



## Development of Large Language Model (LLM) Based Coding Assistant

Gentleman Cadet Abdul Moiz

School of Avionics and Electrical Engineering, College of Aeronautical Engineering, PAF Academy Asghar Khan

Advisor: Sqn Ldr Dr. Raja Farrukh

Co-advisor: Wg Cdr Mohsin Khalil

Date of completion: September, 2025

### Abstract

This project develops a secure, offline coding assistant powered by custom-tuned open-source Large Language Models (LLMs). Designed for mission-critical use in the Pakistan Air Force (PAF), the system ensures confidentiality by operating without internet access. Integrated with a WebUI and a VS Code extension, it enables real-time code generation, debugging, and refactoring while optimizing performance for resource-constrained environments.

### Objective

To design and implement a secure, offline LLM-based coding assistant that integrates with IDEs, supports multi-language programming, and ensures reliable, energy-efficient performance in military-grade environments.

### Project Approach

**Custom Model Development:** Fine-tuned open-source LLMs with optimized parameters for coding tasks and PAF needs.

**Inference Engine:** Built a lightweight backend for fast, low-latency responses in offline environments.

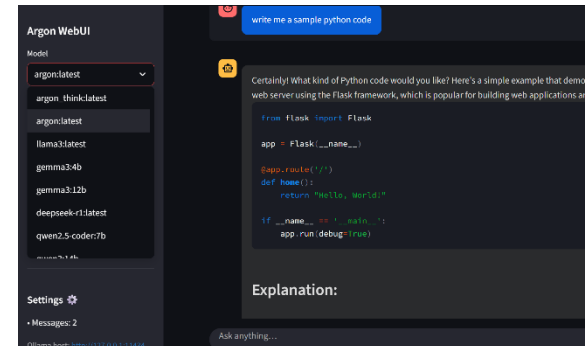
**WebUI:** Designed an interactive GUI for seamless testing, debugging, and visualization.

**VS Code Extension:** Integrated real-time code generation, debugging, and refactoring within the IDE.

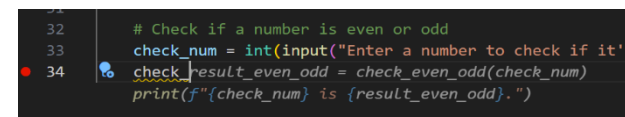
**Security & Testing:** Performed offline unit, integration, and acceptance testing under strict PAF security standards.

### Results

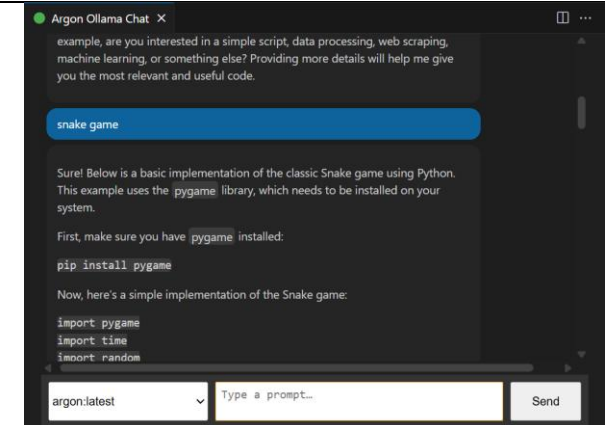
A user friendly web based GUI that has multi model support offering dynamic choice depending upon the nature of the task.



The assistant suggests and completes code directly inside the IDE when assistance is required.



An integrated plugin within VS Code that provides real-time coding assistance



### Conclusion

The project delivered a stand-alone LLM optimized for offline use, supported by a WebUI, backend, and VS Code extension. It provides intelligent coding assistance with low latency and energy efficiency.

This demonstrates the feasibility of deploying LLM-powered assistants in secure environments, improving productivity, reliability, and reducing human error. Future work will expand framework support, enhance debugging features, and refine models through continuous feedback.