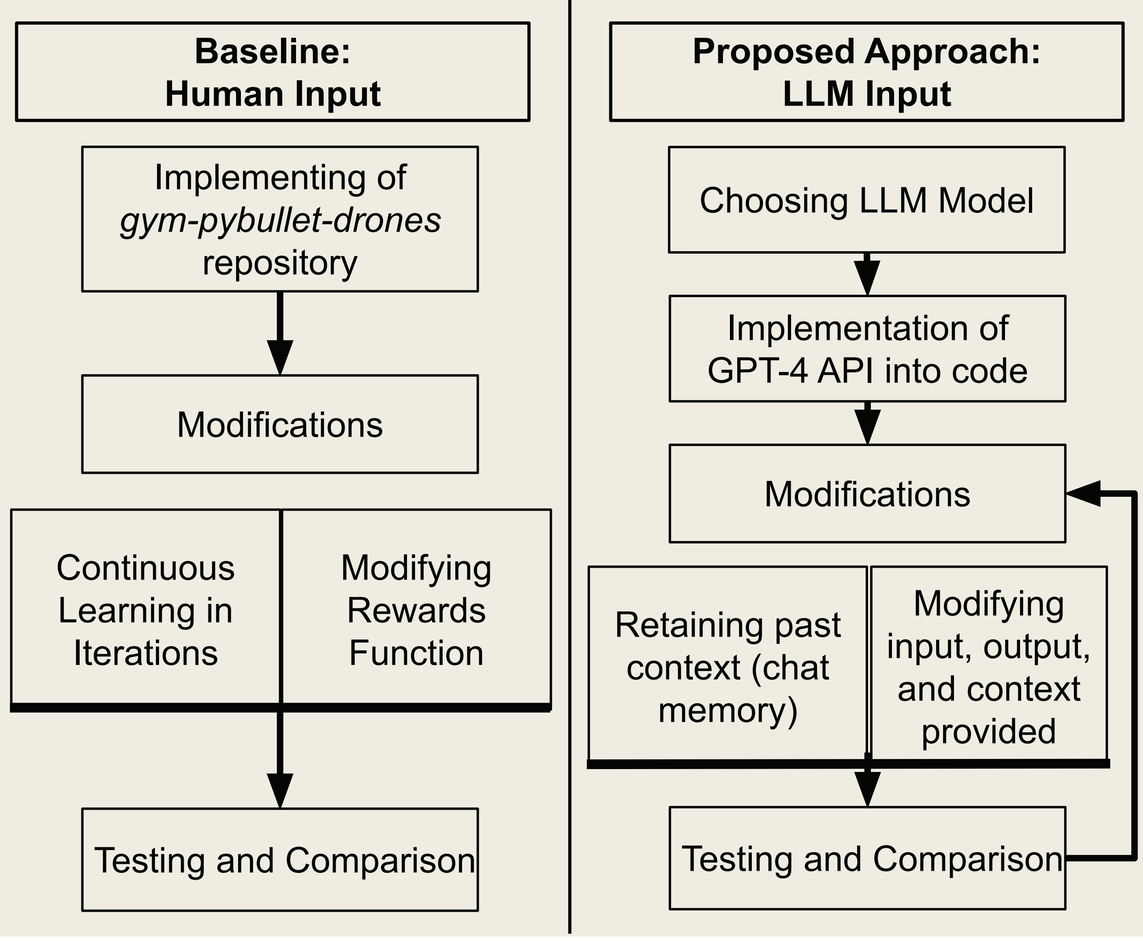
DATA-DRIVEN DRONES: AN ANALYSIS OF LARGE LANGUAGE MODELS IN REINFORCEMENT LEARNING FOR DRONES

Introduction

Project Goal	To investigate the potential of using LLMs in RL drone models for simple navigation
Problem Statement	One of the challenges of RL is the need for crafting of rewards function, for the agent to improve its sequence of decision making, to be done by hand
Proposed Approach	To investigate a novel use of LLMs to craft rewards functions iteratively for RL

