

# AKSHAY GAIKWAD

AI Engineer with 3+ years of experience specializing in Agentic AI systems and multi-agent architectures, with expertise in building and deploying production-grade AI solutions using Python, LangChain, and LangGraph. Experienced in LLM orchestration, RAG pipelines, and prompt engineering, delivering scalable AI applications across enterprise domains with a focus on automation and real-world impact.

✉ akshay.ag544@gmail.com

☎ 919370638163

📍 Pune, Maharashtra, IND

🌐 akshay-gaikwad-3ab619244

👤 AkshayG999

## WORK EXPERIENCE

---

### Nirmitee.io

Software Development Engineer-1

March 2023 - Present

- Architected and deployed autonomous AI agent systems using **Langchain**, **LangGraph** and **Python**, implementing multi-agent, Autonomous Decision-Making workflows that significantly reduced manual intervention in healthcare document processing
- Built production-grade **RAG pipelines** with **LangChain** and **Pinecone**, processing large volumes of medical queries with high accuracy using advanced chunking strategies and semantic routing
- Engineered conversational **AI agents** with advanced NLP capabilities using LLM and custom prompt templates, enhancing customer engagement through context-aware dialogue management
- Implemented MLOps pipelines with Docker, Kubernetes, and **Azure ML** for model deployment, monitoring, and versioning of AI agent systems

### Function Up

Backend Developer Trainee

Jul 2022 - March 2023

- Designed and **deployed** robust APIs using **Node.js** and **MongoDB**, resulting in a 3x increase in system efficiency and a 20% reduction in latency.
- Crafted backend services for BookWise, developing a comprehensive cataloging, searching, and record management system via **RESTful APIs**, which processed 10,000 requests per minute without any issues.
- Conceived and implemented a lightning-fast **URL shortener backend**, leveraging **Node.js** and **MongoDB** to provide instant redirection and link analytics, resulting in a 99.99% uptime and a 50% increase in user engagement.

## SKILLS

---

- **Languages:** Python, JavaScript.
- **Frameworks:** FastAPI, LangChain, LangGraph, CrewAI, OpenAI SDK, Google Gemini, Azure AI, Node.js, React.js, and MCP, RAG, N8N.
- **Cloud & Databases:** AWS, Azure, MongoDB, PostgreSQL, Pinecone, Faiss.
- **Deployment:** Docker, Kubernetes, CI/CD pipelines, Azure ML, LangSmith.
- **Developer Tools:** Git, GitHub, Postman.
- **Agentic AI Systems:** Multi-Agent, Agent Orchestration, Agentic-Tools, Generative AI (GenAI), Large Language Models (LLMs).
- **Prompt Engineering:** Chain-of-Thought (CoT), ReAct Pattern, Few-Shot Learning, Prompt Optimization.

## PROJECT

---

### Sona AI:

Feb 2024

- Architected a multi-agent AI system using **Langchain**, **LangGraph**, and Python for autonomous patient care coordination, orchestrating specialized agents for medical record analysis, symptom assessment, treatment planning, and medication management
- Implemented advanced **RAG** architecture with **LangChain**, **Pinecone**, and **OpenAI**, incorporating medical knowledge and dynamic prompt templates for personalized health recommendations
- Integrated **multiple LLM providers** (OpenAI, Anthropic, Groq, Mistral, Google) for robust multi-model orchestration
- Developed **real-time voice AI agent** using Whisper and streaming **TTS** for low-latency, hands-free interaction
- Developed **medical document processing** using **OCR** and NLP for structured data extraction
- Built scalable backend with **FastAPI**, **PostgreSQL**, and Redis for agent state management, handling high concurrent user sessions with reliable uptime
- Deployed containerized AI services on **Azure** with auto-scaling, reducing patient readmissions through predictive analytics and proactive agent interventions

## Medha AI:

Sep 2023 - Feb 2024

- Engineered autonomous document processing agents using **Python**, **LangChain**, and **Azure AI** for medical record extraction and analysis, achieving high accuracy in entity recognition
- Designed agentic workflows for report generation, orchestrating multi-step reasoning chains for personalized patient insights
- Implemented a hybrid **RAG** system with **Pinecone** vector database and semantic chunking strategies, significantly reducing document processing time while maintaining clinical accuracy
- Built an automated pipeline with Python for continuous model improvement, integrating human-in-the-loop feedback for agent refinement

## Election Police Van's Tracking (Rajasthan Gov.):

Sep 2023 - Oct 2023

- Designed and developed a cutting-edge **real-time police van tracking** system, leveraging **React.js**, **Flutter**, and **Google Maps API** to deliver a comprehensive and efficient solution for monitoring and managing police van activities, resulting in a 30% reduction in response time and a 25% increase in overall efficiency.
- Implemented **geo-location** services, **matrix** route calculations, and **Firebase** integration to enable seamless monitoring and efficient management of police vans during elections, ensuring unparalleled accuracy and reliability, with a 99.9% uptime rate and a 95% reduction in manual data entry errors.

## OTC Agritech:

May 2023 - Aug 2023

- Improved overall **productivity** by 30% and accuracy by 25% through the comprehensive system for **managing farmer** information, **crop stages**, and **purchase history**.
- Facilitated streamlined access and **task execution** for field executives, resulting in enhanced decision-making and improved customer satisfaction by 20%.

## Boongg WhatsApp Bot:

March 2023 - May 2023

- Designed a **WhatsApp-based chatbot** that utilized natural language processing capabilities to **automate motorbike rental processes**, resulting in a 30% reduction in manual labor and a 25% increase in customer satisfaction.
- Developed a WhatsApp chatbot that employed natural language processing to streamline **motorbike rental operations**, processing bookings and receiving a 4.5-star rating from customers for its ease of use and prompt support.

## EDUCATION

---

### Pune University

Bachelor In Computer Science  
**Amrutvahini Polytechnic, Sangamner**

June 2024 - Expected 2026

Diploma(IT) - 77%

## CERTIFICATE

---

Fundamentals of Generative AI [🔗](#)

Create and consume Azure AI services [🔗](#)