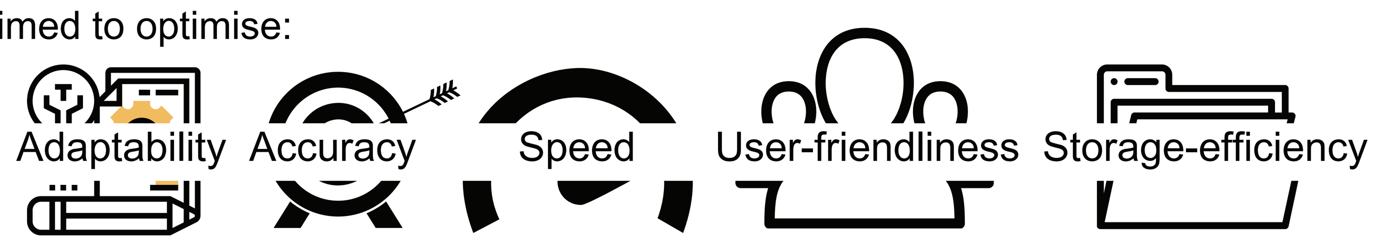
# OPTIMISING UAV DYNAMICS: USER-CENTRIC LARGE LANGUAGE MODEL INTEGRATION FOR DYNAMIC ADAPTATION IN CONTESTED ENVIRONMENTS

## Introduction

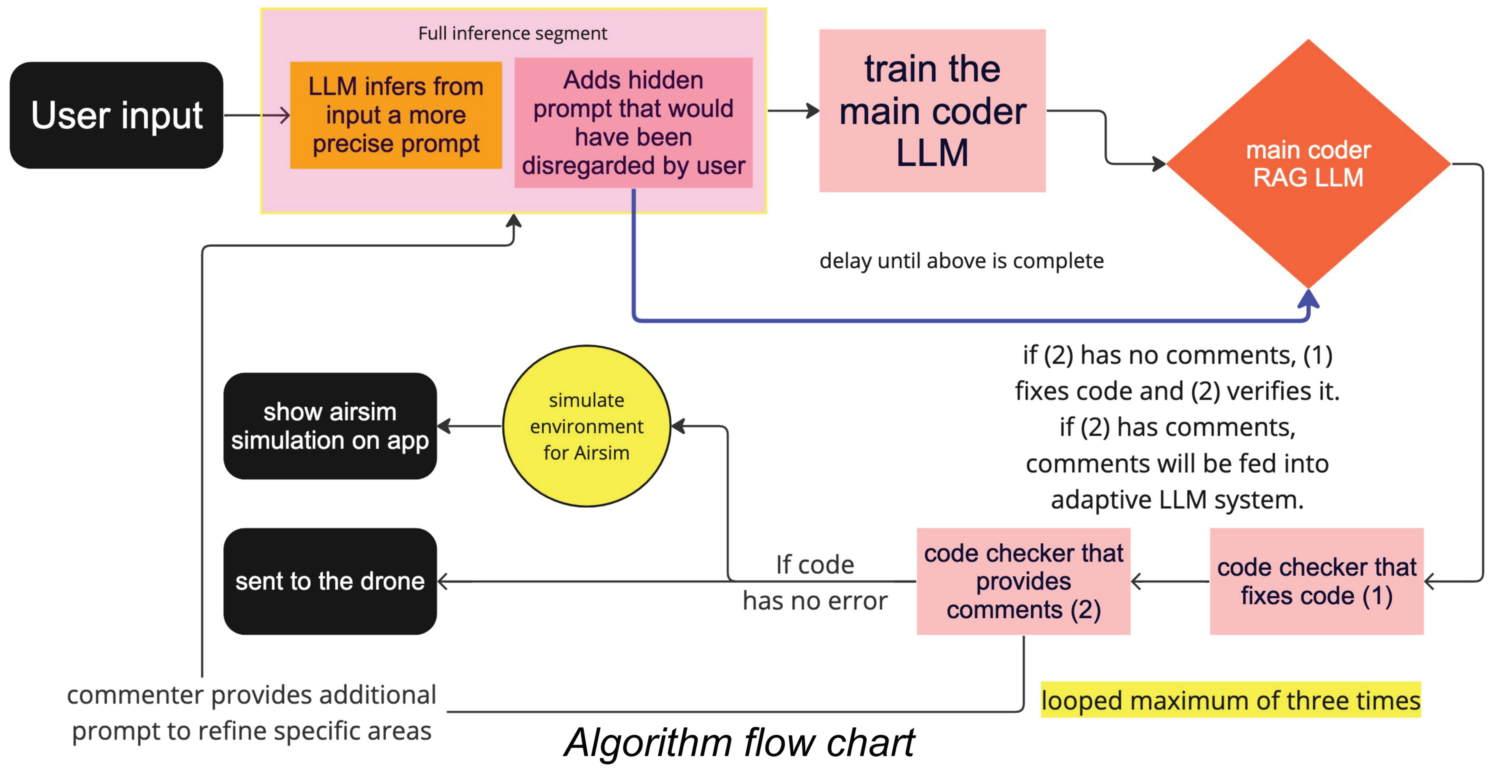
- In the realm of drone-based search and rescue or security missions that operate in **dynamic, contested environments**, no reliable model for coding drones' behaviours to **adapt swiftly** and accomplish novel tasks without prior training exists.
- In our work, we **address current shortcomings** identified in drone systems and Large Language Models (LLMs) through an **adaptive Retrieval Augmented Generation (RAG) system leveraging LLMs.**
- Aimed to optimise:




