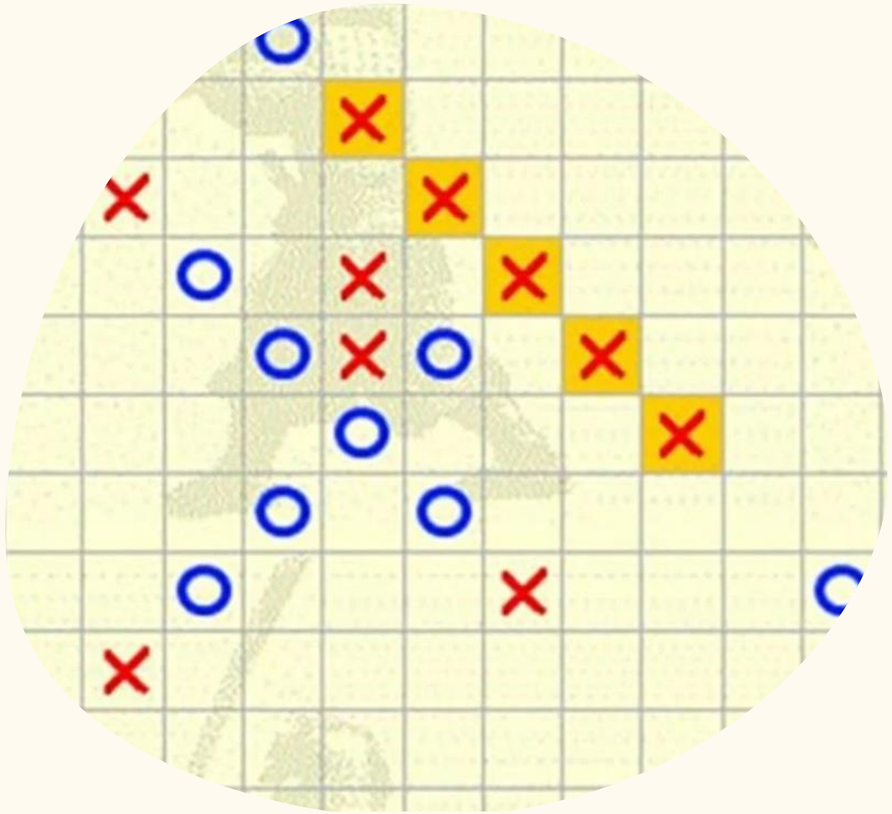
Phenikaa University

Tic-Tac-Toe: Mastering the Classic Game with Reinforcement Learning

Nguyen Vu Phung Anh - 22010994

Table Of Contents

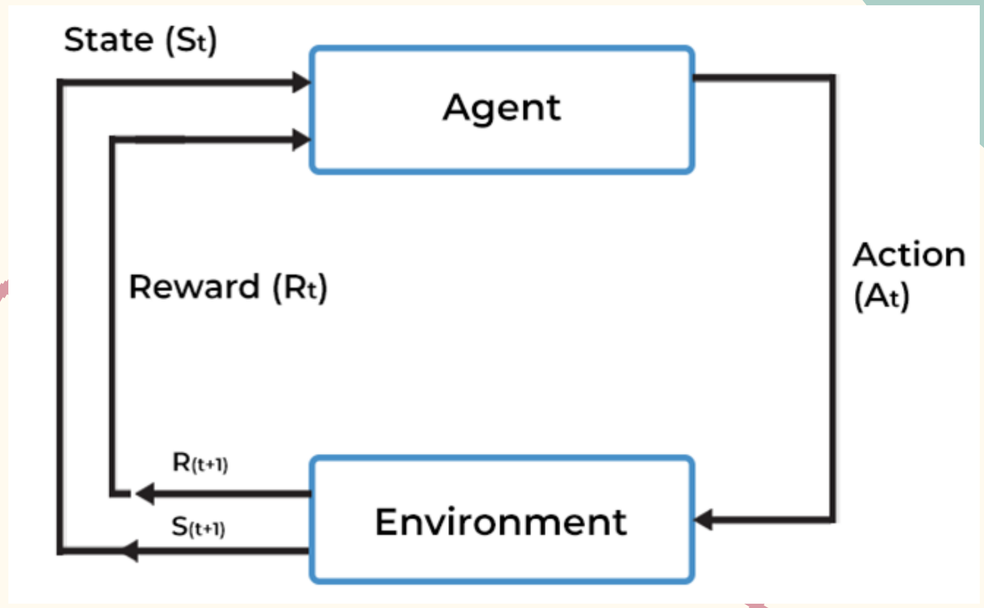


1. OVERVIEW AND OBJECTIVES

REINFORCEMENT LEARNING?

