

USAMA MALOOK

AI Engineer · LLM Agents · Model Context Protocol (MCP) · Distributed Systems

usamasam687@gmail.com · linkedin.com/in/usamamalook · github.com/samitiz · usamamalook.com · Lahore, Pakistan (Remote)

SUMMARY

AI engineer with 5 years building production systems, most recently LLM agents and voice AI used by 2,000+ agents a day. Architected an MCP platform that put 3 CRM products behind one AI client, and led a team of 5 engineers. Builds with coding agents as the primary workflow and ships end to end.

SKILLS

Languages: Python, TypeScript, JavaScript, SQL

AI and Agents: LLM agents, Model Context Protocol (MCP), multi-agent systems, Retrieval-Augmented Generation (RAG), LangChain, OpenAI API, Whisper, ElevenLabs, Claude Code

Backend and Data: Node.js, NestJS, FastAPI, REST APIs, WebSockets, RabbitMQ, Redis, PostgreSQL, MySQL, Elasticsearch

Cloud and Frontend: AWS (ECS Fargate, Lambda, RDS, S3, SQS), GCP (Cloud Run, Pub/Sub, Vertex AI), Docker, CI/CD (GitHub Actions), React, Next.js

EXPERIENCE

Full Stack AI Engineer · Hyly.AI

Jun 2026 – Present

Remote

- Shipped a self-updating documentation system in 72 hours, built with Claude Code and Claude Routines, that re-syncs Notion and repositories on every change for each downstream team.

Senior Software Engineer · Dubizzle Labs

Feb 2023 – Jun 2026

Lahore, Pakistan · Led a team of 5 engineers across CRM, real-time messaging and voice AI

- Architected an MCP platform putting 3 CRM products behind one AI client, routing each request through the product's own auth model; trained 50+ engineers on MCP.
- Co-architected an AI voice agent handling 1,000+ outbound calls a day at 95% intent accuracy, streaming over WebSockets across Whisper, LangChain and ElevenLabs.
- Architected a multi-tenant messaging platform that replaced a third-party vendor for 10+ enterprise brands, delivering 5,000+ daily conversations with at-least-once RabbitMQ delivery.
- Scaled real-time infrastructure to 10K+ concurrent WebSocket connections on Redis pub/sub, with event-driven pipelines processing 100K+ messages a day at 99.9% uptime.
- Cut CRM query latency from 800ms to 120ms for 2,000+ agents by tracing a missing composite index, and lifted lead-search relevance 40% with hybrid vector retrieval.

Software Engineer · Logiks Dev

Mar 2021 – Jan 2023

Remote

- Built a document AI pipeline processing 500+ documents a day at 98% extraction accuracy, pairing Tesseract and OpenCV OCR with a TensorFlow CNN that classifies 10+ document types.
- Scaled async processing to 100+ concurrent documents on Redis and Celery workers, and mentored 3 junior developers through code review and pairing.

EDUCATION

BS Information Technology · Punjab University College of Information Technology (PUCIT)

2019 – 2023

GPA 3.56 / 4.0 · Ehsaas merit scholarship (4 years) · 3rd place, Codefest inter-university programming contest · Teaching Assistant for 3 courses