1. You are developing a reinforcement learning-based system for an autonomous drone tasked with delivering packages in an urban environment. The drone must navigate through a city, avoiding obstacles such as buildings, trees, and other flying objects. It must also optimize its flight path to minimize delivery time while ensuring safety and battery efficiency. The drone is equipped with sensors for detecting obstacles and GPS for navigation.

   Formulate this problem as an RL task. Specify the following components of the associated MDP:

   - State Space (S): Describe how you would define the state space for this problem.
   - Action Space (A): Define the action space available for the drone
   - Reward Function (R): Propose a reward function that encourages efficient and safe deliveries

Answer: Each entity in the State Space (S) shall include combination of details like:

- Drone co-ordinates (latitude, longitude)
- Drone altitude
- Drone velocity
- Weather conditions (like wind speed)
- Details of any close by obstacles, relative to current position of the drone
- Battery balance
- Delivery status

The Action Space (A) available for the drone shall include below possible actions:

- Move up
- Move down
- Move forward
- Move backward
- Turn right
- Turn left
- Deliver package
- Emergency landing

The Reward Function (R) can be proposed such that:

- Positive reward(s) will be awarded for –
    - Delivery of package
    - Avoiding obstacles
    - Less power consumption
    - Shorter time/path taken to deliver
- Negative reward (penalty) will be awarded for –
    - Failing to avoid an obstacle
    - Consuming power beyond reasonable, compared to distance covered
    - Taking longer route/time than reasonable

2. Discuss the trade-off between exploration and exploitation in Reinforcement Learning. Explain the ϵ-greedy strategy and how it accomplishes balancing this trade-off.

Answer: In RL, **Exploration** refers to the agent trying new actions to discover their effects and potentially identifying better policies while risking sub-optimal outcomes. On the other hand, **Exploitation** refers to the agent trying known actions that have already been proven to yield high rewards, at the cost of not discovering potentially better new actions.
For the RL to be effective, it is important to find right balance between both.
The ϵ-greedy strategy attempts to achieve this by proposing to initially start with high value for ϵ (Exploration rate/probability with which a random action should be picked) and then, gradually decrease it as the agent learns more about the environment.

3. In a simple MDP, an agent is in a state s, and the actions it can take can lead to the following outcomes:

- With probability 0.4, the agent transitions to state s', with reward R =10 , and *v(s')* = 5
- With probability 0.6, the agent transitions to state s'', with reward R = 3, and *v(s'')* = 2

The discount factor γ is 0.5. Using Bellman equation, find the expected value of state s.

Answer: As per Bellman's equation:

$$v(s) = P(s', r'|s,a)*(r' + \gamma.v(s')) + P(s'', r'' \mid s, a)*(r'' + \gamma.v(s''))$$

$$= 0.4 * (10 + 0.5 * 5) + 0.6 * (3 + 0.5 * 2)$$

$$= 7.4$$

Therefore, expected value of state **s** is **7.4**