

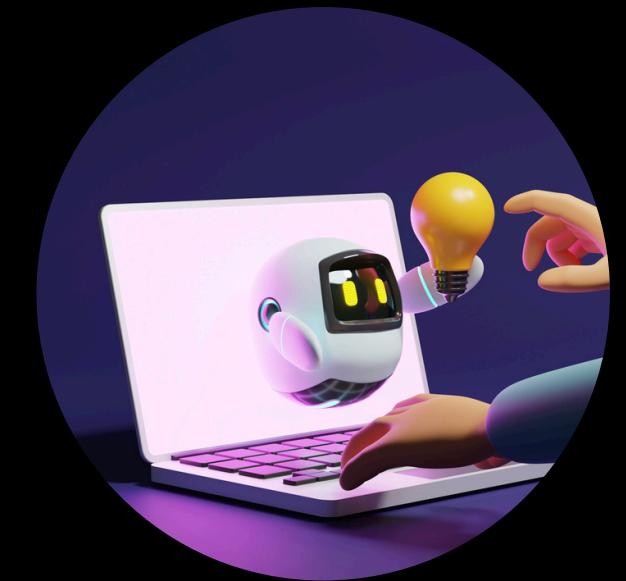
# 物件導向期末報告

組員：黃柏瑋、許家瑋



# 目錄

1. Part 1 展示
2. Part 2 說明與展示
3. Part 3 說明與展示



# Part 1 展示



# Part 2說明

## 參數設定

```
learning_rate_a = 0.12  
min_learning_rate = 0.00007  
discount_factor_g = 0.978  
learning_rate_decay = 0.000007  
epsilon_decay_rate = 0.0000785  
epsilon = 1
```



# Part 2說明

## 訓練方式

```
for i in range(episodes):
    state = env.reset()[0] # states: 0 to 63, 0=top left corner,63=bottom right corner
    terminated = False      # True when fall in hole or reached goal
    truncated = False       # True when actions > 200

    while(not terminated and not truncated):
        if is_training and rng.random() < epsilon:
            action = env.action_space.sample() # actions: 0=left,1=down,2=right,3=up
        else:
            action = np.argmax(q[state,:])

        new_state,reward,terminated,truncated,_ = env.step(action)

        if is_training:
            q[state,action] = q[state,action] + learning_rate_a * (
                reward + discount_factor_g * np.max(q[new_state,:]) - q[state,action]
            )

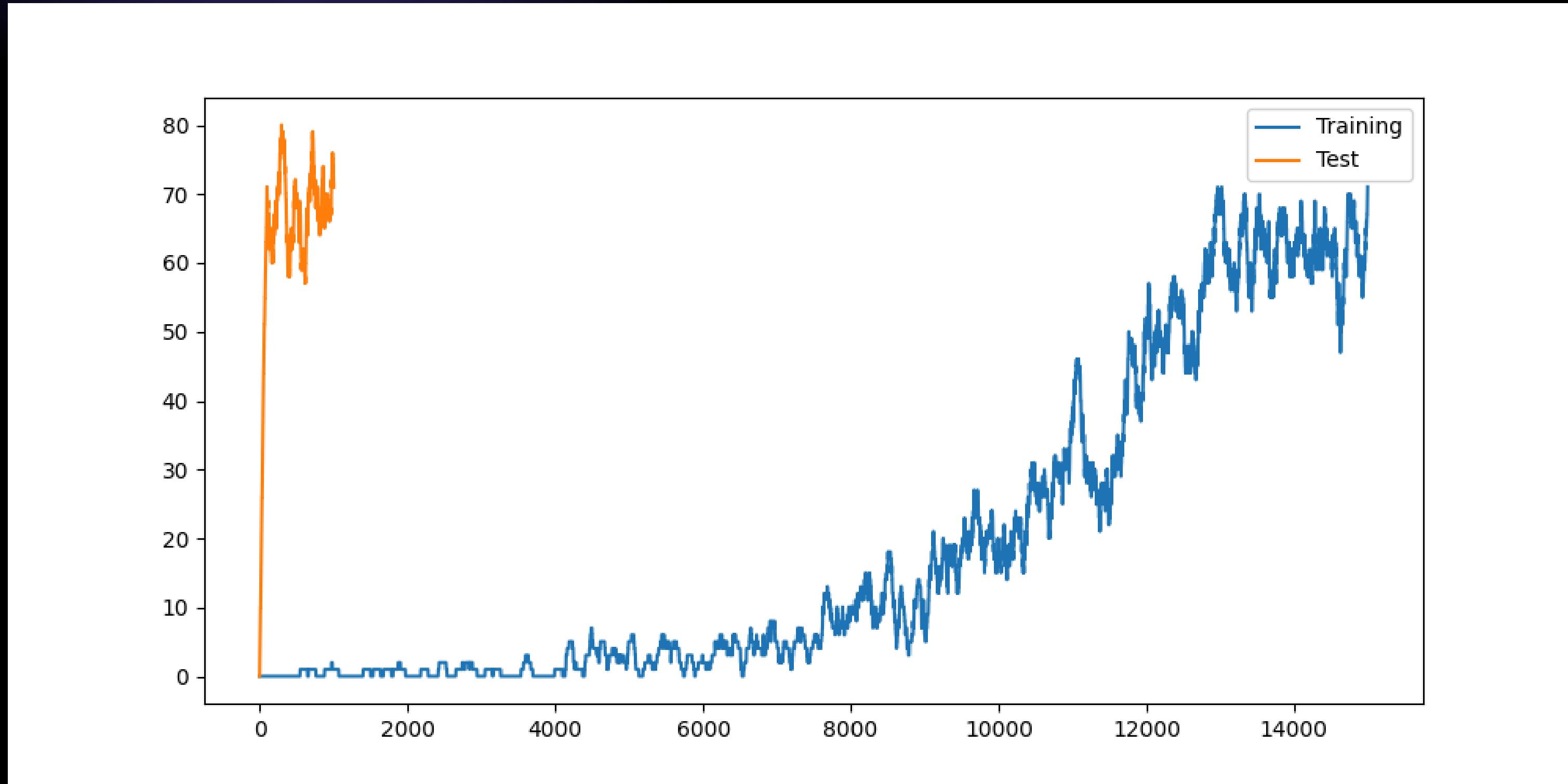
        state = new_state

        epsilon = max(epsilon - epsilon_decay_rate, 0)

    learning_rate_a = max(learning_rate_a - learning_rate_decay, min_learning_rate)

    if reward == 1:
        rewards_per_episode[i] = 1
```

