

# USMAN WARIS

AI Backend Engineer

imosmanwaris.tech@gmail.com | +92 320 078 7777 | ISLAMABAD, PAKISTAN

LinkedIn: <https://linkedin.com/in/usman-waris-0a9b8c7d> | GitHub: <https://github.com/usmancynasure>

Portfolio: <https://usman-portfolio-git-main-usmans-projects-d615b8b3.vercel.app>

## SUMMARY

AI Backend Engineer with hands-on experience designing and deploying production-grade AI systems using LangGraph, LangChain, FastAPI, and microservices architecture. Skilled in building RAG pipelines, agentic workflows with tool-based reasoning, and scalable backend services. Proficient in LLM orchestration, real-time system design, and integrating Voice Models in app. Strong foundation in Python, cloud infrastructure (AWS, GCP (Mainly work with GCP), Docker), and Relational Database (Supabase) and Non relational Database (MongoDB). Looking forward to collaborate in Backend and AI Team

## PROFESSIONAL EXPERIENCE

### AI Backend Engineer

GetSnippet (freeZone LLC Dubai-Based, Remote) | 2025 – Present

- Own and ship production AI systems end-to-end, powering automated trading, agentic platforms, and voice assistants used by real users
- Build high-performance async backends (FastAPI) handling real-time streams, concurrent workloads, and latency-sensitive operations
- Design and deploy LLM-driven microservices for multimodal reasoning (text + vision), structured outputs, and autonomous decision pipelines
- Engineer agentic systems with LangGraph, implementing memory, tool orchestration, and human-in-the-loop controls for reliable automation
- Develop event-driven, real-time architectures using SSE/WebSockets for live system feedback and execution tracking
- Architect fault-tolerant microservices, implementing rate limiting, retries, caching, and observability for production reliability
- Build secure, production-grade auth systems (JWT, OTP, encrypted secrets) with abuse protection and multi-tenant user controls
- Design complex decision engines (e.g., trade validation pipelines) with constraint enforcement, risk checks, and external API coordination
- Optimize LLM cost, latency, and accuracy through prompt engineering, selective inference, and pipeline-level filtering
- Integrate and orchestrate external platforms at scale (Twitter/X, broker APIs, health systems, voice APIs) into unified AI workflows

### Backend & Security Contributor

SNSKIES Ltd, Pakistan | Sep 2024

- Implemented Zero Trust Network Access (ZTNA) frameworks to strengthen organizational security posture.
- Contributed to authentication and verification API development ensuring secure identity management.
- Strengthened backend access control through role-based authorization and secure token handling mechanisms.

## PERSONAL PROJECTS

### MedCon AI Systems - ECG Intelligence & Doctor Verification Platform

- Built end-to-end ECG classification system (OpenCV, PCA + KNN) with automated clinician-ready PDF reporting, fully containerized with Docker.
- Engineered real-time Playwright-based PMDC doctor verification with Firestore logging, reducing manual verification by 90%.

## TECHNICAL & SOFT SKILLS

**Programming Languages:** Python, JavaScript, TypeScript

**Backend Frameworks:** FastAPI, Flask, Django, Node.js, RESTful API Design, WebSocket, JWT Authentication

**AI/ML Technologies:** LangChain, LangGraph, Hugging Face Transformers, Scikit-Learn (PCA, KNN), PyTorch, OpenCV, CNN, RAG Pipelines, Agentic Workflows, LLM Orchestration

**Databases & Cloud:** PostgreSQL, MongoDB, MySQL, Redis, Firebase, Supabase, AWS, Google Cloud, Docker, Kubernetes, CI/CD Pipelines

**Tools & Platforms:** Git, GitHub, Playwright, Postman, Linux, VS Code, Claude Code, Cursor

## EDUCATION

Bachelor of Science in Computer Science

COMSATS UNIVERSITY ISLAMABAD | GPA: 3.06/4.00 | Jan 2022 – Dec 2025

## CERTIFICATIONS

- Introduction to Generative AI – Google, Agentic AI Bootcamp – Udemy, Claude Certifications
- AI Engineering Bootcamp: AI Algorithms, Models, and Real-World Projects - Udemy