Muhammad Ali

p230556@pwr.nu.edu.pk | linkedin.com/in/muhammad-ali-ai-dev | github.com/syedali254 | +92 317 0095943

Summary

AI Engineer with production experience shipping agentic systems, RAG pipelines, voice agents, and full-stack products to real users. Merged contributor to **open-design (100k⋆)** and XGBoost **(28k⋆)**. Ran three concurrent internships in Summer 2026 across AI research, SaaS, and university software.

Education

FAST National University of Computer & Emerging Sciences (NUCES)
B.S. in Computer Science — *CGPA: 3.3/4.00 | 2023 – 2027*

Technical Skills

Languages: Python, JavaScript, TypeScript, C++, Dart, SQL

AI & Agents: Multi-Agent Systems, Agentic Workflows, LangGraph, LangChain, RAG Systems, Vector Search (FAISS), Voice Agents (STT/TTS), MCP, LLM Integration, Jev

ML & Deep Learning: PyTorch, XGBoost, Scikit-learn, OpenCV, Computer Vision, Fine-tuning (BioBERT, DenseNet-121), CNN, Vision Transformers (ViT), Grad-CAM

Backend: FastAPI, Django, REST APIs, Celery, Redis, Docker, Kubernetes

Frontend: React, Next.js, TypeScript, Bootstrap

Databases: PostgreSQL, MongoDB, SQLite, MySQL, Supabase, Neo4j

Automation: n8n, Playwright, RPA

Tools: Git, Linux, AWS/GCP , LiveKit

Experience

Research Intern — Decision Support Group, FAST NUCES — Jun 2026 – Oct 2026
Supervisor: Dr. Nouman Azam — *Onsite*
- Researching multimodal pipeline that converts activity video into structured SOPs and executable automations using vision models and LLMs.
- Co-authoring architectures for automated process discovery and task execution.

Software Engineering Intern — NUSyS Lab, FAST NUCES — Jul – Sep 2026
Supervisor: Dr. Omar Usman Khan — *On-site*
- Built Student Society Management Module for NUtomate, a production university-wide automation platform.
- Delivered full-stack features: society creation, member elections, event calendars, finance integration, and public-facing society pages.

AI Engineering Intern — Intellema — Jun – Aug 2026
Creatives Project: Unified creator analytics platform — *Remote*
- Built the AI layer for a platform connecting YouTube, Instagram, LinkedIn, and Facebook metrics into one dashboard.
- Engineered rule-based alert engine detecting metric drops (reach, engagement) and notifying users with root causes before they notice manually.
- Built RAG-based growth advisor backed by a curated per-platform knowledge base of best practices and engagement triggers (FastAPI, Celery, PostgreSQL).
- Added niche-aware content recommendation layer that searches the web and surfaces ideas tailored to each creator's platform and audience.

Open Source Contributions

nexu-io/open-design (100k⋆) | *TypeScript, LLM Engineering, Prompt Safety* — **Merged PR #581** | Link
- Engineered a production LLM skill for structured SOAP clinical notes with PII de-identification safety gates — 13 commits, 20+ review threads.

dmlc/xgboost (28k⋆) | *Python, R, Documentation* — **In Review PR #12196** | Link
- Authored bilingual documentation on custom loss functions, gradient/hessian formulations, and multi-class metrics.

Projects

SeeFlow | *Python, React, Next.js, n8n, LLMs, Computer Vision, MCP* — 2025
- Upload a screen recording and SeeFlow analyzes the actions, generates the workflow logic, and deploys a working n8n automation — no engineering or n8n knowledge required.
- Agentic gap-finder analyzes an SMB's existing workflows, identifies what they should be automating but aren't, and recommends actionable opportunities.
- Targets small businesses replacing manual repetitive work without hiring automation engineers.

RevOps — B2B Lead Engine | *Python, FastAPI, LangGraph, React, TypeScript* — GitHub | 2026
- Sales intelligence platform that discovers real companies from live sources (Hacker News, RemoteOK, Tavily, Exa) and stores every fact with its source URL and timestamp as verifiable proof.

- It then builds a research profile on each company and decision maker, and prepares personalized outreach .
- LangGraph workflow with human-in-the-loop review where uncertain records route to manual approval before any outreach reaches a customer.
- React + TypeScript dashboard with live workflow runs, evidence records, and connector health.

Multi-Agent Clinical Triage | *LangGraph, XGBoost, DenseNet-121, FAISS* GitHub | 2025

- 9-agent pipeline combining radiological screening, vital-sign deterioration detection using Machine Learning, and FAISS-backed RAG clinical reasoning.
- 89.97% triage classification accuracy on 70,000+ patient records vs. 78% inter-nurse agreement baseline.

DecisionOps — Structured Decision Engine | *Python, Jev, Groq* GitHub | 2026

- Built on Jev (typed, structured model by TypeSafe AI, zero hallucination risk): confidence >0.93 auto-proceeds, mid-range escalates to Groq, low-confidence goes to human review.
- Benchmarked 3 pipelines: Jev alone hit 91% vs. Groq's 81% accuracy; hybrid reached 94% calling Groq only 20% of the time — cost reduced from \$4.20 → \$0.38 per 1k messages.
- Full observability layer logs every request with confidence score, route, latency, and model called — live dashboard with version comparison and regression tracking.

Sentix.AI — Product Intelligence | *Python, Django, LangGraph, Gemini, Bootstrap 5* GitHub | 2025

- Multi-agent pipeline that cross-checks reviewer claims in YouTube videos against real user sentiment in comments — produces a weighted sentiment score with Strengths/Weaknesses tags.
- AI-driven alternatives agent maps user pain points to competitor products.
- Exportable PDF executive reports; live data via YouTube Data API v3 and Reddit.

AeroCheck — Easypaisa Fake Screenshot Detector | *OpenCV, XGBoost, Gemini Pro* GitHub | 2025

- 94.2% accuracy detecting fraudulent Easypaisa transaction screenshots using 108 forensic image features — Error Level Analysis, Local Binary Patterns, and DCT frequency analysis.
- Gemini Pro verification layer as a second-pass check on ambiguous cases.

Smart Recruiter | *LangChain, LiveKit, Gemini Live, Django* GitHub | 2025

- AI-powered recruitment platform with voice-based interviews where candidates speak their responses and the system conducts, evaluates, and scores the interview autonomously using STT/TTS via LiveKit and Gemini Live.
- Agentic RAG ranking pipeline ranks candidates against job requirements before routing shortlisted ones to the voice interview stage.