Reinforcement learning (RL) is an area of machine learning concerned with how software agents ought to take actions in an environment in order to maximize the notion of cumulative reward.

OR: RL is teaching a software agent how to behave in an environment by telling it how good it's doing.



Objectives

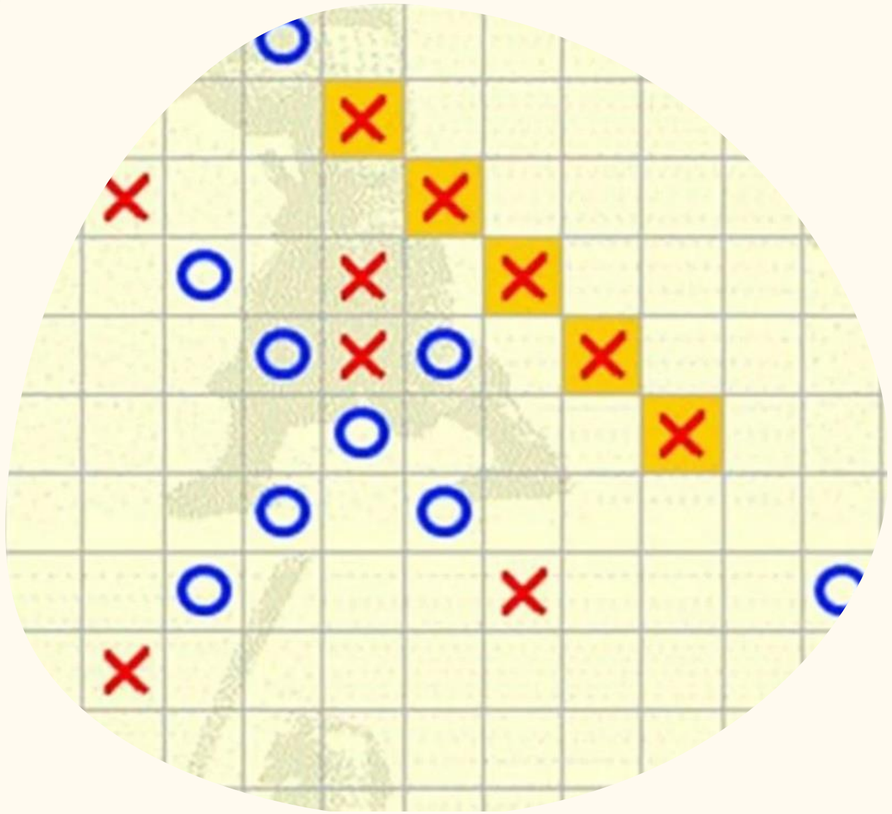
Value Iteration

Policy Iteration

Monte Carlo

SARSA

Q LEARNING



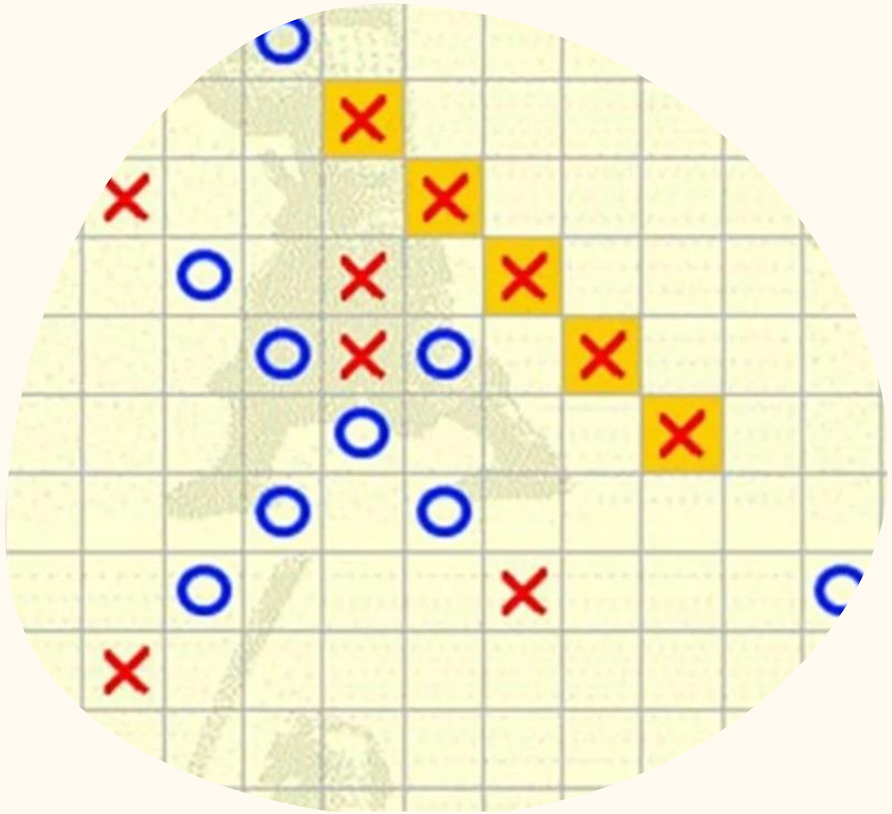
2.REWARD, ACTION, STATE OF TIC-TAC-TOE AGENT

