

# Syed Muhammad Sarim Ahsan

sarim.ahsan101@gmail.com | +92 323 703 4858 | [GitHub](#) | [LinkedIn](#) | [HuggingFace](#) | [sarimahsan.dev](#)

## SUMMARY

Third-year Computer Engineering student at SSUET (GPA 3.5) with practical experience building and deploying AI/ML systems and full-stack web applications. Comfortable working across the full stack — from LLM pipelines and RAG systems to containerised production deployments. Currently supporting two live UK-deployed platforms as an Application Support Engineer while developing AI products independently.

## EDUCATION

<b>Sir Syed University of Engineering &amp; Technology — Karachi</b>	2023 – Present (Expected 2027)
B.S. in Computer Engineering · GPA: 3.5 / 4.00 · Currently in 3rd Year (6th Semester)	
<b>Beaconhouse School System</b>	Completed 2021
Matriculation · Grade A	

## CERTIFICATIONS & ACHIEVEMENTS

- Associate Data Scientist — [DataCamp](#)
- [Machine Learning Specialization](#) (Supervised ML + Advanced Algorithms) — Stanford / DeepLearning.AI (Andrew Ng)
- AMD AI Academy — 6 hands-on courses: LLM Serving (vLLM / MI300X), Diffusion Models, R1 Fine-tuning (Unsloth), RL with LLMs, AI Agents (MCP), HuggingFace on AMD
- Participant, [Meta Hacker Cup](#) — Global Competitive Programming
- 180+ LeetCode problems solved — Arrays, BSTs, Stack, BFS, DFS
- AI / ML Technical Writing — [Medium](#)

## EXPERIENCE

<b>Application Support Engineer — <a href="#">GCRD (UK)</a></b>	2025 – Present
<ul style="list-style-type: none"><li>• Sole technical owner of portalersm.com, a live student recruitment portal serving a UK-based education company</li><li>• Manages VPS infrastructure, Docker Compose deployments, Nginx reverse proxy, MySQL, and automated backup scripts with Gmail notifications</li><li>• Handles bug fixes, feature updates, and CI/CD pipeline maintenance via GitHub Actions</li></ul>	
<b>Application Support Engineer — <a href="#">Uni Student Hub (UK)</a></b>	2024 – 2025
<ul style="list-style-type: none"><li>• End-to-end ownership of unistudenthub.net, a call-centre platform for university student support</li><li>• Handles full deployment lifecycle: Docker orchestration, Nginx configuration, and ongoing feature development</li></ul>	
<b>IT Intern — IFFCO Pakistan</b>	2025
<ul style="list-style-type: none"><li>• Gained exposure to enterprise IT infrastructure and systems administration at a large multinational corporation</li></ul>	

## PROJECTS

AI / ML

<b>FlowForge — <a href="#">Multi-Agent Enterprise Sales AI</a></b>	2026
<ul style="list-style-type: none"><li>• A LangGraph pipeline of researcher, analyst, and writer agents that automates B2B outreach end-to-end</li><li>• Stack: FastAPI, Groq (LLaMA 3.3 70B), Tavily web search, React/Vite frontend; includes PDF export, Gmail integration, and lead scoring</li></ul>	
<b>MeezanGPT — <a href="#">Corrective RAG Chatbot</a></b>	2026
<ul style="list-style-type: none"><li>• Islamic banking Q&amp;A; chatbot built on a Corrective RAG (CRAG) pipeline — automatically verifies retrieved context before answering to minimise hallucinations</li><li>• Stack: FAISS / Pinecone, OpenAI / LLaMA, BeautifulSoup, PyPDF2, FastAPI</li></ul>	
<b>ArXiv Lens — <a href="#">AI Research Tool</a></b>	2026
<ul style="list-style-type: none"><li>• Summarises arXiv papers, generates force-directed citation knowledge graphs, and enables RAG-powered chat with any paper</li></ul>	

- Stack: FastAPI, React, LangChain, ChromaDB, Groq LLaMA 3; async parallelisation and glassmorphism frontend

**LexCorp RAG — [Legal Chatbot](#)** 2025

- RAG chatbot built for a law firm client using PyMuPDF, Qdrant, all-MiniLM-L6-v2 embeddings, Groq LLaMA 3, and FastAPI with GitHub Actions CI/CD

**Doctor AI — [Medical Symptom Checker](#)** 2024

- Web application where users describe symptoms or upload X-ray images to receive AI-assisted diagnostic suggestions — NLP + CNN backend via FastAPI and HuggingFace
- Published models on HuggingFace Hub: `medical_symptom_transformer` and `xray_disease_cnn_model`

*MCP Servers & Agentic Tools*

**[PSX MCP Server](#)** 2026

- MCP server providing LLMs (Claude, GPT) with real-time Pakistan Stock Exchange data — live prices, market trends, top gainers/losers, and financial insights

**[GitHub MCP Server](#)** 2026

- Read-only MCP server exposing GitHub profiles, repositories, file trees, content, search, and commit history to LLM agents

*LLM Fine-tuning & Model Training*

**[HuggingFace Fine-tuned Models](#)** 2026

- Fine-tuned multiple open-source LLMs using QLoRA on instruction-following and domain-specific datasets
- TinyLlama-1.1B on Alpaca dataset: `sarimahsan101/TinyLlama-1.1B-Alpaca-QLoRA`
- Qwen 2.5-7B for structured reasoning: `sarimahsan101/qwen2.5-7b-thinking-esp`
- OPT-125M QLoRA baseline: `sarimahsan101/opt-125m-alpaca-qlora`

**AMD AI Academy — [Hands-on Notebooks](#)** 2026

- Completed 6 hands-on courses covering LLM inference, diffusion models, fine-tuning, and RL-based alignment
- Ran notebooks on AMD MI300X GPUs via the ROCm ecosystem — topics include vLLM serving, Unsloth QLoRA, GRPO, MCP-based agents, and HuggingFace on AMD

*Full-Stack Web*

**GCRD Student Portal — [portalersm.com](#)** 2025

- Full-stack student recruitment portal for UK client: role-based auth (Students, Agents, Sub-Agents, Admins), application tracking, document upload, REST APIs
- Stack: React, Node.js, Express, MySQL — deployed live on VPS with Docker and Nginx

**University Student Hub — [unistudenthub.net](#)** 2025

- Call-centre communication platform for UK university students — MERN-style stack, GitHub Actions CI/CD, Docker Compose, live in production

*Mobile Application*

**[Mates Food Delivery App](#) — Flutter, Dart, Firebase** 2024

- Cross-platform mobile food delivery app with real-time order tracking, interactive menus, and user authentication
- Firebase Realtime Database backend for live order status updates and persistent user data

**TECHNICAL SKILLS**

---

<b>AI / ML:</b>	LangChain, LangGraph, RAG, Corrective RAG, QLoRA, LoRA, PyTorch, scikit-learn, HuggingFace, bitsandbytes, Unsloth, vLLM
<b>Backend:</b>	FastAPI, Node.js, Express, REST APIs, MySQL, MongoDB, Qdrant, ChromaDB, FAISS, Pinecone
<b>Frontend:</b>	React, Vite, Tailwind CSS, HTML5, CSS3
<b>Languages:</b>	Python, JavaScript (ES6+), C++, Java
<b>DevOps &amp; Infra:</b>	Docker, Docker Compose, Nginx, GitHub Actions, VPS / Linux
<b>LLM APIs &amp; Tools:</b>	Groq, OpenAI API, Gemini API, Tavily, HuggingFace Hub, ROCm