# A logo with black text Description automatically generatedReinforcement Learning and Sequential Decision Making – CS5180 LAB

**Varun Samala**

# EX-0

**3**. In situations marked by uncertainty, a strategy informed by human expertise generally proves more effective than one relying on random choices. When individuals make decisions, they can thoughtfully assess various options, choosing actions with a higher probability of positive outcomes, thereby increasing the chances of achieving the intended objective. Conversely, a random approach lacks strategic thinking, resulting in unpredictable and often suboptimal results. Additionally, human-driven strategies hold the advantage of adaptability and evolution based on new information and experiences, while a random approach remains static and unable to enhance its effectiveness over time. This fundamental distinction underscores the superior reliability of human-guided methods in accomplishing predetermined goals compared to random strategies.

PLOT:

A graph of a graph

Description automatically generated

**Cumulative reward vs Steps for Random policy**

**4.Better Policy Strategy**:

The improved policy is designed to detect walls and corners, guiding the agent toward the goal efficiently. When the agent is in the middle of the room without walls or corners, the policy resorts to a random approach. The key feature is that, in the presence of walls or corners, the policy follows the wall, significantly increasing the probability of taking the correct action (0.5) compared to the random policy (0.25). This adaptive strategy enhances performance, enabling the agent to reach the goal more effectively.

Worse Policy Strategy:

The inferior policy, on the other hand, adopts a simplistic strategy by consistently moving to the right, irrespective of the agent's current state. This deterministic approach makes it highly unlikely for the agent to reach the goal position and receive a reward. Unlike the random policy that exhibits adaptability, the worse policy's rigid behavior results in poor performance over time, as it consistently takes incorrect steps that do not lead the agent to the goal position.

A graph showing a number of different colored lines

Description automatically generated

**Cumulative reward vs Steps for Better policy, Random policy and Worse policy**