ACTIONS AND REWARDS

ACTION	REWARD
Placing X/O	0
Game Draw	+0.5
Winning Move	+1

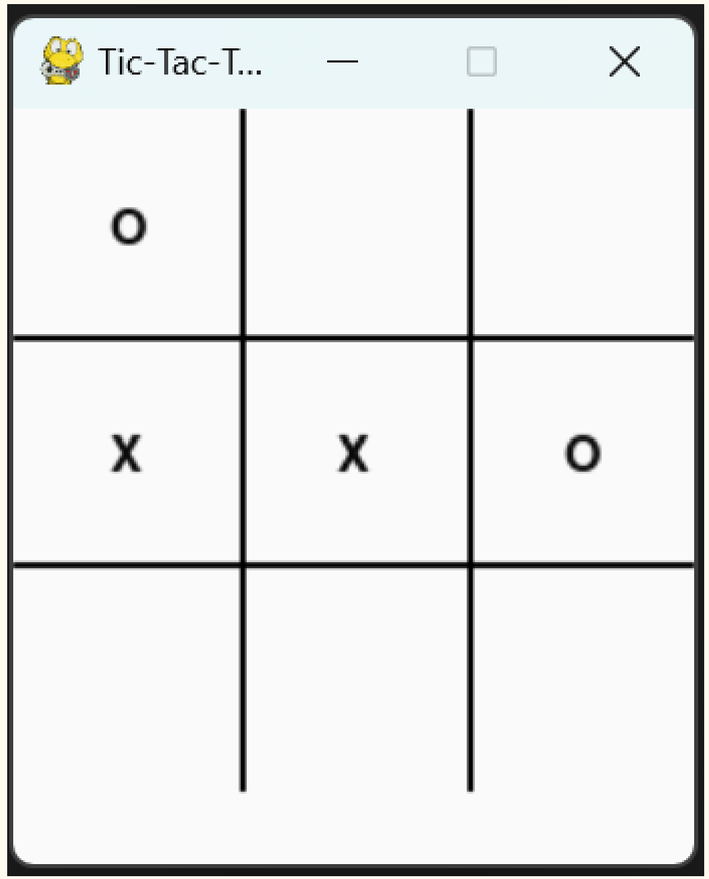
STATES

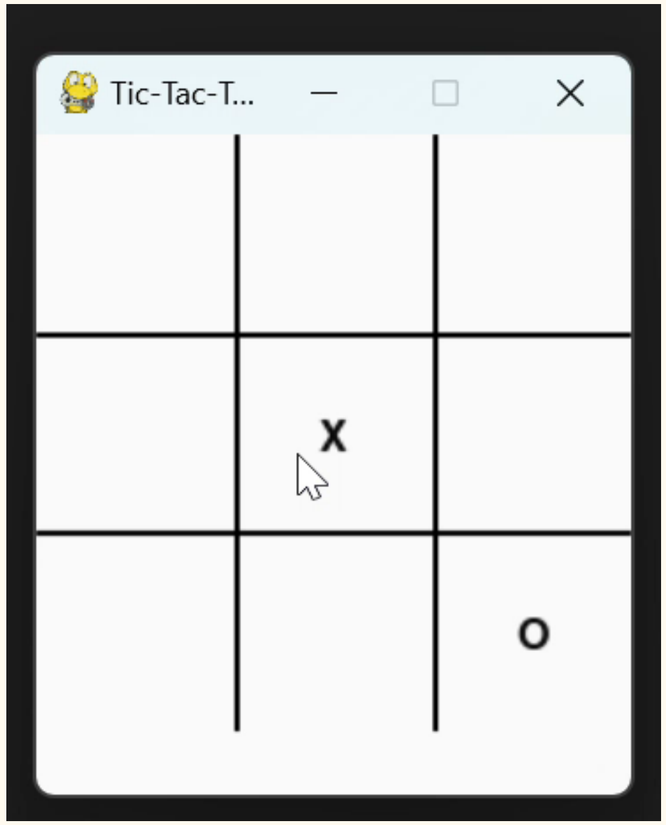
State	Description
Initial State	The board is empty, all 9 squares are unmarked.
In-progress State	The game is ongoing, some squares are filled with "X" or "O", no one has won yet, and it's not a draw.
Winning State (X/O)	A row, column, or diagonal has 3 identical "X"s or "O"s, the game ends with X or O as the winner.
Draw State	All squares are filled, but no one has won, resulting in a draw.



