

AHSAN YOUSAF

+92 307 9429420 ◇ Lahore, Pakistan

ahsan.yousaf1828@gmail.com ◇ linkedin.com/in/ahsanmalick-

ABOUT

Full-Stack AI Engineer specializing in **Generative AI, agentic RAG systems, and SLM fine-tuning**. Proven experience building **production AI agents with MCP-style tool calling**, evaluation pipelines, and scalable LLM services using LangChain, LangGraph, and modern AI stacks.

EXPERIENCE

Full Stack AI Engineer @ Scaylar

Jun 2025 – Present

- Built **production-grade agentic AI systems** using LangChain and LangGraph with reasoning loops, memory, and **MCP-style tool calling**.
- Architected and optimized **RAG pipelines** using advanced chunking, embedding optimization, reranking, and prompt strategies to improve accuracy and latency.
- Designed scalable **LLM orchestration layers** supporting multi-tenant SaaS workloads, tool isolation, fault tolerance, and secure access control.
- Integrated vector databases and external LLM providers to deliver reliable, cost-efficient, and extensible AI-driven backend services.

Associate Software Engineer @ Cyberify

Apr 2024 – May 2025

- Developed and deployed **RAG-based knowledge systems** on **AWS EC2**, processing 10K+ documents with 95% retrieval accuracy using PGVector.
- Implemented **SLM fine-tuning and inference pipelines**, including embeddings, domain adaptation, and automated evaluation workflows.
- Integrated OpenAI GPT-4, ElevenLabs TTS, Stripe, and **tool-invoking AI agents** with robust error handling and observability.

EDUCATION

Bachelor of Computer Engineering

2020 – 2024

COMSATS University Islamabad, Lahore Campus

SKILLS

Generative AI & LLMs: LangChain, LangGraph, Agentic RAG, MCP Tool Calling, Prompt Engineering, LLM Evaluation

SLMs & Training: Fine-Tuning (LoRA, PEFT), Embeddings, Inference Optimization

Vector Databases: PGVector, Supabase, ChromaDB, Qdrant

Backend & Infra: Node.js, Python, PostgreSQL, MongoDB, AWS

Integrations: OpenAI, Replicate, ElevenLabs, Stripe, WebSockets

PROJECTS

Verder (Elunic) – Enterprise RAG & AI Agent Platform

- Built **enterprise-grade RAG pipelines** for internal knowledge search and document intelligence using vector databases.
- Developed autonomous **AI agents with tool calling and reasoning workflows**, improving retrieval accuracy and response latency.

Jack n Jill Health (Portals.care) – Healthcare AI Automation

- Automated clinical and administrative workflows using **AI-driven agents and tool-based orchestration**.
- Implemented intelligent pipelines for data processing and decision support, reducing manual intervention.

INTERESTS

Agentic AI, RAG Systems, MCP Tooling, SLM Fine-Tuning, LLM Evaluation, Intelligent Pipelines, System Architecture