Reinforcement Learning: Balancing a CartPole using Q-Learning

Nazar Khan
CVML Lab
University of The Punjab

This tutorial helps you get started with OpenAI Gymnasium (the updated version of OpenAI Gym) for reinforcement learning. This tutorial will provide a visual, hands-on experience, where you can see how an agent learns in a simple environment. We'll use **CartPole** as the example environment, which is one of the classic environments in RL.

Getting Started with OpenAI Gymnasium: A Visual Tutorial

In this tutorial, you will learn how to:

1. Install OpenAI Gymnasium and dependencies.
2. Understand the CartPole environment.
3. Create and train a reinforcement learning agent using Q-learning.
4. Visualize how the agent learns over time.

Step 1: Install OpenAI Gymnasium and Dependencies

First, you need to install **OpenAI Gymnasium** (Gym's newer version) and some other dependencies.

Install the necessary libraries:

```
pip install gymnasium[all] numpy matplotlib
```

- `gymnasium[all]` : This installs all the environments (including the classic CartPole environment) and necessary dependencies.
- `numpy` : For array and matrix manipulations.
- `matplotlib` : For visualizing the training process.

Step 2: Import Libraries and Set Up the CartPole Environment

Let's start by importing the necessary libraries and initializing the **CartPole** environment.

In [ ]:
```
import gymnasium as gym
```

```python
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.animation import FuncAnimation
import time

# Create the CartPole environment
env = gym.make("CartPole-v1", render_mode='rgb_array')
print(env)

# Reset the environment to start
observation, info = env.reset()
print("Observation Space:", env.observation_space)
print("Action Space:", env.action_space)
```

```
<TimeLimit<OrderEnforcing<PassiveEnvChecker<CartPoleEnv<CartPole-v1>>>>>
Observation Space: Box([-4.8               -inf -0.41887903        -inf],
[4.8               inf 0.41887903        inf], (4,), float32)
Action Space: Discrete(2)
```

- `gym.make("CartPole-v1")` : This initializes the CartPole environment.
- `render_mode='human'` : This ensures that the environment renders a visual representation for human viewers.
- `env.reset()` : Resets the environment to its initial state.

The output should display information about the observation and action spaces. For CartPole:

- **Observation space** is a continuous space with 4 elements (Cart position, Cart velocity, Pole angle, Pole velocity).
- **Action space** is discrete: 0 (move left) or 1 (move right).

Step 3: Define Q-Learning Algorithm

We'll now define a simple **Q-learning** algorithm for training the agent to balance the pole.

Key elements for Q-learning:

1. **Q-table**: A table that stores Q-values for each state-action pair.
2. **Learning Rate (α)**: Determines how quickly the agent updates its Q-values.
3. **Discount Factor (γ)**: Determines the importance of future rewards.
4. **Exploration-Exploitation (ε)**: Determines the agent's strategy of exploration (random actions) versus exploitation (choosing the best-known action).

In [ ]:
```python
# Parameters for Q-Learning
alpha = 0.1      # Learning rate
gamma = 0.99     # Discount factor
epsilon = 0.1    # Exploration rate
n_episodes = 30000 # Number of episodes for training

# Initialize Q-table (for discrete states)
n_actions = env.action_space.n
```

```
q_table = np.zeros((24, 24, 24, 24, n_actions))  # For CartPole, discreti
print("Shape of Q-table: ", q_table.shape)
```

```
Shape of Q-table:  (24, 24, 24, 24, 2)
```

- **Discretizing the continuous state space**: CartPole's state space is continuous, but we'll discretize it to make Q-learning feasible. Here, the 4 dimensions of the state space are divided into 24 bins each.

Step 4: Discretize the Continuous State Space

To apply Q-learning, we need to convert the continuous state space into discrete states. We'll use `numpy` 's `linspace` to create bins.

In [ ]:
```
# Define state space boundaries and number of bins for each dimension
state_bins = [
    np.linspace(-2.4, 2.4, 24),  # Cart position
    np.linspace(-3.0, 3.0, 24),  # Cart velocity
    np.linspace(-0.5, 0.5, 24),  # Pole angle
    np.linspace(-2.0, 2.0, 24)   # Pole velocity
]

def discretize_state(state):
    """
    Discretize the continuous state to an index in the Q-table.
    """
    state_discretized = []
    for i, (s, bins) in enumerate(zip(state, state_bins)):
        state_discretized.append(np.digitize(s, bins) - 1)
    return tuple(state_discretized)
```

- `np.digitize(s, bins)` maps each continuous state value to a bin index.
- This discretizes the 4-dimensional state space into 4 indices, each ranging from 0 to 23 (as we have 24 bins for each dimension).

Step 5: Train the Agent with Q-learning

Now we will implement the Q-learning training loop. In each episode, the agent will:

1. Choose an action based on an ε-greedy policy.
2. Take the action and observe the new state and reward.
3. Update the Q-table using the Q-learning update rule.