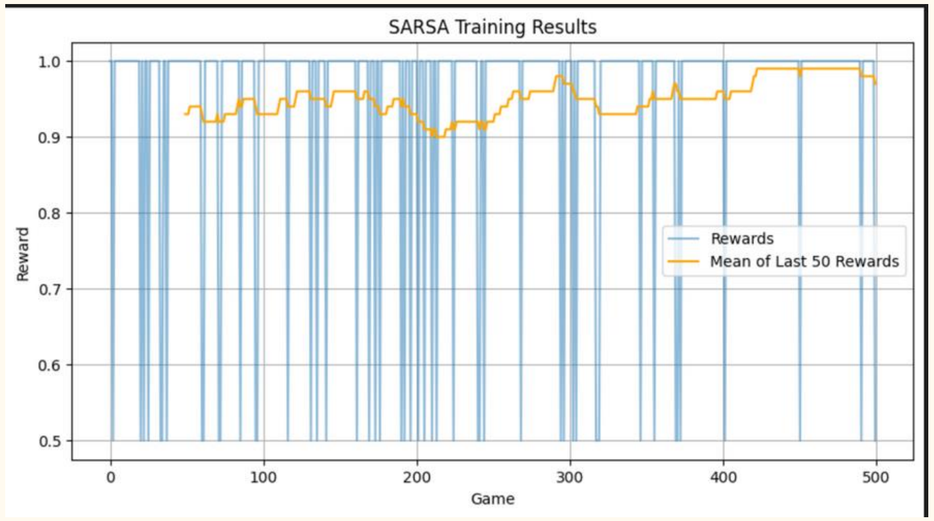
3. RESULTS AND CONCLUSION

USER INTERFACE AND DEMO

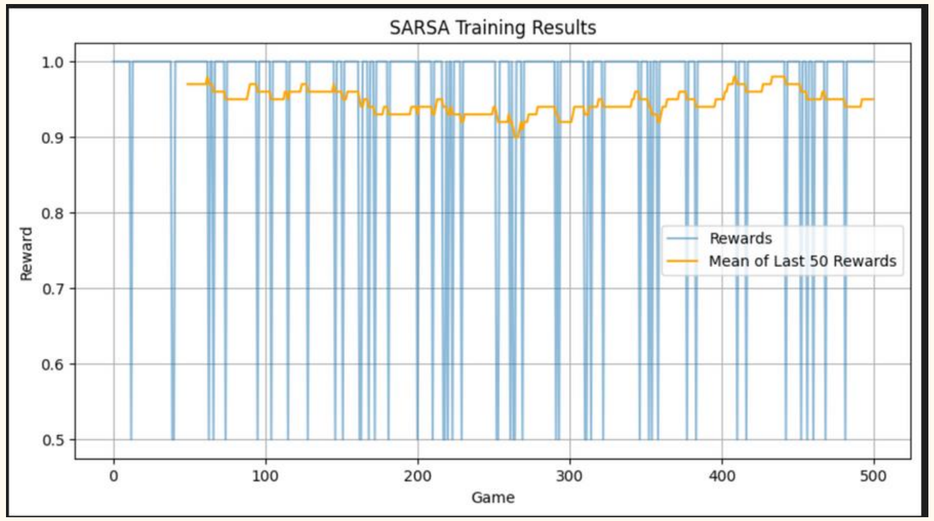




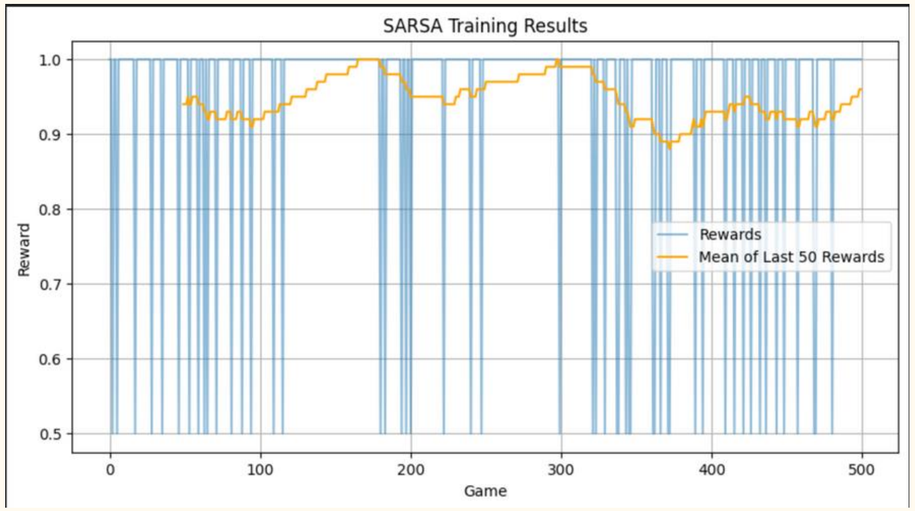
SARSA



Epsilon = 0.1

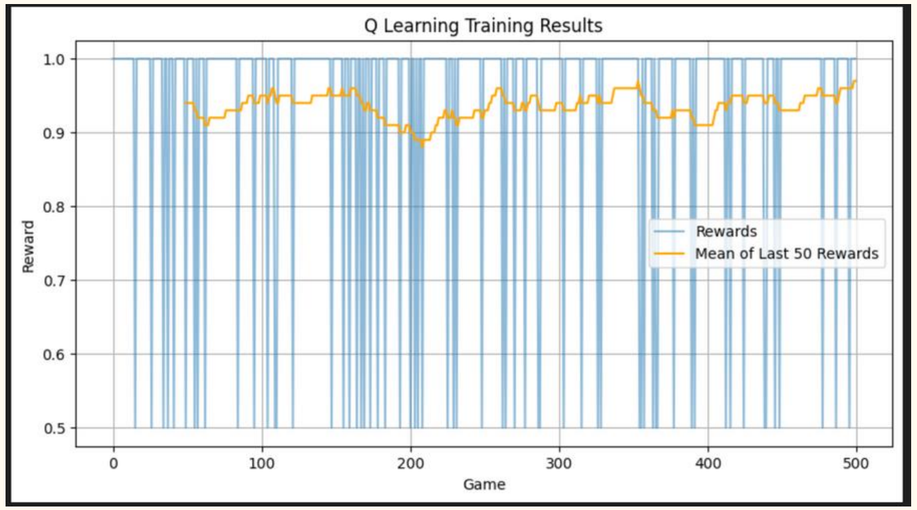


