

Aliyan Zafar

AI / ML Engineer

Github: github.com/aliyanz85 | LinkedIn: [aliyan-zafar-ai-al7y4n](https://www.linkedin.com/in/aliyan-zafar-ai-al7y4n/) | malik.aliyan85@gmail.com | Mobile: +92 304 8522171

Islamabad, Pakistan

Agentic AI and full-stack engineer with production experience building multi-agent systems, MCP servers, and applied computer vision pipelines. Comfortable across FastAPI/Docker backends, LLM orchestration, and independently deployed AI products.

EXPERIENCE

AI Engineer — [Suits Consultants](#) Islamabad, Pakistan

September 2025 – Present

- Led development of Genie, an agentic AI platform automating Shopify store operations, including inventory synchronization and sentiment-aware customer fulfillment.
- Architected multi-agent orchestration to handle high-concurrency workloads for e-commerce clients, coordinating task delegation across autonomous agents.
- Designed and deployed MCP (Model Context Protocol) servers enabling AI agents to securely access local data and external API tools.
- Automated income tax filing and bookkeeping workflows using AI agents, reducing manual data entry and reconciliation effort for clients.
- Built and containerized backend services (FastAPI + Docker) supporting production AI-driven applications.

AI Intern — [Airloop.Ai](#) Islamabad, Pakistan

June 2025 – August 2025

- Engineered a pixel-level pavement crack segmentation pipeline (ConvNeXt-UPerNet encoder-decoder with FPN + Pyramid Pooling) on a 118-image UAV dataset, reaching 96.7% pixel accuracy using a combined Dice-Focal loss to handle severe class imbalance (cracks occupy <3% of image area).
- Implemented a configurable backbone (ConvNeXt Tiny/Small/Base) with mixed-precision training and test-time augmentation to improve inference robustness.
- Built classification logic to determine crack type and severity, producing structured output for structural-assessment reporting.

PROJECTS

- Genie — Agentic E-Commerce Platform: Full-stack AI chatbot automating Shopify order processing and customer interaction via autonomous, sentiment-aware agents. (Proprietary, built at Suits Consultants.)
- A2A + MCP Multi-Agent Communication Framework: Designed a production-style multi-agent architecture (FastAPI microservices) with an asynchronous Agent-to-Agent broker and a centralized MCP context registry; agents share context via pointer-based references with token-based authentication, powered by Groq LLaMA 3.3 70B.
- MultiHopRAG — Hybrid Retrieval-Generation Framework: Built a hybrid RAG system combining FAISS dense vector retrieval with NetworkX knowledge-graph traversal and iterative query refinement for multi-hop question answering; achieved 85% accuracy vs. 35% for a single-step RAG baseline (34% precision improvement).
- VOICE — Bi-Directional Pakistani Sign Language Translation (FYP): Final Year Project: developed and deployed a real-time platform translating between Pakistani Sign Language, text, and speech to reduce communication barriers for the Deaf and Hard-of-Hearing community.
- CrisisSim — Multi-Agent Emergency Response Simulation: Built a research-grade simulation combining Mesa agent-based modeling with five LLM reasoning strategies (ReAct, Reflexion, Plan-Execute, Chain-of-Thought, Tree-of-Thought) to benchmark disaster-response coordination under resource-constrained, multi-agent conditions.
- Superfriend AI — Codebase Visualization Tool: Developer tool for visualizing and interacting with codebases via AI; implemented Celery + Redis for asynchronous task queuing and Stripe for monetization.
- LSTM vs. Transformer — Comparative Deep Learning Study: Implemented Transformer and LSTM architectures from scratch in PyTorch for English-to-Urdu machine translation (evaluated with BLEU/ROUGE), plus a Vision Transformer vs. CNN/ResNet-18 comparison on CIFAR-10.
- AI-Powered Email Triage Assistant: Built an LLM-driven email classification and auto-draft system integrated with the Microsoft Graph API to automate inbox triage and response generation for high-volume mailboxes.
- Deep Learning Applications (CNN + LSTM): CNN-based signature verification and character-level LSTM language modeling on Shakespearean text.
- Osly x N8N: Generated executable n8n workflows from natural-language prompts, covering both frontend and backend integration.
- HelpMeLaw: Legal support chatbot using GPT-4 for conversational logic, deployed on AWS EC2.

- Mortgage Chatbot: Multimodal assistant using LangChain and GPT-4 Vision to analyze financial documents and automate loan eligibility checks.
- Mini Dentist: Diagnostic chatbot using a JSON-based decision tree with Qdrant vector search for intelligent triage responses.
- Git Operations Automation: One-click pipeline automating static website updates, removing manual frontend deployment steps.
- TimeLeft API: Backend APIs for a matchmaking platform using cosine-similarity scoring, built with FastAPI for improved execution speed.

EDUCATION

FAST National University of Computer and Emerging Sciences (FAST-NUCES) Islamabad, Pakistan

Bachelor of Science, Artificial Intelligence Graduated May 2026

- Relevant Coursework: Generative AI, Agentic AI, Artificial Neural Networks, Machine Learning, Computer Vision, Programming for AI, Software Engineering

SKILLS SUMMARY

- AI & Agentic Systems: Agentic AI, Multi-Agent Systems, MCP (Model Context Protocol), LangChain, RAG (FAISS, knowledge graphs), n8n, Hugging Face, OpenAI API, Structured Outputs
- Machine Learning & CV: PyTorch, TensorFlow/Keras, Scikit-learn, Computer Vision (segmentation, classification), NLP, ANNs, Transformers, LSTMs
- Languages & Backend: Python, C++, JavaScript, FastAPI, Django, Flask, Node.js
- DevOps & Tools: Docker, AWS (EC2, S3), Redis, Celery, Git, Jupyter, Kaggle
- Databases: PostgreSQL, MongoDB, MySQL, Supabase, Qdrant (Vector DB)