

Cairo University Faculty of Engineering

Computer Engineering Department

CMPS458 Reinforcement Learning Report Template

## Team 1:

## Marwan Ahmed

## Noor Ahmed

Khalid Ali

*Supervisor:* Ayman AboElhassan

October 21, 2025

# Deliverables

Repo link: https://github.com/KhalidAli44/RL\_assignemnt\_1.git

Video record link: https://github.com/KhalidAli44/RL\_assignemnt\_1/tree/main/Assignment%201/videos

Questions:

1. ***What is the state-space size of the 5x5 Grid Maze problem?***

The size of the state-space in this assignment is 25, which are the 25 different positions available for the agent (5 \* 5).

1. ***How to optimize the policy iteration for the Grid Maze problem?***

By fixing the goal cell and the two bad cells, the size of the state-space was reduced to 25 as it only tracks the agent, instead of the general number of states which would be equal to 25 \* 24 \* 23 \* 25 = 345,000. This assumption allows policy iteration to be feasible and converge quickly on the Grid Maze problem.

1. ***How many iterations did it take to converge on a stable policy for 5x5 maze?***

From experiment observations, policy converged within 2 to 4 iterations (3 iterations on average), and did not exceed 5 iterations.

1. ***Explain, with an example, how policy iteration behaves with multiple goal cells.***

For multiple goal cells, policy iteration will favor actions leading to the nearest goal to get the maximum reward by crossing the shortest path. In case, for example, two goal cells are equidistant from a certain cell, the policy will likely favour the action leading towards the goal cell with the safest path due to the stochastic nature of the environment.

1. ***Can policy iteration work on a 10x10 maze? Explain why?***

Yes, for a 10×10 grid with only agent position (100 states) policy iteration is practical and will converge. Complexity increases but remains tractable (100 states × 4 actions).

1. ***Can policy iteration work on a continuous-space maze? Explain why?***

Not directly**.** Tabular policy iteration requires finite discrete states and explicit transition probabilities. Continuous states are uncountable so you cannot run tabular policy iteration on raw continuous space. This is why discretizations and approximations are used to formulate a tractable environment.

1. ***Can policy iteration work with moving bad cells (like Packman moving ghosts)? Explain why?***

Yes, if you model ghost positions in the state and know their transition model. Then the MDP state becomes an augmented tuple (agent position, first bad cell position, second bad cell position), and policy iteration can be applied — but the state space grows multiplicatively making DP often infeasible. In this case (with no overlapping terminal cells), the size of the state-space will be equal to 25 \* 24 \* 23 \* 25 = 345,000.