Final Project Instructions

This document outlines the requirements and tasks for the final project. All necessary materials have been provided, including:

- Report template
- Code files containing the PPO example, SAC/DDPG/TD3 implementations, and the UR5 environment
- A collection of related research papers in **reference_papers.zip**
- A detailed PPO project description in **Final_Project_PPO.pdf**

The objective of this project is to control a UR5 robotic manipulator to reach a target pose using reinforcement learning. Students are required to complete the following three components:

1. PPO Implementation

Students must work with the baseline PPO implementation and improve its performance by completing the following steps:

- Modify the reward function in the UR5 environment.
- Tune PPO hyperparameters to obtain stable and successful learning.
- Analyze training results using metrics such as episodic return curves and success rates.

For more detailed instructions on the PPO setup, training procedure, and reward configuration, please refer to **Final_Project_PPO.pdf**.

2. Off-Policy Algorithm Implementation (Choose One)

Select one of the following reinforcement learning algorithms:

- **SAC**
- **DDPG**
- **TD3**

Adapt the selected algorithm to work with the UR5 environment. (refer to ppo_continuous_action.py)
Train the agent and evaluate its performance using appropriate metrics and qualitative observations.

Note:
Example implementations of SAC, DDPG, and TD3 are included in the provided code package.
These files are located under the CleanRL directory:

```
cleanrl/cleanrl/sac_continuous_action.py
cleanrl/cleanrl/ddpg_continuous_action.py
cleanrl/cleanrl/td3_continuous_action.py
```

Students should also refer to **reference_papers.zip** for the original research papers related to the chosen algorithm.

3. Report

A written report must follow the required structure provided in: **final_report_{student_id}.docx**
The report should include:

- Theoretical background (RL, PPO, and the chosen off-policy algorithm)
- Implementation details (reward design, hyperparameters, algorithm-specific changes)
- Experimental results (training curves, success rates, etc)
- Discussion and conclusions

Students need to use the papers in **reference_papers.zip** when writing the Background/Theory section and when comparing PPO with the chosen algorithm.

Tips for Good Reports (Evaluation Criteria)

Your report will be evaluated based on the following key criteria.
Please ensure that each criterion is addressed clearly in your writing.

1. Explanation of Methods

You should clearly explain PPO and the chosen off-policy algorithm **in your own words**, including the main ideas and how they differ.

2. Implementation Details

Describe the important implementation decisions you made, including:

- How you modified the reward function
- Which hyperparameters you tuned
- Why you made those choices

3. Experimental Results & Analysis

Include training curves and relevant performance metrics, and **interpret the results**.
Explain:

- How PPO and the chosen algorithm compare
- Why certain behaviors or differences appeared