In [ ]:
```
#env = gym.make("CartPole-v1")#, render_mode='human')
# Training loop
rewards = []

for episode in range(n_episodes):
    state, _ = env.reset()  # Reset environment to start a new episode
    total_reward = 0
```

```
    done = False

    while not done:
        # Discretize the state
        state_discretized = discretize_state(state)

        # Exploration vs Exploitation: Choose action
        if np.random.rand() < epsilon:
            action = env.action_space.sample()  # Explore: Random action
        else:
            action = np.argmax(q_table[state_discretized])  # Exploit: Be

        # Step in the environment
        next_state, reward, terminated, truncated, _ = env.step(action)

        # Discretize next state
        next_state_discretized = discretize_state(next_state)

        # Q-learning update rule
        q_table[state_discretized + (action,)] = q_table[state_discretize
            alpha * (reward + gamma * np.max(q_table[next_state_discretiz

        total_reward += reward
        state = next_state

        if terminated or truncated:
            done = True

    rewards.append(total_reward)

    if episode % 50 == 0:
        print(f"Episode {episode}/{n_episodes}, Total Reward: {total_rewa

# Plot the reward curve over episodes
plt.plot(rewards)
plt.xlabel('Episode')
plt.ylabel('Total Reward')
plt.title('Total Rewards Over Training Episodes')
plt.savefig("training_rewards.png")
plt.show()
```

```
Episode 0/30000, Total Reward: 10.0
Episode 50/30000, Total Reward: 13.0
Episode 100/30000, Total Reward: 9.0
Episode 150/30000, Total Reward: 11.0
Episode 200/30000, Total Reward: 11.0
Episode 250/30000, Total Reward: 10.0
Episode 300/30000, Total Reward: 16.0
Episode 350/30000, Total Reward: 14.0
Episode 400/30000, Total Reward: 9.0
Episode 450/30000, Total Reward: 10.0
Episode 500/30000, Total Reward: 12.0
Episode 550/30000, Total Reward: 9.0
Episode 600/30000, Total Reward: 10.0
Episode 650/30000, Total Reward: 15.0
Episode 700/30000, Total Reward: 10.0
Episode 750/30000, Total Reward: 15.0
Episode 800/30000, Total Reward: 13.0
Episode 850/30000, Total Reward: 8.0
Episode 900/30000, Total Reward: 15.0
Episode 950/30000, Total Reward: 15.0
Episode 1000/30000, Total Reward: 15.0
Episode 1050/30000, Total Reward: 11.0
Episode 1100/30000, Total Reward: 16.0
Episode 1150/30000, Total Reward: 14.0
Episode 1200/30000, Total Reward: 18.0
Episode 1250/30000, Total Reward: 15.0
Episode 1300/30000, Total Reward: 19.0
Episode 1350/30000, Total Reward: 20.0
Episode 1400/30000, Total Reward: 11.0
Episode 1450/30000, Total Reward: 18.0
Episode 1500/30000, Total Reward: 37.0
Episode 1550/30000, Total Reward: 17.0
Episode 1600/30000, Total Reward: 14.0
Episode 1650/30000, Total Reward: 11.0
Episode 1700/30000, Total Reward: 16.0
Episode 1750/30000, Total Reward: 12.0
Episode 1800/30000, Total Reward: 16.0
Episode 1850/30000, Total Reward: 21.0
Episode 1900/30000, Total Reward: 13.0
Episode 1950/30000, Total Reward: 29.0
Episode 2000/30000, Total Reward: 14.0
Episode 2050/30000, Total Reward: 19.0
Episode 2100/30000, Total Reward: 27.0
Episode 2150/30000, Total Reward: 57.0
Episode 2200/30000, Total Reward: 74.0
Episode 2250/30000, Total Reward: 32.0
Episode 2300/30000, Total Reward: 57.0
Episode 2350/30000, Total Reward: 88.0
Episode 2400/30000, Total Reward: 59.0
Episode 2450/30000, Total Reward: 42.0
Episode 2500/30000, Total Reward: 54.0
Episode 2550/30000, Total Reward: 36.0
Episode 2600/30000, Total Reward: 44.0
Episode 2650/30000, Total Reward: 76.0
Episode 2700/30000, Total Reward: 35.0
Episode 2750/30000, Total Reward: 55.0
Episode 2800/30000, Total Reward: 34.0
Episode 2850/30000, Total Reward: 60.0
Episode 2900/30000, Total Reward: 76.0
Episode 2950/30000, Total Reward: 35.0
```

```
Episode 3000/30000, Total Reward: 51.0
Episode 3050/30000, Total Reward: 39.0
Episode 3100/30000, Total Reward: 36.0
Episode 3150/30000, Total Reward: 79.0
Episode 3200/30000, Total Reward: 63.0
Episode 3250/30000, Total Reward: 50.0
Episode 3300/30000, Total Reward: 76.0
Episode 3350/30000, Total Reward: 87.0
Episode 3400/30000, Total Reward: 33.0
Episode 3450/30000, Total Reward: 85.0
Episode 3500/30000, Total Reward: 42.0
Episode 3550/30000, Total Reward: 38.0
Episode 3600/30000, Total Reward: 55.0
Episode 3650/30000, Total Reward: 61.0
Episode 3700/30000, Total Reward: 41.0
Episode 3750/30000, Total Reward: 64.0
Episode 3800/30000, Total Reward: 68.0
Episode 3850/30000, Total Reward: 76.0
Episode 3900/30000, Total Reward: 89.0
Episode 3950/30000, Total Reward: 67.0
Episode 4000/30000, Total Reward: 156.0
Episode 4050/30000, Total Reward: 65.0
Episode 4100/30000, Total Reward: 163.0
Episode 4150/30000, Total Reward: 117.0
Episode 4200/30000, Total Reward: 51.0
Episode 4250/30000, Total Reward: 75.0
Episode 4300/30000, Total Reward: 55.0
Episode 4350/30000, Total Reward: 51.0
Episode 4400/30000, Total Reward: 193.0
Episode 4450/30000, Total Reward: 48.0
Episode 4500/30000, Total Reward: 180.0
Episode 4550/30000, Total Reward: 56.0
Episode 4600/30000, Total Reward: 203.0
Episode 4650/30000, Total Reward: 57.0
Episode 4700/30000, Total Reward: 89.0
Episode 4750/30000, Total Reward: 105.0
Episode 4800/30000, Total Reward: 110.0
Episode 4850/30000, Total Reward: 72.0
Episode 4900/30000, Total Reward: 73.0
Episode 4950/30000, Total Reward: 127.0
Episode 5000/30000, Total Reward: 44.0
Episode 5050/30000, Total Reward: 67.0
Episode 5100/30000, Total Reward: 76.0
Episode 5150/30000, Total Reward: 107.0
Episode 5200/30000, Total Reward: 154.0
Episode 5250/30000, Total Reward: 71.0
Episode 5300/30000, Total Reward: 52.0
Episode 5350/30000, Total Reward: 80.0
Episode 5400/30000, Total Reward: 78.0
Episode 5450/30000, Total Reward: 79.0
Episode 5500/30000, Total Reward: 99.0
Episode 5550/30000, Total Reward: 114.0
Episode 5600/30000, Total Reward: 70.0
Episode 5650/30000, Total Reward: 98.0
Episode 5700/30000, Total Reward: 307.0
Episode 5750/30000, Total Reward: 82.0
Episode 5800/30000, Total Reward: 130.0
Episode 5850/30000, Total Reward: 21.0
Episode 5900/30000, Total Reward: 72.0
Episode 5950/30000, Total Reward: 57.0
```

```
Episode 6000/30000, Total Reward: 44.0
Episode 6050/30000, Total Reward: 95.0
Episode 6100/30000, Total Reward: 58.0
Episode 6150/30000, Total Reward: 103.0
Episode 6200/30000, Total Reward: 59.0
Episode 6250/30000, Total Reward: 72.0
Episode 6300/30000, Total Reward: 133.0
Episode 6350/30000, Total Reward: 21.0
Episode 6400/30000, Total Reward: 44.0
Episode 6450/30000, Total Reward: 73.0
Episode 6500/30000, Total Reward: 92.0
Episode 6550/30000, Total Reward: 82.0
Episode 6600/30000, Total Reward: 57.0
Episode 6650/30000, Total Reward: 30.0