# Part 2說明

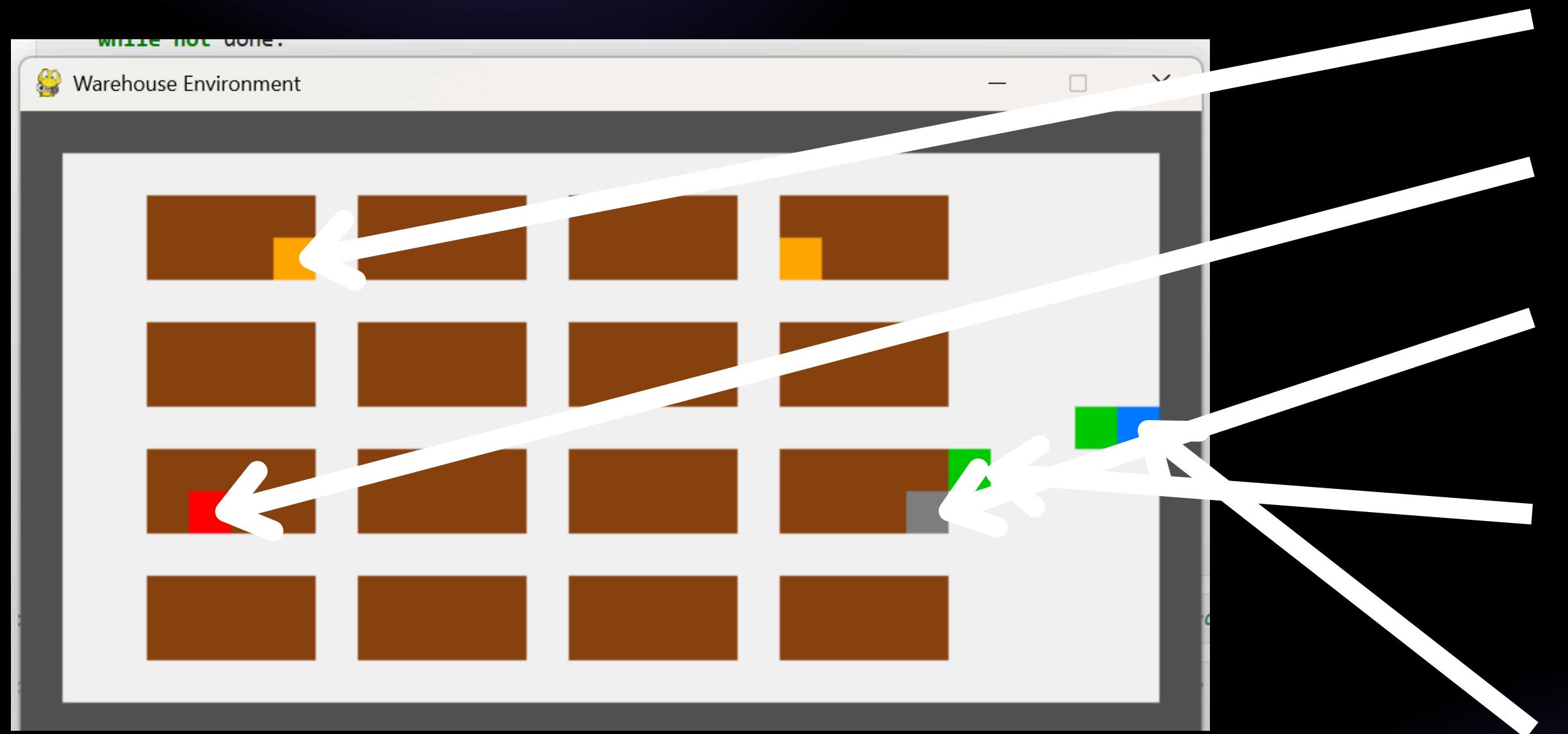


# Part 2 展示



# Part 3 custom warehouse simulation with env and RL

implements the Env class from gymnasium with custom rewards and behaviors



Regular package(orange)

Express package(red)

Heavy package(grey)

agent(green)

offload(blue)

Train RL model with PPO from stable baselines\_3

# Part 3 demo



報告結束