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. 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 meter.

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

The list should look like this: timestep -> reason

1. In how many of the actions does the UAV get closer to the target?
2. How many of the actions are not aligned?
3. Could the agent reach the target faster?
4. What is your evaluation of the selected path by the agent considering all timesteps in this episode?
5. Provide a list of timesteps that agent could select a better action.

The list should look like this: timestep -> reason