**Hypothesis** Our multi-agent adaptive RAG model will outperform existing LLMs in responding reliably to dynamic scenarios with simple natural language prompts.

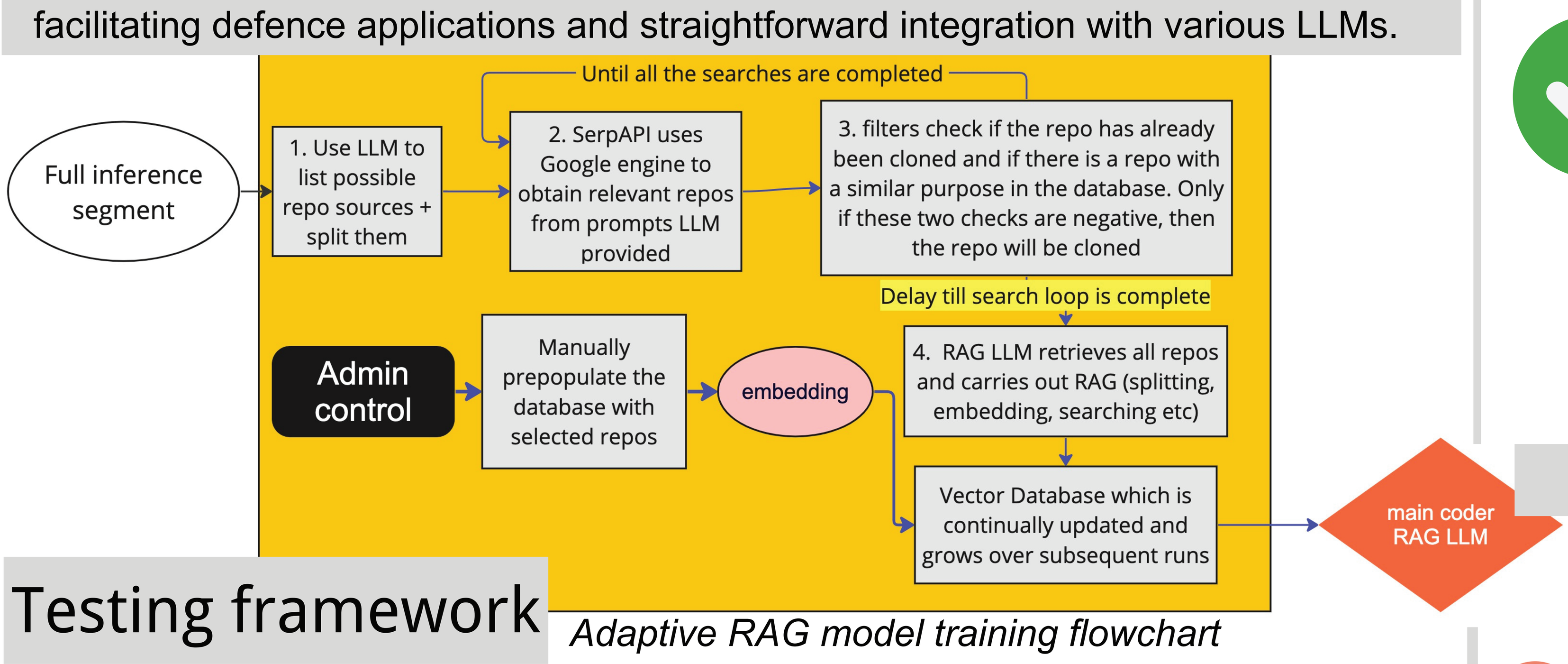
## Methodology

### Algorithm development




*Algorithm flow chart*

- Our approach combines an **adaptive RAG model with multi-agents**: 2 error correction modules, an inference engine, and a simulation component.
- The RAG system actively learns from **relevant Github repositories** downloaded through a **Google search engine**; the inference engine **processes** the simple natural language input into a more **precise prompt** for mission-specificity; LLM agents in the **error correction modules** fix syntax and logic mistakes.
- The system's design is compatible with **Airsim and Dronekit** and is **modular**, facilitating defence applications and straightforward integration with various LLMs.




*Adaptive RAG model training flowchart*