Epsilon = 0.5

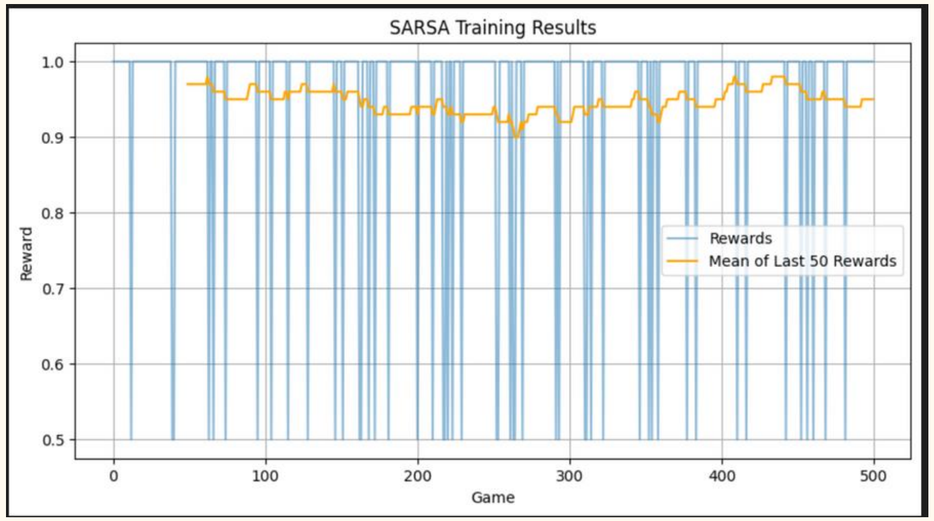


Epsilon = 0.9

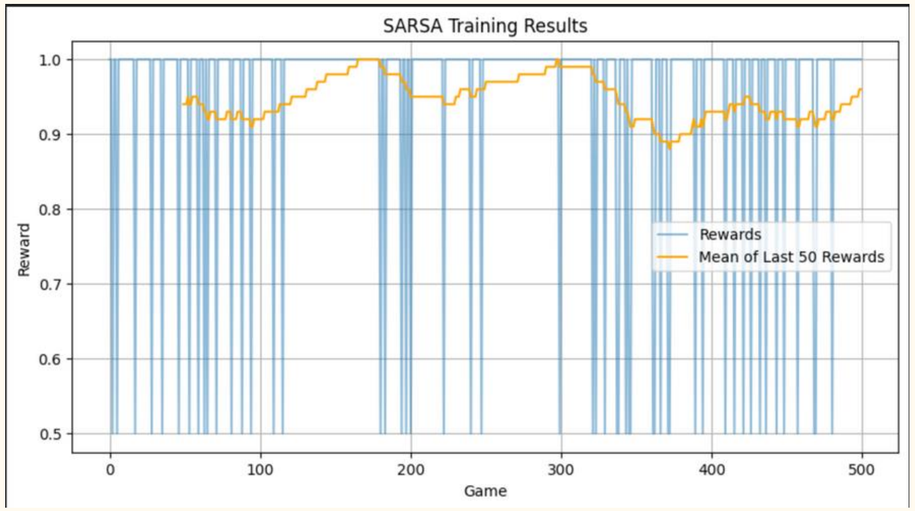
Q LEARNING



Epsilon = 0.1

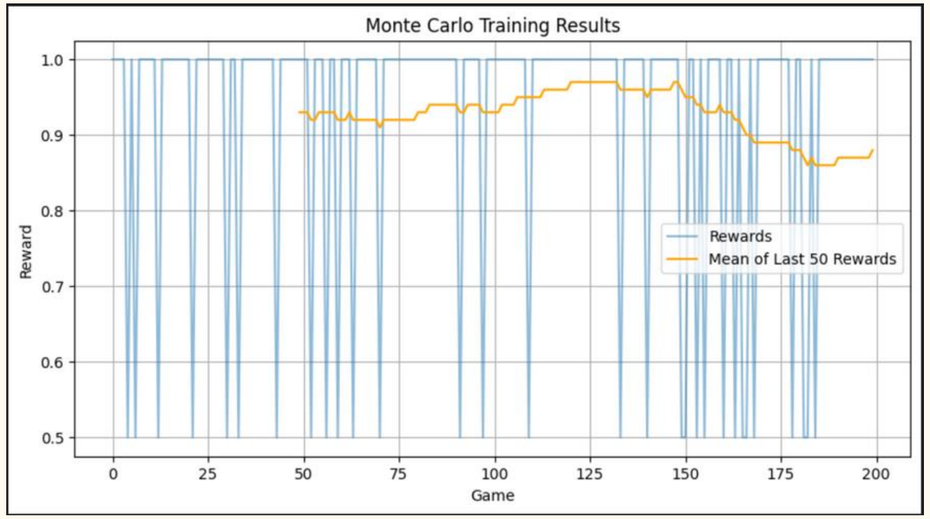


Epsilon = 0.5

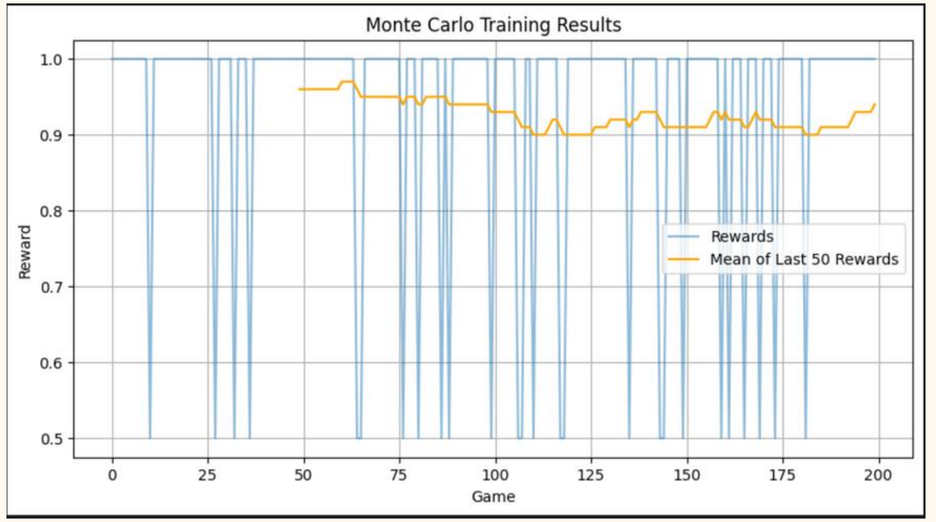