Episode 6700/30000, Total Reward: 93.0
Episode 6750/30000, Total Reward: 57.0
Episode 6800/30000, Total Reward: 170.0
Episode 6850/30000, Total Reward: 62.0
Episode 6900/30000, Total Reward: 52.0
Episode 6950/30000, Total Reward: 143.0
Episode 7000/30000, Total Reward: 69.0
Episode 7050/30000, Total Reward: 113.0
Episode 7100/30000, Total Reward: 115.0
Episode 7150/30000, Total Reward: 94.0
Episode 7200/30000, Total Reward: 103.0
Episode 7250/30000, Total Reward: 116.0
Episode 7300/30000, Total Reward: 91.0
Episode 7350/30000, Total Reward: 87.0
Episode 7400/30000, Total Reward: 48.0
Episode 7450/30000, Total Reward: 110.0
Episode 7500/30000, Total Reward: 106.0
Episode 7550/30000, Total Reward: 117.0
Episode 7600/30000, Total Reward: 158.0
Episode 7650/30000, Total Reward: 45.0
Episode 7700/30000, Total Reward: 61.0
Episode 7750/30000, Total Reward: 172.0
Episode 7800/30000, Total Reward: 207.0
Episode 7850/30000, Total Reward: 97.0
Episode 7900/30000, Total Reward: 55.0
Episode 7950/30000, Total Reward: 69.0
Episode 8000/30000, Total Reward: 87.0
Episode 8050/30000, Total Reward: 197.0
Episode 8100/30000, Total Reward: 70.0
Episode 8150/30000, Total Reward: 101.0
Episode 8200/30000, Total Reward: 31.0
Episode 8250/30000, Total Reward: 141.0
Episode 8300/30000, Total Reward: 161.0
Episode 8350/30000, Total Reward: 45.0
Episode 8400/30000, Total Reward: 53.0
Episode 8450/30000, Total Reward: 111.0
Episode 8500/30000, Total Reward: 170.0
Episode 8550/30000, Total Reward: 175.0
Episode 8600/30000, Total Reward: 89.0
Episode 8650/30000, Total Reward: 118.0
Episode 8700/30000, Total Reward: 92.0
Episode 8750/30000, Total Reward: 119.0
Episode 8800/30000, Total Reward: 98.0
Episode 8850/30000, Total Reward: 163.0
Episode 8900/30000, Total Reward: 56.0
Episode 8950/30000, Total Reward: 88.0
```

```
Episode 9000/30000, Total Reward: 96.0
Episode 9050/30000, Total Reward: 83.0
Episode 9100/30000, Total Reward: 73.0
Episode 9150/30000, Total Reward: 151.0
Episode 9200/30000, Total Reward: 155.0
Episode 9250/30000, Total Reward: 98.0
Episode 9300/30000, Total Reward: 157.0
Episode 9350/30000, Total Reward: 113.0
Episode 9400/30000, Total Reward: 118.0
Episode 9450/30000, Total Reward: 93.0
Episode 9500/30000, Total Reward: 113.0
Episode 9550/30000, Total Reward: 45.0
Episode 9600/30000, Total Reward: 106.0
Episode 9650/30000, Total Reward: 110.0
Episode 9700/30000, Total Reward: 116.0
Episode 9750/30000, Total Reward: 168.0
Episode 9800/30000, Total Reward: 139.0
Episode 9850/30000, Total Reward: 115.0
Episode 9900/30000, Total Reward: 59.0
Episode 9950/30000, Total Reward: 67.0
Episode 10000/30000, Total Reward: 197.0
Episode 10050/30000, Total Reward: 102.0
Episode 10100/30000, Total Reward: 151.0
Episode 10150/30000, Total Reward: 115.0
Episode 10200/30000, Total Reward: 135.0
Episode 10250/30000, Total Reward: 136.0
Episode 10300/30000, Total Reward: 156.0
Episode 10350/30000, Total Reward: 204.0
Episode 10400/30000, Total Reward: 92.0
Episode 10450/30000, Total Reward: 119.0
Episode 10500/30000, Total Reward: 101.0
Episode 10550/30000, Total Reward: 117.0
Episode 10600/30000, Total Reward: 107.0
Episode 10650/30000, Total Reward: 94.0
Episode 10700/30000, Total Reward: 88.0
Episode 10750/30000, Total Reward: 112.0
Episode 10800/30000, Total Reward: 114.0
Episode 10850/30000, Total Reward: 159.0
Episode 10900/30000, Total Reward: 102.0
Episode 10950/30000, Total Reward: 111.0
Episode 11000/30000, Total Reward: 53.0
Episode 11050/30000, Total Reward: 111.0
Episode 11100/30000, Total Reward: 26.0
Episode 11150/30000, Total Reward: 58.0
Episode 11200/30000, Total Reward: 22.0
Episode 11250/30000, Total Reward: 81.0
Episode 11300/30000, Total Reward: 49.0
Episode 11350/30000, Total Reward: 19.0
Episode 11400/30000, Total Reward: 52.0
Episode 11450/30000, Total Reward: 57.0
Episode 11500/30000, Total Reward: 66.0
Episode 11550/30000, Total Reward: 85.0
Episode 11600/30000, Total Reward: 76.0
Episode 11650/30000, Total Reward: 92.0
Episode 11700/30000, Total Reward: 36.0
Episode 11750/30000, Total Reward: 107.0
Episode 11800/30000, Total Reward: 43.0
Episode 11850/30000, Total Reward: 108.0
Episode 11900/30000, Total Reward: 106.0
Episode 11950/30000, Total Reward: 78.0
```

```
Episode 12000/30000, Total Reward: 46.0
Episode 12050/30000, Total Reward: 69.0
Episode 12100/30000, Total Reward: 77.0
Episode 12150/30000, Total Reward: 35.0
Episode 12200/30000, Total Reward: 137.0
Episode 12250/30000, Total Reward: 96.0
Episode 12300/30000, Total Reward: 38.0
Episode 12350/30000, Total Reward: 109.0
Episode 12400/30000, Total Reward: 55.0
Episode 12450/30000, Total Reward: 62.0
Episode 12500/30000, Total Reward: 133.0
Episode 12550/30000, Total Reward: 70.0
Episode 12600/30000, Total Reward: 101.0
Episode 12650/30000, Total Reward: 45.0
Episode 12700/30000, Total Reward: 80.0
Episode 12750/30000, Total Reward: 67.0
Episode 12800/30000, Total Reward: 49.0
Episode 12850/30000, Total Reward: 70.0
Episode 12900/30000, Total Reward: 66.0
Episode 12950/30000, Total Reward: 70.0
Episode 13000/30000, Total Reward: 69.0
Episode 13050/30000, Total Reward: 112.0
Episode 13100/30000, Total Reward: 83.0
Episode 13150/30000, Total Reward: 21.0
Episode 13200/30000, Total Reward: 101.0
Episode 13250/30000, Total Reward: 62.0
Episode 13300/30000, Total Reward: 114.0
Episode 13350/30000, Total Reward: 110.0
Episode 13400/30000, Total Reward: 87.0
Episode 13450/30000, Total Reward: 154.0
Episode 13500/30000, Total Reward: 141.0
Episode 13550/30000, Total Reward: 71.0
Episode 13600/30000, Total Reward: 128.0
Episode 13650/30000, Total Reward: 121.0
Episode 13700/30000, Total Reward: 30.0
Episode 13750/30000, Total Reward: 50.0
Episode 13800/30000, Total Reward: 116.0
Episode 13850/30000, Total Reward: 78.0
Episode 13900/30000, Total Reward: 123.0
Episode 13950/30000, Total Reward: 27.0
Episode 14000/30000, Total Reward: 77.0
Episode 14050/30000, Total Reward: 56.0
Episode 14100/30000, Total Reward: 70.0
Episode 14150/30000, Total Reward: 75.0
Episode 14200/30000, Total Reward: 50.0
Episode 14250/30000, Total Reward: 53.0
Episode 14300/30000, Total Reward: 82.0
Episode 14350/30000, Total Reward: 115.0
Episode 14400/30000, Total Reward: 55.0
Episode 14450/30000, Total Reward: 79.0
Episode 14500/30000, Total Reward: 141.0
Episode 14550/30000, Total Reward: 78.0
Episode 14600/30000, Total Reward: 137.0
Episode 14650/30000, Total Reward: 62.0
Episode 14700/30000, Total Reward: 93.0
Episode 14750/30000, Total Reward: 111.0
Episode 14800/30000, Total Reward: 59.0
Episode 14850/30000, Total Reward: 89.0
Episode 14900/30000, Total Reward: 82.0
Episode 14950/30000, Total Reward: 70.0
```

