**Reinforcement Learning Lab - Part 2: Implementation Report**

**Task 1: Modified Gridworld Implementation**

**Environment Description**

* 8×8 grid with stochastic movement (80% intended direction, 20% random)
* State space: 64 discrete states
* Action space: 4 actions (UP, DOWN, LEFT, RIGHT)
* Rewards: -1 per step, +10 for gold (terminal), -10 for bomb (terminal)

**Implementation**

* **Random Agent**: Baseline comparison
* **DQN Agent**: Neural network 2→64→64→4 with experience replay
* **Hyperparameters**: Learning rate 0.001, batch size 32, buffer size 10,000
* **Epsilon values tested**: 0.1, 0.5, 0.9
* **Training**: 50 episodes per epsilon value

**Results**

* **Random Agent**: Poor, inconsistent performance
* **DQN Performance**:
* ε = 0.1: Final = -59.68, Best = -3.00
* ε = 0.5: Final = -116.72, Best = -7.00
* ε = 0.9: Final = -181.32, Best = -15.00
* **Best Performance**: ε = 0.1 achieved fastest convergence
* **Policy Visualization**: Q-value table showing optimal actions per state

**Task 2: LunarLander DQN Implementation**

**Environment Description**

* LunarLander-v3 environment (OpenAI Gym)
* State space: 8-dimensional continuous
* Action space: 4 discrete actions
* Success threshold: Average score ≥ 200 over 100 episodes

**Implementation**

* **Two Neural Networks**: Online and target networks
* **Architecture**: 8→64→64→4
* **Features**: Experience replay (100K buffer), epsilon decay (1.0→0.01)
* **Hyperparameters**: LR=5e-4, batch=64, target update every 10 steps
* **Training**: 500 episodes (increased from initial 200)

**Results**

* **Target**: ≥200 average score
* **Achieved**: **Environment solved in 322 episodes!**
* **Final Performance**: Average Score = 201.24
* **Learning Progress**:
* Episode 100: -133.73
* Episode 200: -60.71
* Episode 300: 23.44
* Episode 400: 170.50
* Episode 422: 201.24