Epsilon = 0.9

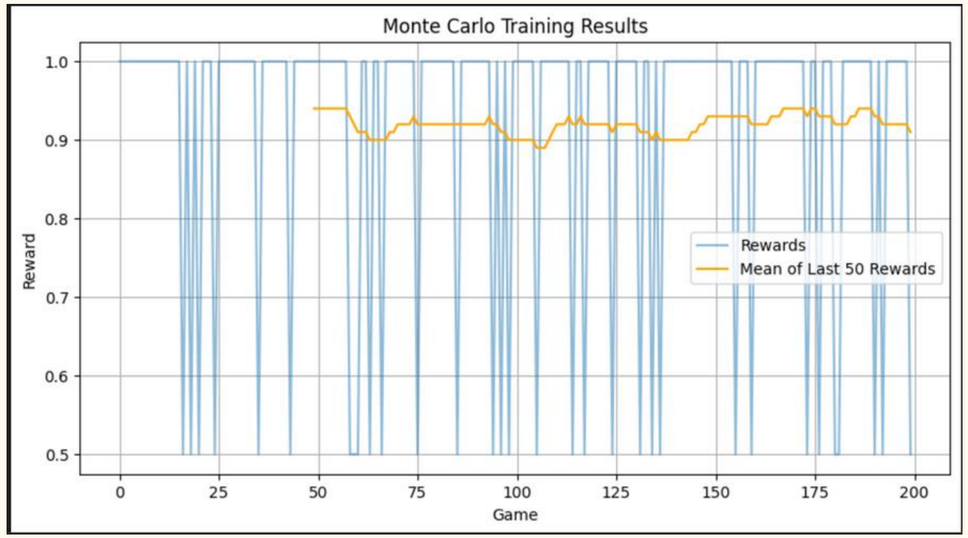
MONTE CARLO



Epsilon = 0.1

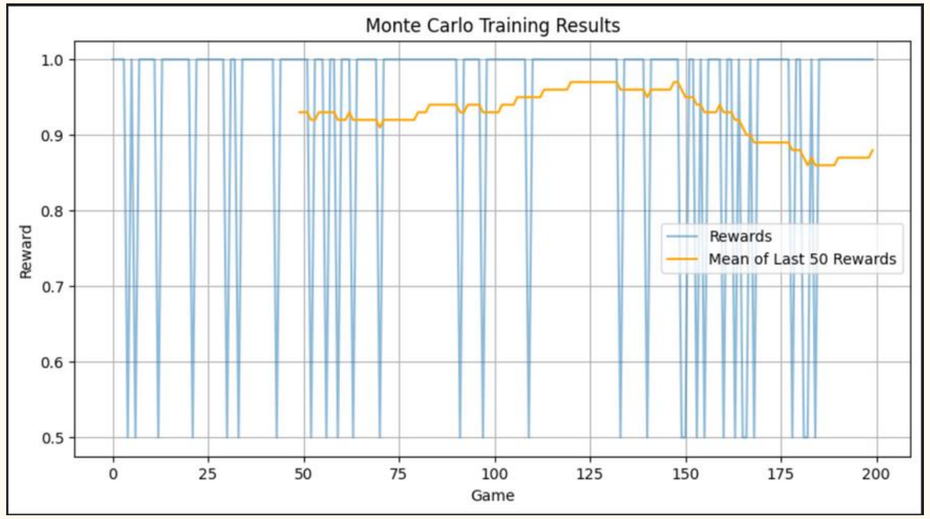


Epsilon = 0.5

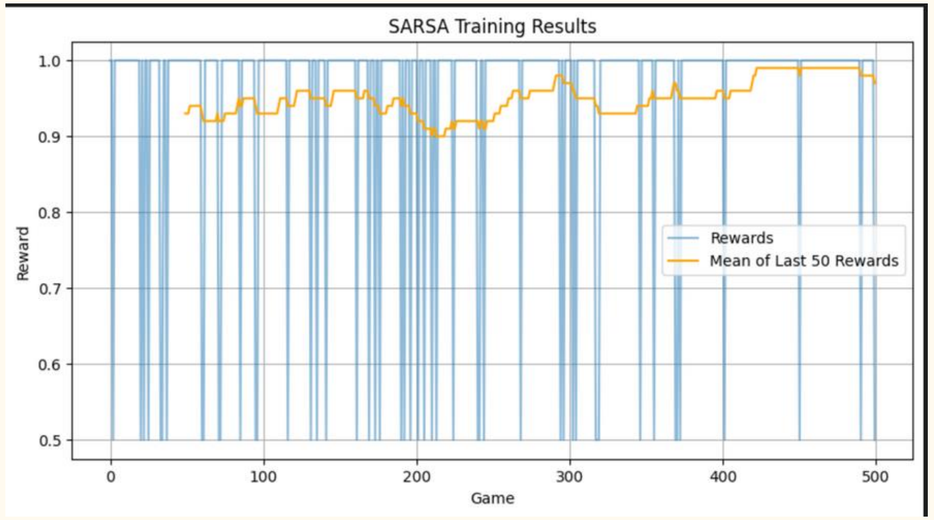


Epsilon = 0.9

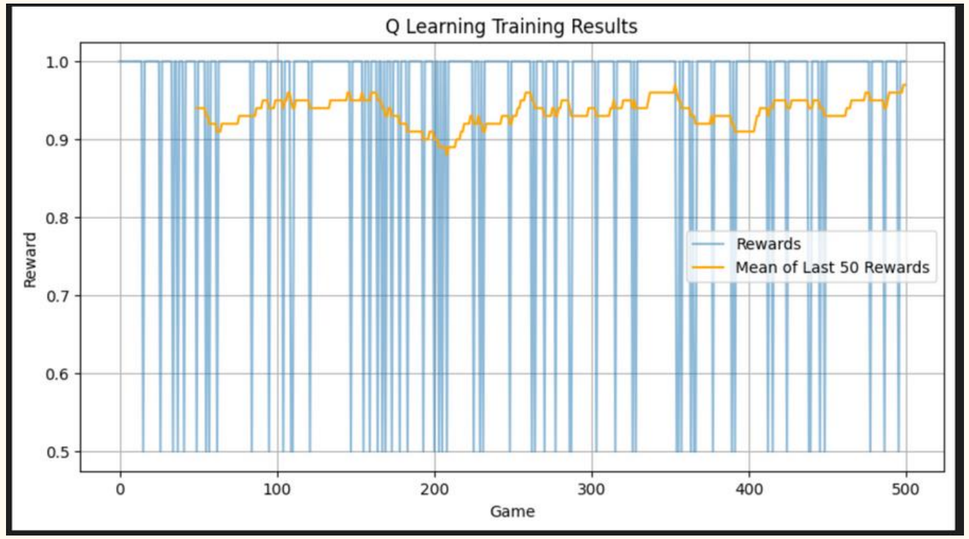
Compare Q-Learning, SARSA, Monte Carlo



Monte Carlo



SARSA



Q Learning

Nam

CONCLUSION

Q-Learning, which emerged as the most successful algorithm, was applied consistently across various scenarios with stable success.

Besides, SARSA and Monte Carlo also achieved relatively good results and achieved their goals.

Unfortunately, we excluded the results of Value Iteration and Policy Iteration due to their relatively lower performance in this context.

THANK YOU FOR LISTENING