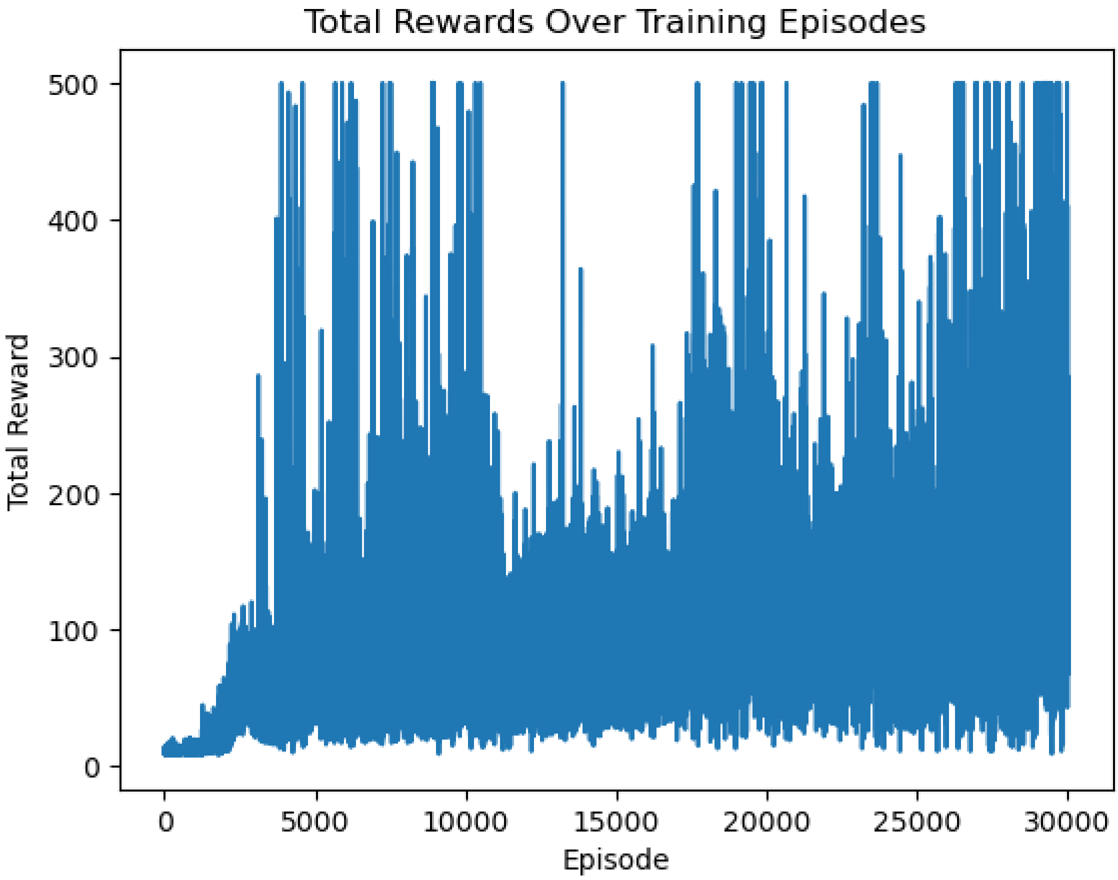
```
Episode 15000/30000, Total Reward: 74.0
Episode 15050/30000, Total Reward: 81.0
Episode 15100/30000, Total Reward: 59.0
Episode 15150/30000, Total Reward: 59.0
Episode 15200/30000, Total Reward: 67.0
Episode 15250/30000, Total Reward: 87.0
Episode 15300/30000, Total Reward: 54.0
Episode 15350/30000, Total Reward: 108.0
Episode 15400/30000, Total Reward: 82.0
Episode 15450/30000, Total Reward: 76.0
Episode 15500/30000, Total Reward: 79.0
Episode 15550/30000, Total Reward: 95.0
Episode 15600/30000, Total Reward: 90.0
Episode 15650/30000, Total Reward: 101.0
Episode 15700/30000, Total Reward: 90.0
Episode 15750/30000, Total Reward: 128.0
Episode 15800/30000, Total Reward: 134.0
Episode 15850/30000, Total Reward: 97.0
Episode 15900/30000, Total Reward: 93.0
Episode 15950/30000, Total Reward: 117.0
Episode 16000/30000, Total Reward: 87.0
Episode 16050/30000, Total Reward: 129.0
Episode 16100/30000, Total Reward: 113.0
Episode 16150/30000, Total Reward: 86.0
Episode 16200/30000, Total Reward: 184.0
Episode 16250/30000, Total Reward: 96.0
Episode 16300/30000, Total Reward: 90.0
Episode 16350/30000, Total Reward: 103.0
Episode 16400/30000, Total Reward: 144.0
Episode 16450/30000, Total Reward: 87.0
Episode 16500/30000, Total Reward: 150.0
Episode 16550/30000, Total Reward: 162.0
Episode 16600/30000, Total Reward: 125.0
Episode 16650/30000, Total Reward: 94.0
Episode 16700/30000, Total Reward: 63.0
Episode 16750/30000, Total Reward: 110.0
Episode 16800/30000, Total Reward: 53.0
Episode 16850/30000, Total Reward: 80.0
Episode 16900/30000, Total Reward: 77.0
Episode 16950/30000, Total Reward: 128.0
Episode 17000/30000, Total Reward: 87.0
Episode 17050/30000, Total Reward: 126.0
Episode 17100/30000, Total Reward: 105.0
Episode 17150/30000, Total Reward: 255.0
Episode 17200/30000, Total Reward: 119.0
Episode 17250/30000, Total Reward: 38.0
Episode 17300/30000, Total Reward: 130.0
Episode 17350/30000, Total Reward: 125.0
Episode 17400/30000, Total Reward: 126.0
Episode 17450/30000, Total Reward: 131.0
Episode 17500/30000, Total Reward: 87.0
Episode 17550/30000, Total Reward: 187.0
Episode 17600/30000, Total Reward: 206.0
Episode 17650/30000, Total Reward: 176.0
Episode 17700/30000, Total Reward: 113.0
Episode 17750/30000, Total Reward: 150.0
Episode 17800/30000, Total Reward: 295.0
Episode 17850/30000, Total Reward: 120.0
Episode 17900/30000, Total Reward: 185.0
Episode 17950/30000, Total Reward: 163.0
```

```
Episode 18000/30000, Total Reward: 256.0
Episode 18050/30000, Total Reward: 148.0
Episode 18100/30000, Total Reward: 134.0
Episode 18150/30000, Total Reward: 190.0
Episode 18200/30000, Total Reward: 135.0
Episode 18250/30000, Total Reward: 141.0
Episode 18300/30000, Total Reward: 144.0
Episode 18350/30000, Total Reward: 117.0
Episode 18400/30000, Total Reward: 199.0
Episode 18450/30000, Total Reward: 133.0
Episode 18500/30000, Total Reward: 114.0
Episode 18550/30000, Total Reward: 178.0
Episode 18600/30000, Total Reward: 225.0
Episode 18650/30000, Total Reward: 213.0
Episode 18700/30000, Total Reward: 172.0
Episode 18750/30000, Total Reward: 142.0
Episode 18800/30000, Total Reward: 102.0
Episode 18850/30000, Total Reward: 113.0
Episode 18900/30000, Total Reward: 118.0
Episode 18950/30000, Total Reward: 147.0
Episode 19000/30000, Total Reward: 129.0
Episode 19050/30000, Total Reward: 180.0
Episode 19100/30000, Total Reward: 97.0
Episode 19150/30000, Total Reward: 135.0
Episode 19200/30000, Total Reward: 169.0
Episode 19250/30000, Total Reward: 124.0
Episode 19300/30000, Total Reward: 134.0
Episode 19350/30000, Total Reward: 102.0
Episode 19400/30000, Total Reward: 127.0
Episode 19450/30000, Total Reward: 250.0
Episode 19500/30000, Total Reward: 105.0
Episode 19550/30000, Total Reward: 130.0
Episode 19600/30000, Total Reward: 166.0
Episode 19650/30000, Total Reward: 86.0
Episode 19700/30000, Total Reward: 202.0
Episode 19750/30000, Total Reward: 171.0
Episode 19800/30000, Total Reward: 159.0
Episode 19850/30000, Total Reward: 247.0
Episode 19900/30000, Total Reward: 135.0
Episode 19950/30000, Total Reward: 108.0
