unlike models such as PPO or A2C that are more appropriate for continuous or stochastic problems, DQN performs well in small, discrete, tabular environments.

Performance Graph



✓ Best model at 10000 timesteps with mean reward 1.00

Timesteps vs. Reward Performance

Timesteps	Mean Reward	Std Dev
5,000	0.00	0.00
10,000	1.00	0.00
20,000	0.00	0.00
50,000	1.00	0.00
100,000	1.00	0.00
200,000	1.00	0.00

How Changing Timesteps Affected Rewards

At lower training timesteps (e.g., 5,000), the DQN agent was unable to explore the environment effectively, resulting in zero rewards. However, starting at 10,000 timesteps, the agent consistently learned the optimal policy to reach the goal, achieving a mean reward of 1.00.

Increasing the training time beyond 10,000 timesteps (to 50,000 and above) maintained this perfect performance but did not provide additional improvement. Therefore, 10,000 timesteps offered the best trade-off between training time and agent performance.

Project Summary: FrozenLake-v1 with DQN

In this project, we applied **Deep Q-Networks (DQN)** to solve the **FrozenLake-v1** environment using reinforcement learning. FrozenLake is a grid-based environment where the agent must learn to navigate from a start point to a goal while avoiding holes.

We selected the **non-slippery version** of the environment to make learning more stable. The observation space consists of 16 discrete states, and the agent can take 4 possible actions: left, down, right, and up.

We trained the agent using DQN over varying total timesteps: **5,000 to 200,000**. The model's performance was evaluated using the mean reward over 10 episodes. It achieved a **perfect reward (1.00)** starting from **10,000 timesteps**, showing successful learning. Further increases in training time maintained high performance but did not significantly improve results.

We also created an animation to visualize the agent's learned policy in action, confirming that it successfully navigates the environment to reach the goal.