

Aafi Malek

aafimalek2032@gmail.com — +91 7574067294 — LinkedIn — GitHub

Professional Summary

AI Engineer specializing in agentic AI systems, multi-agent orchestration, and real-time voice platforms. Experienced in building tool-using agents capable of multi-step reasoning, workflow execution, and dynamic decision-making in production environments. Strong background in FastAPI-based backend systems, event-driven architectures, and cloud-native deployments on AWS. Hands-on experience with LLM-powered applications, retrieval systems, and real-time communication pipelines, focusing on scalability, latency optimization, and reliable system design.

Education

Bachelor of Engineering in Artificial Intelligence & Machine Learning

L.D. College of Engineering, Ahmedabad

May 2026

Experience

Datumquest

AI Engineer Intern

Jan 2026 – Present

- Built agentic AI systems enabling tool usage, reasoning, and multi-step task execution across real-world workflows.
- Designed multi-agent architectures for task routing, tool invocation, and dynamic interaction pipelines.
- Developed backend systems using FastAPI for agent orchestration, API integrations, and event-driven services.
- Implemented RAG pipelines using vector databases to enable knowledge-grounded responses.
- Integrated real-time communication systems using WebSockets, WebRTC, and webhook-based triggers.
- Deployed scalable AI services on AWS focusing on reliability, latency, and production readiness.

Projects

Prompt to Animation

- Built an agent-driven system to convert natural language prompts into structured animation generation workflows.
- Designed orchestration pipelines for multi-step generation with asynchronous execution and job scheduling.
- Implemented scalable infrastructure using AWS (S3, ALB, ECS Fargate, ECR) for compute-heavy workloads.
- Developed FastAPI backend with Next.js frontend and integrated Redis + MongoDB for job/state management.
- *Tech Stack: FastAPI, Next.js, AWS (S3, ALB, ECS Fargate, ECR), Redis, MongoDB*

Multi-Agent Voice Platform

- Built a multi-agent voice AI platform and dashboard for real-time customer workflow automation.
- Designed agent orchestration with tool calling, decision pipelines, and workflow automation.
- Integrated VAPI and n8n for voice workflows and event-driven execution pipelines.
- Developed FastAPI backend and Next.js dashboard with analytics, session tracking, and control systems.
- Enabled real-time communication using WebSockets, WebRTC, and webhook integrations.
- *Tech Stack: FastAPI, LangFuse, Next.js, n8n, VAPI, Render, Neon, WebSockets, WebRTC*

Advanced AI Chat System

- Built an agent-based AI chat system with memory, retrieval augmented generation, tool usage, and streaming responses.
- Designed orchestration workflows for context handling, retrieval, and dynamic response generation.
- Implemented backend using FastAPI and deployed on AWS (EC2, S3) with CI/CD pipelines.
- Built full-stack app using Next.js with persistent sessions and memory with real-time chat streaming capabilities.
- *Tech Stack: Next.js, FastAPI, AWS (EC2, S3), GitHub Actions CI/CD*

Skills

Agentic AI: LangGraph, CrewAI, MCP, Tool Calling, Agent Orchestration

LLM / GenAI: OpenAI GPT, Claude, Gemini, Llama, Prompt Engineering, Function Calling, Structured Outputs

RAG & Graph: Pinecone, Chroma, Neo4j, vector search, hybrid retrieval

Backend & APIs: Python, FastAPI, Flask, REST APIs, WebSockets, WebRTC, Webhooks

Databases: MongoDB, PostgreSQL, Redis

Cloud & DevOps: AWS (EC2, S3, ECS, Fargate, ECR, ALB), Docker, GitHub Actions, CI/CD