Episode 20000/30000, Total Reward: 96.0
Episode 20050/30000, Total Reward: 24.0
Episode 20100/30000, Total Reward: 266.0
Episode 20150/30000, Total Reward: 138.0
Episode 20200/30000, Total Reward: 141.0
Episode 20250/30000, Total Reward: 119.0
Episode 20300/30000, Total Reward: 29.0
Episode 20350/30000, Total Reward: 189.0
Episode 20400/30000, Total Reward: 127.0
Episode 20450/30000, Total Reward: 140.0
Episode 20500/30000, Total Reward: 191.0
Episode 20550/30000, Total Reward: 135.0
Episode 20600/30000, Total Reward: 189.0
Episode 20650/30000, Total Reward: 129.0
Episode 20700/30000, Total Reward: 66.0
Episode 20750/30000, Total Reward: 218.0
Episode 20800/30000, Total Reward: 112.0
Episode 20850/30000, Total Reward: 142.0
Episode 20900/30000, Total Reward: 104.0
Episode 20950/30000, Total Reward: 134.0
```

```
Episode 21000/30000, Total Reward: 76.0
Episode 21050/30000, Total Reward: 130.0
Episode 21100/30000, Total Reward: 91.0
Episode 21150/30000, Total Reward: 134.0
Episode 21200/30000, Total Reward: 149.0
Episode 21250/30000, Total Reward: 24.0
Episode 21300/30000, Total Reward: 101.0
Episode 21350/30000, Total Reward: 198.0
Episode 21400/30000, Total Reward: 68.0
Episode 21450/30000, Total Reward: 86.0
Episode 21500/30000, Total Reward: 147.0
Episode 21550/30000, Total Reward: 83.0
Episode 21600/30000, Total Reward: 68.0
Episode 21650/30000, Total Reward: 84.0
Episode 21700/30000, Total Reward: 29.0
Episode 21750/30000, Total Reward: 96.0
Episode 21800/30000, Total Reward: 58.0
Episode 21850/30000, Total Reward: 133.0
Episode 21900/30000, Total Reward: 110.0
Episode 21950/30000, Total Reward: 117.0
Episode 22000/30000, Total Reward: 84.0
Episode 22050/30000, Total Reward: 135.0
Episode 22100/30000, Total Reward: 78.0
Episode 22150/30000, Total Reward: 133.0
Episode 22200/30000, Total Reward: 39.0
Episode 22250/30000, Total Reward: 88.0
Episode 22300/30000, Total Reward: 92.0
Episode 22350/30000, Total Reward: 94.0
Episode 22400/30000, Total Reward: 99.0
Episode 22450/30000, Total Reward: 89.0
Episode 22500/30000, Total Reward: 99.0
Episode 22550/30000, Total Reward: 132.0
Episode 22600/30000, Total Reward: 69.0
Episode 22650/30000, Total Reward: 96.0
Episode 22700/30000, Total Reward: 155.0
Episode 22750/30000, Total Reward: 133.0
Episode 22800/30000, Total Reward: 113.0
Episode 22850/30000, Total Reward: 124.0
Episode 22900/30000, Total Reward: 145.0
Episode 22950/30000, Total Reward: 103.0
Episode 23000/30000, Total Reward: 95.0
Episode 23050/30000, Total Reward: 84.0
Episode 23100/30000, Total Reward: 107.0
Episode 23150/30000, Total Reward: 170.0
Episode 23200/30000, Total Reward: 218.0
Episode 23250/30000, Total Reward: 162.0
Episode 23300/30000, Total Reward: 135.0
Episode 23350/30000, Total Reward: 52.0
Episode 23400/30000, Total Reward: 155.0
Episode 23450/30000, Total Reward: 188.0
Episode 23500/30000, Total Reward: 146.0
Episode 23550/30000, Total Reward: 93.0
Episode 23600/30000, Total Reward: 189.0
Episode 23650/30000, Total Reward: 97.0
Episode 23700/30000, Total Reward: 270.0
Episode 23750/30000, Total Reward: 162.0
Episode 23800/30000, Total Reward: 84.0
Episode 23850/30000, Total Reward: 51.0
Episode 23900/30000, Total Reward: 175.0
Episode 23950/30000, Total Reward: 133.0
```

```
Episode 24000/30000, Total Reward: 110.0
Episode 24050/30000, Total Reward: 129.0
Episode 24100/30000, Total Reward: 66.0
Episode 24150/30000, Total Reward: 104.0
Episode 24200/30000, Total Reward: 111.0
Episode 24250/30000, Total Reward: 72.0
Episode 24300/30000, Total Reward: 92.0
Episode 24350/30000, Total Reward: 112.0
Episode 24400/30000, Total Reward: 104.0
Episode 24450/30000, Total Reward: 130.0
Episode 24500/30000, Total Reward: 36.0
Episode 24550/30000, Total Reward: 164.0
Episode 24600/30000, Total Reward: 180.0
Episode 24650/30000, Total Reward: 137.0
Episode 24700/30000, Total Reward: 180.0
Episode 24750/30000, Total Reward: 108.0
Episode 24800/30000, Total Reward: 52.0
Episode 24850/30000, Total Reward: 86.0
Episode 24900/30000, Total Reward: 158.0
Episode 24950/30000, Total Reward: 114.0
Episode 25000/30000, Total Reward: 110.0
