Github

SYED AHAD ALI

github.com/Kagura-Ahad
linkedin.com/in/ahad-ali-01085a22a/
ahad2ali3@gmail.com
+92 334 3433587
Karachi, Pakistan

LinkedIn

EDUCATION

Habib University (Fresh Graduate)
BSc. in Computer Science
CGPA: **3.26/4.00**

Government Dehli College (2021)
Pre-Engineering (Intermediate)
Percentage: **93.9%**

SKILLS

Languages: Python, C++, SQL,
Frameworks/Libraries: FastAPI, MERN Stack, React, Tailwind CSS, Vite.js
Developer Tools: GitHub, LaTeX, Postman
Databases: MSSQL, SQLite
AI/ML Ecosystem: NLP, Model Training, Zero-shot Analysis, Multi-shot Analysis, CUDA, TensorRT

FINAL YEAR PROJECT (WON BEST FYP AWARD)

- Engineered a high-performance, real-time object detection pipeline in **C++** for a critical workplace safety system, initially leveraging GStreamer, OpenCV DNN (YOLOv9 ONNX with CUDA backend), and a custom **C++ ByteTrack** implementation for robust multi-object tracking.
- Achieved a **2x inference speed-up** by re-architecting the detection module using **NVIDIA TensorRT**; involved compiling the ONNX model to a highly optimized TensorRT engine and replacing OpenCV calls with a low-level, direct **CUDA API** implementation (managing cudaMallocAsync, cudaStream_t, and manual FP16/32 tensor parsing) crucial for **high-frequency** data processing and **low-latency** responses.
- Developed the full-stack solution by integrating the core **C++** video processing engine with a **FastAPI** backend for system control and data dissemination, and a **Vite.js/React** frontend for displaying live alerts and system status.
- Counted as an **Internship** at **DAWLANCE**.

CERTIFICATIONS

Introduction to Concurrent Programming with GPUs (CREDENTIALS) **August 2025**

- Understood the hardware differences that make GPUs ideal for parallel tasks.
- Learned how to develop concurrent software in Python and C++.
- Learned specifics of the CUDA hardware/software ecosystem by understanding the difference between is **CUDA Runtime API** and **CUDA Driver API**.

PROMINENT ACADEMIC PROJECTS

AURA: GraphRAG-Based GenAI Framework **Spring 2025**

- **Techstack:** Python, LangChain, Neo4j GraphRAG, Neo4j, Ollama (Llama 3.1), Azure AI, Nomic Embed
- **Description:** Implemented a GraphRAG system using structured knowledge graph for better performance on "global-level"questions usually hard for traditional vector-search RAG achieving **50%** better answers in terms of generality.
- **My Contribution:** Data Scraping and chunking strategy, then analyzed against a baseline RAG using an LLM-as-a-judge model to measure performance on metrics like answer diversity and detail.

Urdu GEC NLP Research (WON 2nd PRIZE in DURS 2025) **Spring 2024 – Present**

Tech Stack: Python, Hugging Face Transformers, mT0 (Transformer Model), ERRANT (adapted), UrduHack, Pandas, NumPy, Git.
Description: Led a research initiative to address data scarcity in a low-resource language (Urdu) by generating a large-scale synthetic corpus (240k+ pairs) to train a high-performing Grammatical Error Correction (GEC) model.
My Contribution: Engineered the core data generation pipeline by adapting the English ERRANT toolkit for Urdu, trained the final model, and curated a gold-standard test set via manual annotation with linguistic experts.

PROMINENT FREELANCE PROJECTS

Data Analysis Task Crime Pattern Analysis **February 2025**

Tech Stack: Python, Pandas, Scikit-learn (K-Means, PCA), MLxtend (Apriori), Matplotlib, Seaborn, Jupyter.
Description: Executed a freelance data mining project on UK police data to identify crime hotspots and co-occurrence patterns, delivering actionable insights into neighborhood crime profiles and their potential implications for policing.
My Contribution: Applied K-Means clustering to segment neighborhoods by crime frequency, used PCA for dimensionality reduction and visualization, and implemented the Apriori algorithm to discover high-confidence association rules (e.g., 86% confidence) between crime types.