Reinforcement Learning: Reinforcement learning is a method within machine learning where an agent learns to make decisions through interactions with an environment. In this approach, the agent performs actions based on the current state of the environment and receives either rewards or penalties in response. These rewards or penalties help the agent learn how to effectively achieve a goal. Typically, reinforcement learning operates episodically, beginning at a specific state and concluding when the agent reaches the goal or meets other defined conditions, such as losing in a game. Each episode is made up of timesteps, during which the agent takes action and subsequently receives a reward.

Scenario:

We have a reinforcement learning agent that learns to find the shortest and safest path from a start position to target for a UAV. There are obstacles on the way to the target, so the agent should consider the safety of the path and avoid obstacles. There are 5 obstacles numbered 0-4. The agent has two goals:

1- finding the shortest possible path

2- finding the safest possible path by avoiding the obstacle

At each timestep, the agent selects an action that ends up in a new location and it receives a reward that evaluates the goodness of the agent's choice.

By taking a new action, the values of the following variables in the environment change:

- UAV position: (x,y) coordinate, - changes\_in\_distance\_with\_respect\_to\_the\_target:

- relative\_yaw\_UAV\_to\_objective, relative\_distance\_UAV\_to\_objectives

- relative\_yaw\_UAV\_to\_obstacle[0-4], and relative\_distance\_UAV\_to\_obstacle[0-4]

The positive value of changes in distance to the target shows a decrease in distance and vice versa.

We consider the agent in an unsafe location if the distance from the UAV to any obstacle is below 0.175. - Assume units in metric system-

Hints:

* Consider an action aligned if it moves towards the target.
* An action that is not aligned might cause a shift in direction, thus there could be a more direct path.
* Evaluate the actions based on the changes in distance, safety, alignment.

Answer the following questions.

1. At timestep 02, did the agent select a good action?

**Answer**: while the action at timestep 2 was effective in moving the UAV closer to the target, it was not a fully 'good' decision because it did not maintain the requisite safety margin from all obstacles, particularly obstacle0. Balancing both goals more effectively would be preferable in future actions.

Strongly agree  Agree Neutral Disagree  Strongly disagree

1. Provide a list of timesteps that agent could select a safer action.

The list should look like this: timestep -> reason

**Answer**:

1. \*\*Timestep 2\*\* -> The distance from UAV to obstacle0 decreased significantly to a risky proximity of 0.12, well below the unsafe threshold of 0.175. Safer actions could be selected to increase the distance from this and other closely positioned obstacles like obstacle2.

2. \*\*Timestep 3\*\* -> Although there is improvement from the previous step, the distance to obstacle0 is still just at the edge of safety at 0.19, suggesting that further action away from obstacle0 would enhance UAV safety.

Strongly agree  Agree Neutral Disagree  Strongly disagree

1. In how many of the actions does the UAV get closer to the target?

**Answer**: All actions

Strongly agree  Agree Neutral Disagree  Strongly disagree

1. How many of the actions are not aligned?

**Answer**:

To determine if the actions are aligned, we need to consider the changes in the variable "relative\_yaw\_UAV\_to\_objective," which indicates the angle between the UAV's heading and the direction towards the objective. There are \*\*2 definitely unaligned actions\*\* (timesteps 1 and 2) and one potentially unaligned action (timestep 3). Timestep 4 results in an aligned position.

Strongly agree  Agree Neutral Disagree  Strongly disagree

1. Could the agent reach the target faster?

**Answer**: Yes, the agent could potentially have reached the target faster if it had taken a more efficient, less deviated route while still maintaining the requisite safety margins for obstacle avoidance. This would involve improved path planning algorithms that optimize both the speed and safety criteria more effectively.

Strongly agree  Agree Neutral Disagree  Strongly disagree

1. What is your evaluation of the selected path by the agent considering all timesteps in this episode?

**Answer**: Overall, the agent managed both goals reasonably well but can improve on consistent safety margins while navigating towards objectives.

Strongly agree  Agree Neutral Disagree  Strongly disagree

1. Provide a list of timesteps that agent could select a better action.

The list should look like this: timestep -> reason

Answer: 1. Timestep 1 -> The distance to obstacle0 decreased from 0.66 to 0.41, nearing the unsafe threshold (below 0.175). Actions to increase safety buffer without deviating considerably from target path could improve performance.

2. Timestep 2 -> The distance to obstacle0 decreased further to 0.12, putting the UAV in an unsafe location (below 0.175). Selecting an action to increase this distance while still progressing towards the target would have been a better choice. Additionally, the relative distance to the target could potentially be maintained or reduced in a safer manner.

3. Timestep 3 -> Despite nearing the target effectively, the proximity to obstacle0 (0.19) and obstacle2 (0.24) threaten safety, suggesting that a slight adjustment in angle or positioning that could respect safety measures would be more optimal, even if it reduces speed towards the target slightly.

Strongly agree  Agree Neutral Disagree  Strongly disagree

1. Given explanation helps to understand the agent’s behavior easier.

Strongly agree  Agree Neutral Disagree  Strongly disagree