Episode 25050/30000, Total Reward: 87.0
Episode 25100/30000, Total Reward: 99.0
Episode 25150/30000, Total Reward: 116.0
Episode 25200/30000, Total Reward: 48.0
Episode 25250/30000, Total Reward: 96.0
Episode 25300/30000, Total Reward: 137.0
Episode 25350/30000, Total Reward: 133.0
Episode 25400/30000, Total Reward: 106.0
Episode 25450/30000, Total Reward: 87.0
Episode 25500/30000, Total Reward: 134.0
Episode 25550/30000, Total Reward: 107.0
Episode 25600/30000, Total Reward: 126.0
Episode 25650/30000, Total Reward: 102.0
Episode 25700/30000, Total Reward: 203.0
Episode 25750/30000, Total Reward: 250.0
Episode 25800/30000, Total Reward: 224.0
Episode 25850/30000, Total Reward: 107.0
Episode 25900/30000, Total Reward: 84.0
Episode 25950/30000, Total Reward: 124.0
Episode 26000/30000, Total Reward: 138.0
Episode 26050/30000, Total Reward: 188.0
Episode 26100/30000, Total Reward: 105.0
Episode 26150/30000, Total Reward: 145.0
Episode 26200/30000, Total Reward: 128.0
Episode 26250/30000, Total Reward: 135.0
Episode 26300/30000, Total Reward: 68.0
Episode 26350/30000, Total Reward: 177.0
Episode 26400/30000, Total Reward: 226.0
Episode 26450/30000, Total Reward: 126.0
Episode 26500/30000, Total Reward: 117.0
Episode 26550/30000, Total Reward: 107.0
Episode 26600/30000, Total Reward: 114.0
Episode 26650/30000, Total Reward: 98.0
Episode 26700/30000, Total Reward: 103.0
Episode 26750/30000, Total Reward: 143.0
Episode 26800/30000, Total Reward: 174.0
Episode 26850/30000, Total Reward: 149.0
Episode 26900/30000, Total Reward: 198.0
Episode 26950/30000, Total Reward: 155.0
```

```
Episode 27000/30000, Total Reward: 105.0
Episode 27050/30000, Total Reward: 157.0
Episode 27100/30000, Total Reward: 177.0
Episode 27150/30000, Total Reward: 189.0
Episode 27200/30000, Total Reward: 33.0
Episode 27250/30000, Total Reward: 127.0
Episode 27300/30000, Total Reward: 173.0
Episode 27350/30000, Total Reward: 251.0
Episode 27400/30000, Total Reward: 238.0
Episode 27450/30000, Total Reward: 250.0
Episode 27500/30000, Total Reward: 166.0
Episode 27550/30000, Total Reward: 221.0
Episode 27600/30000, Total Reward: 218.0
Episode 27650/30000, Total Reward: 133.0
Episode 27700/30000, Total Reward: 242.0
Episode 27750/30000, Total Reward: 166.0
Episode 27800/30000, Total Reward: 181.0
Episode 27850/30000, Total Reward: 160.0
Episode 27900/30000, Total Reward: 145.0
Episode 27950/30000, Total Reward: 171.0
Episode 28000/30000, Total Reward: 181.0
Episode 28050/30000, Total Reward: 34.0
Episode 28100/30000, Total Reward: 221.0
Episode 28150/30000, Total Reward: 175.0
Episode 28200/30000, Total Reward: 86.0
Episode 28250/30000, Total Reward: 97.0
Episode 28300/30000, Total Reward: 46.0
Episode 28350/30000, Total Reward: 109.0
Episode 28400/30000, Total Reward: 144.0
Episode 28450/30000, Total Reward: 182.0
Episode 28500/30000, Total Reward: 107.0
Episode 28550/30000, Total Reward: 158.0
Episode 28600/30000, Total Reward: 177.0
Episode 28650/30000, Total Reward: 165.0
Episode 28700/30000, Total Reward: 131.0
Episode 28750/30000, Total Reward: 245.0
Episode 28800/30000, Total Reward: 275.0
Episode 28850/30000, Total Reward: 138.0
Episode 28900/30000, Total Reward: 146.0
Episode 28950/30000, Total Reward: 170.0
Episode 29000/30000, Total Reward: 235.0
Episode 29050/30000, Total Reward: 500.0
Episode 29100/30000, Total Reward: 356.0
Episode 29150/30000, Total Reward: 179.0
Episode 29200/30000, Total Reward: 238.0
Episode 29250/30000, Total Reward: 133.0
Episode 29300/30000, Total Reward: 113.0
Episode 29350/30000, Total Reward: 163.0
Episode 29400/30000, Total Reward: 169.0
Episode 29450/30000, Total Reward: 355.0
Episode 29500/30000, Total Reward: 233.0
Episode 29550/30000, Total Reward: 132.0
Episode 29600/30000, Total Reward: 172.0
Episode 29650/30000, Total Reward: 154.0
Episode 29700/30000, Total Reward: 269.0
Episode 29750/30000, Total Reward: 211.0
Episode 29800/30000, Total Reward: 179.0
Episode 29850/30000, Total Reward: 121.0
Episode 29900/30000, Total Reward: 121.0
Episode 29950/30000, Total Reward: 157.0
```




```
<Figure size 640x480 with 0 Axes>
```