Methodology



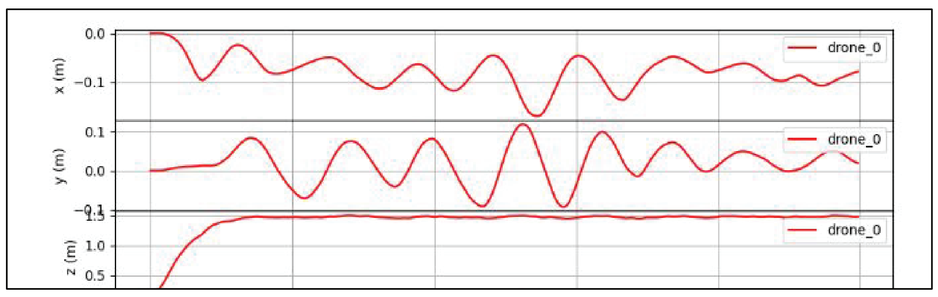
Results

Human Input

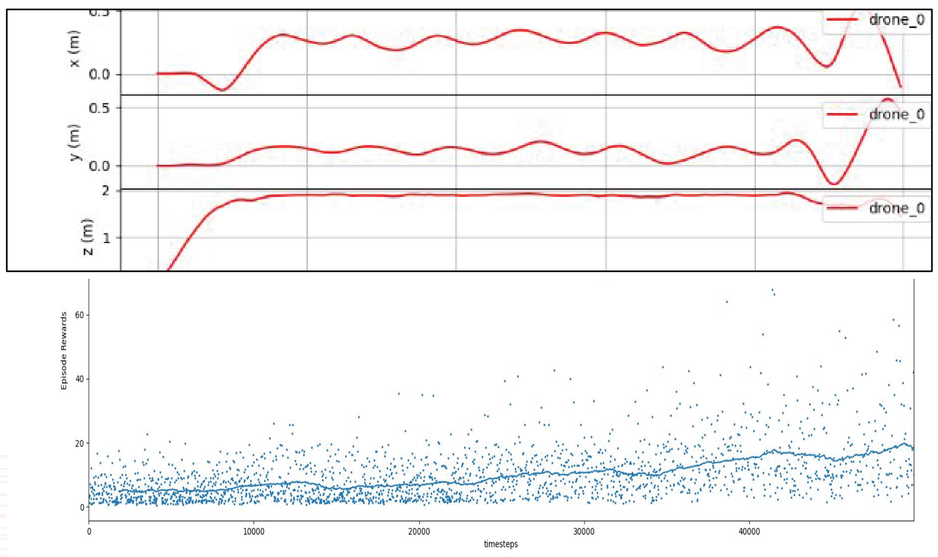
- Achieved stable flight to the coordinates (0,0,1) within **$2x10^5$ timesteps**
- Achieved stable flight to the coordinates (0,0,2) within **$3x10^5$ timesteps**
- **Human input** overall has **slower training output**

LLM Input

- Achieved stable flight to the coordinates (0,0,1) within **$1x10^5$ timesteps**
- Achieved stable flight to the coordinates (0,0,2) within **$2x10^5$ timesteps**
- **LLM** overall has **faster training output**



Human Input (0,0,2) Drone Dynamics



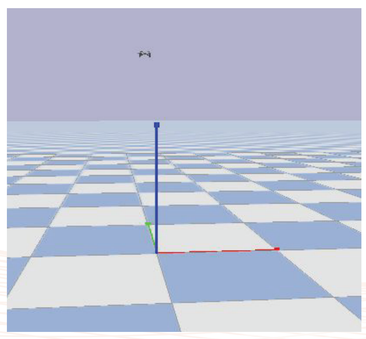
LLM Input (0,0,2) Drone Dynamics

Research Findings

- It is important for simulations to be able to **run faster than real time**, and/or multiple simulations to run in **parallel** as at least 10^5 timesteps were generally required to train the drone in RL
- LLM requires **clear and specified prompts** for desired output
- There are differences with using the API for ChatGPT4 as compared to the online version - there is **no memory** of the previous query and there is a **token limit** for each query. Since the GPT-4 model takes in an input called “context’ where one can add a dictionary of past conversations and set it as a context, each time a new query is input, we added it’s response into the dictionary.

Discussion and Future Work

- LLMs are more **efficient** and **effective** as compared to human input
- A possibility would be to include **human feedback** periodically in the RL process, further shortening training process
- A limitation of this analysis would be that testing was only done for **simple hover sequences**, so different flight paths and integrating obstacle avoidance can be explored
- This LLM training sequence has the potential to branch beyond simulation, and onto real life drone platforms as well



Members:
Ng Shi Qing Eugenia, River Valley High School
Vera Ong Liwen, Raffles Institution
Mentor:
Kuan Qi Heng Benson, DSO National Laboratories

YDSP Young Defence Scientists Programme
DSTA Defence Science & Technology Agency
DSO NATIONAL LABORATORIES