### Testing framework

We focused primarily on the **AirSim simulator** for our tests which is widely recognised for its trusted drone tests and realistic environments.

- There is no **widely accepted benchmarking framework** to evaluate our pipeline (given LLM integration with UAVs is an emerging domain),

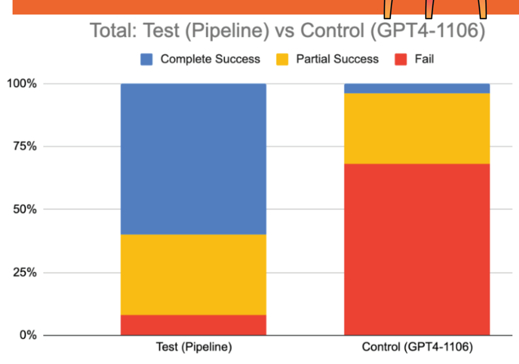
--> We developed **a new testing framework** showcasing reliability and task accuracy.

- This testing framework is categorised into **three stages** of varying difficulties

| 1. Basic UAV capabilities | 2. Operational requirements within contested environments | 3. Advanced tests involving technical requirements for challenging operations |
|---|---|---|
| basic manouevres such as takeoff and landing | such as flying in an equilateral triangle formation | such as obstacle detection and avoidance |

- We benchmarked our system against **GPT-4-1106** (which is a part of our pipeline).
- Results are classified as "Successfully Complete", "Partially Complete", or "Fail".