- **Q-learning update**: After each action, we update the Q-value for the state-action pair using the Q-learning rule.
- **ε-greedy policy**: The agent explores the environment by taking random actions with probability ε and exploits the best-known action with probability 1-ε.
- **Reward visualization**: After training, we plot the total reward achieved by the agent in each episode.

When does an episode end?

In our "CartPole-v1" environment, an **episode** ends when one of the following conditions is met:

1. **Pole Angle Exceeds Limit**:
   The pole's angle exceeds a threshold (±12 degrees from vertical). This threshold is measured in radians internally.

2. **Cart Position Exceeds Bounds**:
   The cart moves too far to the left or right from the center. Specifically, the cart's position exceeds ±2.4 units from the center of the track.

3. **Maximum Episode Steps**:
   The environment has a maximum step limit (500 steps for "CartPole-v1"). If this limit is reached without any of the above failures, the episode ends successfully.

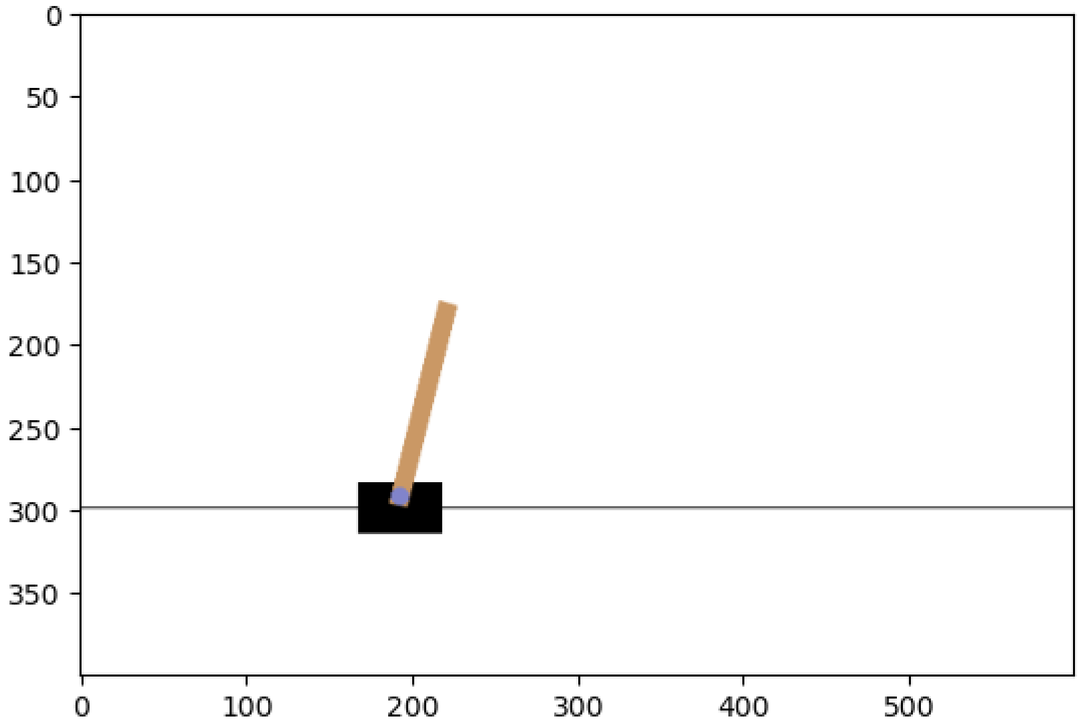
Rewards:
For every step the pole remains upright, the agent receives a reward of +1. Therefore, the total reward in an episode reflects how long the pole was balanced.

