Research Plan for UAV Dataset Generation and Model Fine-Tuning

Name: Ibrahim Odat

Github:

- https://github.com/3odat/PX4-Autopilot [main Project].
- https://github.com/3odat/Paper-Summarization [Literature Review].

Objective: Develop a dataset for the **PX4 Autopilot System**, fine-tune a model, and analyze security vulnerabilities of **LLMs in UAV systems**.

---

1. Dataset Generation

Goal:

Create a high-quality dataset tailored to UAV (PX4 Autopilot System) commands, interactions, and scenarios.

Approach:

1. **Synthetic Data Generation**
    - Use **GPT-4/DeepSeek** to generate/Search **structured UAV mission data**.
    - Define **task templates** based on UAV use cases (**e.g., takeoff, landing, navigation, object tracking**) / working on existing templates.

- Generate **Scene Descriptions** and **Task Descriptions** dynamically.
- Ensure dataset diversity by simulating **different environmental conditions and error scenarios**..
- Validate dataset **with domain experts** for realism.

2. **Dataset Preprocessing (Week 3)**

   - **Filter, clean, and standardize** dataset formats.
   - Convert data into **structured format (JSON, CSV, or MiniSpec-compatible format)**.
   - Evaluate dataset **using statistical distribution analysis** to ensure balance.

---

2. Model Fine-Tuning

Goal:

Fine-tune a **smaller LLM** (e.g., **Phi-2**) using the generated dataset to ensure **accurate UAV task generation**.

Approach:

1. **Preliminary Testing (Week 4)**

   - Run **LLAMA3.3 on existing UAV datasets** (if available) as a baseline.
   - Use prompt engineering to **evaluate performance before fine-tuning**.

2. **Fine-Tuning Process (Week 4 - 5)**

    - Finetune any smaller model (e.g., **Phi-2/TinyLLaMa** on **Scene Descriptions + Task Descriptions**.
    - Implement **Online Knowledge Distillation** from **LLAMA3.3 (Teacher)** → **Phi-2 (Student)** if fine tuning result is not accurate.
    - Use **KL Divergence Loss** for **logit-based distillation**.(Realistic).

3. **Evaluation & Optimization**

    - **Metrics:** Compare model **outputs** with **ground truth UAV commands**.
    - **Error Analysis:** Identify **failure cases & adversarial weaknesses**.
    - **Performance Tuning:** Adjust hyperparameters for **better accuracy & efficiency**.

---

3. Cybersecurity Analysis of LLMs in UAV Systems

Goal:

Investigate **security threats & vulnerabilities** of LLMs integrated with UAVs.

Approach:

1. **Attacks on LLMs in UAV Systems (Week 3-4)**

    - **Prompt Injection** (such as Task Manipulation in PX4).
    - **Adversarial Attacks** (malicious task modifications).

    - **Data Poisoning** (compromising training datasets)..
2. **Security Mitigation Strategies (Week 6)**
    - Implement **defensive prompt engineering**.
    - Develop **LLM access control & monitoring frameworks**.
    - Evaluate **cyber threat modeling for UAV LLMs**.

---

4. Literature Review (4-6 Papers per Week) Goal:

Develop a deep understanding of UAV AI models and their security.

Approach:

- LLM fine-tuning techniques & dataset generation strategies.
- **Adversarial attacks on LLMs** & **AI security research**.
- **Security implications of LLMs in UAVs** and PX4 autopilot research.
- Summarize findings in a **structured knowledge base**.

---

5. Reporting & Documentation Plan

Weekly Reporting (Every Sunday Night)

- **Progress updates** on dataset, model training, and security research.
- **Experimental results & observations**.
- **Challenges faced & proposed solutions**.

Daily Updates (Slack/In-Person)

- Quick notes on the day's **progress**.
- **Key takeaways & next steps**.

Final Research Deliverables

1. **Dataset & Model Report**
    - Final dataset description + model fine-tuning results.
2. **Security Analysis Report**
    - LLM vulnerability assessment in UAVs.
3. **Final Presentation & Paper Draft**
    - Paper draft on UAV dataset + model fine-tuning + security.
    - Slide deck summarizing **findings, challenges, and future work**.

---

6. Research Timeline Overview

- Synthetic Dataset Generation (GPT/DeepSeek) + Start Literature Review
- Handmade Dataset Generation + Initial Model Testing
- Dataset Preprocessing & Security Threat Analysis (Prompt Injection, Data Poisoning)
- Fine-Tuning (Phi-2) + LLM Security Analysis (Adversarial Attacks) Model Optimization & Security Mitigation Strategies
- Final Experiments + Report Writing + Paper Drafting