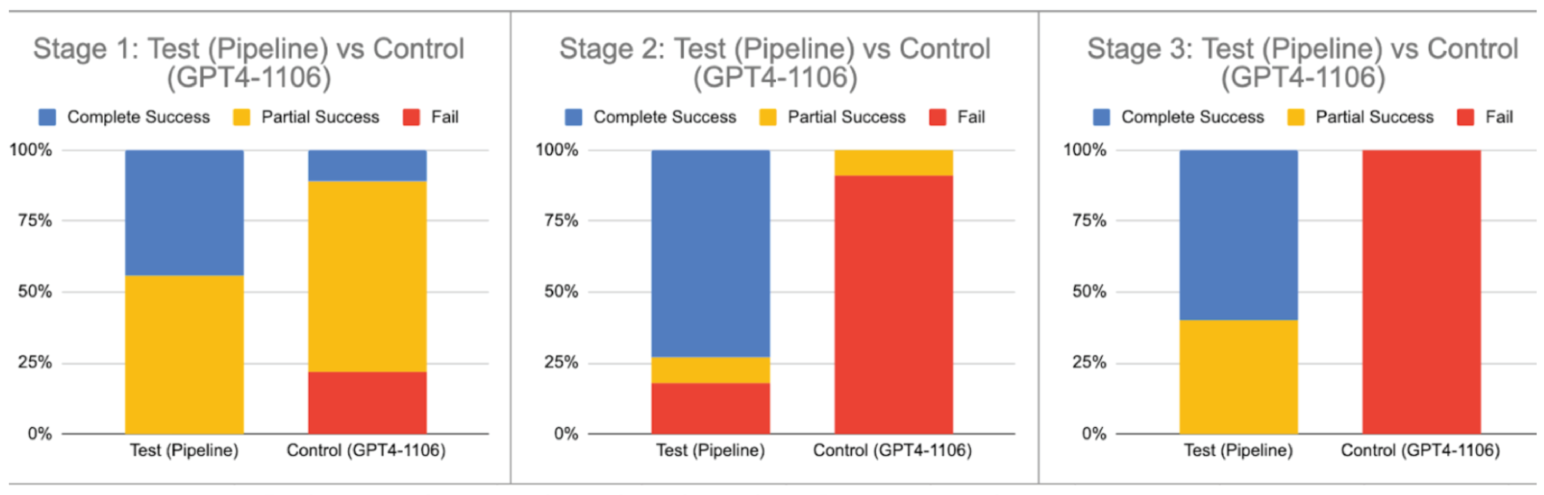
## Results




- Our pipeline **completely succeeded 60%** of the time, whereas GPT-4-1106 **completely succeeded 4%** of the time.
- We have success rates of **44%, 73%, and 60%** in basic, operational, and advanced technical tasks, respectively.

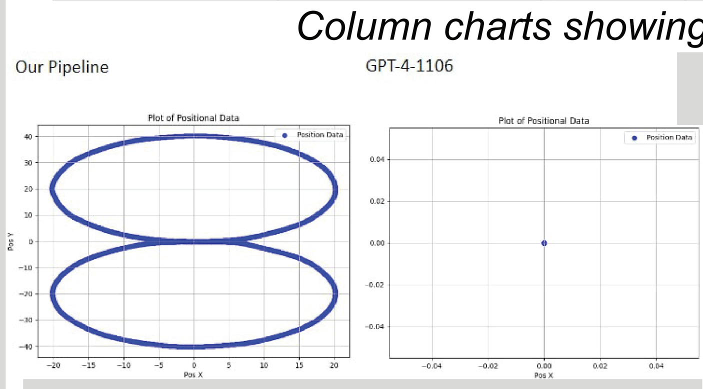
--> Significantly **<u>outperforms</u>** GPT-4-1006

*Level of mission success as a proportion of tests*




*Column charts showing levels of success by percentage*

### Example of test result




- The drone was instructed to **fly in a figure of 8**. No further instructions on how the drone was to plot its path nor was there a relevant repository in the initial database.
- System was **able to generate code**, which successfully flew a figure of 8 in AirSim.

### Discussion

- These findings support our theory that our multi-agent adaptive RAG model surpasses current LLMs in challenging scenarios, achieving a **<u>15-fold improvement</u>** in task success over GPT-4-1106 overall.

## Conclusion and Future work

Proved our initial hypothesis that our novel, context-specific model can **significantly improve the accuracy and reliability** of the drone code generated, with respect to existing powerful LLMs such as GPT-4-1106.

- Deployment in real life applications, ie.

Modern homeland security use-cases

Defence and search and rescue missions

### Future work

- Our algorithm is at a Technology Readiness Level (TRL) of TRL 4.
- Our future efforts will be directed at:

1. Real drone testing
2. User interface
3. Shortened processing duration

- to altogether push our algorithm to TRL 5-7.


Members:
Zhang Ning An Serene, Raffles Girls' School
Zeaus Koh Jin Rui, Hwa Chong Institution
Mentors:
Jeremy Wong, Defence Science and Technology Agency
David Wong, Defence Science and Technology Agency
Lim En-Lye Perrie, Defence Science and Technology Agency
Lim Gang Le, Defence Science and Technology Agency


Acknowledgements:
Benedict Lee Rui Yang, NUS High School of Mathematics and Science


YDSP Young Defence Scientists Programme

DSTA Defence Science & Technology Agency

DSO National Laboratories