Done Flag:
When the episode ends, the environment returns `done=True`. This indicates that the episode has concluded, and the agent should reset the environment before continuing.

Step 6: Visualize the Trained Agent and Compare with an Untrained Agent

To see how the agent performs, you can run the trained agent in the environment for a few episodes and visualize its behavior.

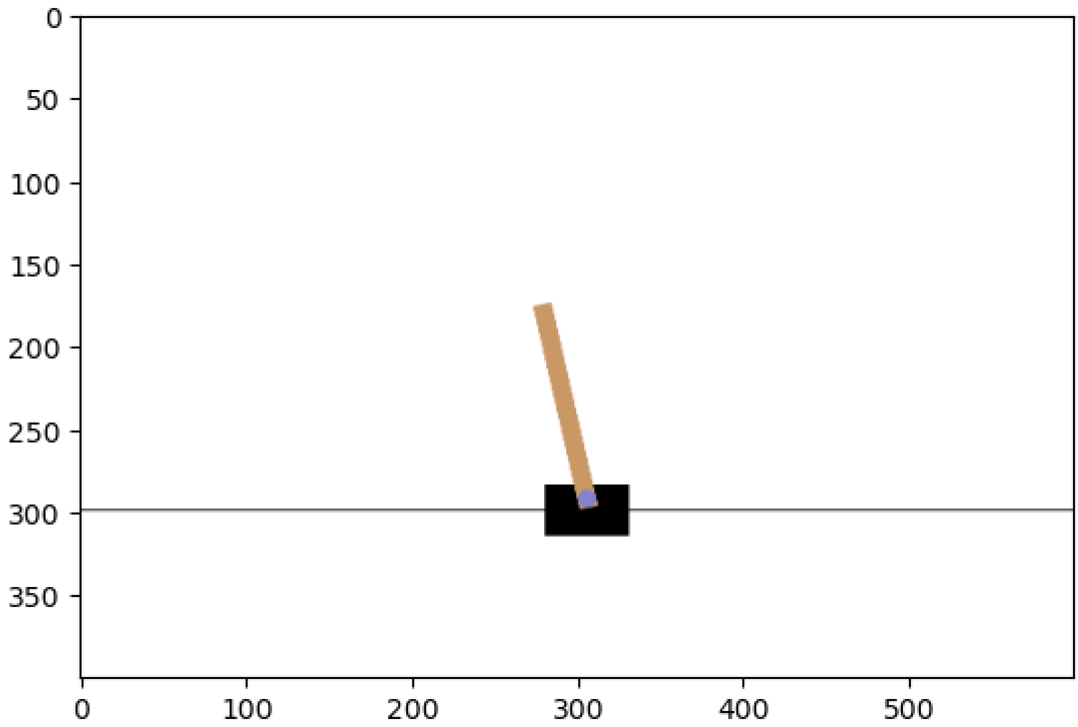
In [ ]:
```
env = gym.make("CartPole-v1", render_mode='rgb_array')
# Test the trained agent
frames=[]
for episode in range(10):
    state, _ = env.reset()
    done = False
    while not done:
        # Discretize the state
        state_discretized = discretize_state(state)

        # Choose the action with the highest Q-value (exploitation)
        action = np.argmax(q_table[state_discretized])

        # Step in the environment
        state, reward, terminated, truncated, _ = env.step(action)

        # Render the environment (to visualize the agent)
        im = env.render()
        frames.append(im)

        if terminated or truncated:
            done = True

fig = plt.figure()
img = plt.imshow(frames[0])

def update(frame):
        img.set_data(frames[frame])
        return img

ani = FuncAnimation(fig, update, frames=len(frames), interval=50)
ani.save("CartPoleTrained.mp4", fps=10, writer='ffmpeg')
plt.show()

env.close()
```



- **Visualizing agent behavior**: After training, the agent will follow the learned policy and you will see it balancing the pole using the **CartPole-v1** environment.
- `env.render()` : This will render the environment so you can see the agent's movements during each step.

In [ ]:
```
env = gym.make("CartPole-v1", render_mode='rgb_array')
# Test an untrained agent that acts randomly
frames=[]
for episode in range(10):
    state, _ = env.reset()
    done = False
    while not done:
        # Discretize the state
        state_discretized = discretize_state(state)

        # Act randomly
        action = np.random.randint(0,2) #randomly generates either 0 or 1

        # Step in the environment
        state, reward, terminated, truncated, _ = env.step(action)

        # Render the environment (to visualize the agent)
        im = env.render()
        frames.append(im)

        if terminated or truncated:
            done = True

fig = plt.figure()
img = plt.imshow(frames[0])

def update(frame):
        img.set_data(frames[frame])
        return img
```

```
ani = FuncAnimation(fig, update, frames=len(frames), interval=50)
ani.save("CartPoleUntrained.mp4", fps=10, writer='ffmpeg')
plt.show()

env.close()
```



Final Thoughts

- **Q-learning**: This basic Q-learning algorithm allows you to see how the agent improves over time. As the agent learns, the total reward should increase.
- **Exploration vs Exploitation**: You'll observe that early in training, the agent explores a lot, but over time, it starts exploiting what it has learned.
- **Visualization**: The plots and animations show how the agent's performance improves, and the rendered environment shows its progress in real-time.

This tutorial provided a hands-on, visual approach for teaching reinforcement learning. It allows you to experiment with Q-learning in a simple environment and understand how the agent learns through interaction with the environment.

Next Steps:

- Modify the exploration rate (`epsilon`) and observe how it impacts learning.
- Experiment with the number of episodes and see how the agent's performance changes.
- Try training the agent in different environments in Gymnasium, such as **MountainCar-v0** or **FrozenLake-v1**.