ABRAR AHMAD

AI/ML Engineer | AWS Cloud Engineer | Agentic AI Developer

Lahore, Pakistan | hello@iamabrarahmad.com | iamabrarahmad.com

linkedin.com/in/abrar-ahmad-24045338b | github.com/code-by-abrar

PROFESSIONAL SUMMARY

AI/ML and AWS Cloud Engineer specializing in Generative AI, Agentic AI, Retrieval-Augmented Generation (RAG), Computer Vision, and serverless cloud architecture. Builds and deploys end-to-end systems with Python, FastAPI, LangChain, LangGraph, YOLO, and AWS, covering multi-agent LLM orchestration, real-time object detection, and asynchronous backends.

TECHNICAL SKILLS

Programming Languages: Python, SQL

Generative AI & LLMs: Large Language Models (LLMs), Agentic AI, Multi-Agent Systems, Retrieval-Augmented Generation (RAG), LangChain, LangGraph, Groq, LLaMA 3, OpenAI, Hugging Face, ChromaDB, Prompt Engineering

Machine Learning & Computer Vision: Machine Learning, Deep Learning, PyTorch, TensorFlow, YOLO, OpenCV, Object Detection, Image Segmentation, Natural Language Processing (NLP)

Cloud Computing (AWS): Amazon S3, Amazon CloudFront, AWS Lambda, Serverless Architecture

Backend & Databases: FastAPI, RESTful APIs, Celery, Redis, PostgreSQL, ClickHouse

MLOps & DevOps: Docker, MLflow, CI/CD, Git, GitHub

Core Concepts: System Design, Asynchronous Processing, Performance Optimization

PROJECTS

Serverless Portfolio Website | *Amazon S3, Amazon CloudFront, AWS Lambda* iamabrarahmad.com

- Deployed a serverless portfolio on AWS, serving static assets from Amazon S3 via a CloudFront CDN with HTTPS.
- Implemented backend logic with AWS Lambda, removing server management with a pay-per-use cost model.

AI-RFP-Bidder: Multi-Agent Proposal Generator | *FastAPI, Groq LLaMA 3.3 70B, ChromaDB, Celery* GitHub

- Built a multi-agent AI system that turns Request for Proposal (RFP) documents into technical proposals and cost estimates.
- Built a RAG pipeline on ChromaDB and offloaded long-running LLM tasks to Celery workers behind FastAPI.

Lumina Twin: Enterprise AI Simulator | *LangGraph, Groq, SimPy, Python* GitHub

- Developed a multi-agent simulation platform modeling organizational roles and workflows with LangGraph and Groq.
- Orchestrated agent state machines with SimPy, enabling inter-agent communication and multi-hop reasoning chains.

Clinical RAG System: Healthcare AI Assistant | *LangChain, Groq, Vector Database, Python* GitHub

- Built a medical Q&A assistant with a RAG pipeline over local hospital records using LangChain and Groq.
- Engineered a document ingestion layer with optimized chunking and vector indexing for precise clinical retrieval.

LinkTrim: High-Scale URL Analytics Platform | *FastAPI, Redis, ClickHouse, Celery* GitHub

- Architected a URL shortening and click-analytics backend for high concurrency using FastAPI, Redis, and ClickHouse.
- Designed a write-optimized ClickHouse schema with Celery workers for asynchronous event ingestion.

AI Resume Roaster: ATS Resume Analyzer | *FastAPI, LangChain, Groq LLaMA 3, PDFPlumber* GitHub

- Built an AI resume analyzer and ATS optimizer that detects buzzwords, scores layout, and rewrites weak bullets.
- Parsed PDF resumes with PDFPlumber and delivered LLM-generated feedback through a FastAPI backend.

Computer Vision Detection Suite | *YOLO, OpenCV, PyTorch, Python* GitHub

- Fine-tuned YOLO models for real-time detection of fire, smoke, potholes, helmets, number plates, and plant diseases.
- Built YOLO segmentation for concrete cracks and a low-latency fire alert pipeline with non-blocking inference.

EDUCATION

University of Management and Technology (UMT) Lahore, Pakistan

Bachelor of Science (BSc) in Computer Science

LANGUAGES

English, Urdu