PRATIK SONTAKKE

Agentic AI Engineer

CONTACT

+91 8888648824

pratikass488@gmail.com

GitHub

LinkedIn

Portfolio

YouTube

X

IT SKILLS

Agentic AI & GenAI: Multi-Agent Orchestration, LangGraph, LangChain, Model Context Protocol (MCP), RAG, LLM Evaluation (Ragas), Tool Calling

Back-end & Data: Python, Java, FastAPI, Spring Boot, SQL, Vector DBs (pgvector, Chroma), PostgreSQL,Redis

Cloud & DevOps: AWS (ECS, Lambda), Docker, Kubernetes (k8s), Terraform, GitHub Actions, CI/CD

CORE COMPETENCIES

- Autonomous Agent Orchestration
- RAG Architecture
- LLM Evaluation & Optimization
- Back-end Engineering (Python/Java)
- Cloud-Native Development (AWS,Docker, k8s)
- Vector Database Engineering (pgvector, Chroma)
- Microservices & API Design
- Secure Multi-Tenant Architecture
- CI/CD & DevOps Automation
- Cross-Functional Collaboration

EDUCATION

AI Engineering Specialization – *MisogiAI by Masai, Bengaluru*
Jun 2025 – Sept 2025
- Focus: AI/ML, Autonomous Agents, LangGraph, RAG
- Outcome: Designed and implemented **complex, multi-step tool-calling workflows**

Full Stack Backend Specialization – *Masai School, Bengaluru*
Apr 2022 – Apr 2023
- 1,200+ hours of intensive training in **Java, Spring Boot, Data Structures & Algorithms, and System Design**

B.Sc. in Computer Science – *Sant Gadge Baba Amravati University, Maharashtra*
Apr 2017 – Nov 2020

PERSONAL DETAILS

Address	: Bengaluru
Date of Birth	: 18th November 1999
Languages Known	: English, Marathi & Hindi

ABOUT ME

To leverage expertise in **Agentic AI Engineering, autonomous agent development, RAG orchestration, and cloud-native back-end engineering** to build secure, scalable, and high-impact AI systems that drive operational automation and accelerate enterprise decision-making.

PROFILE SUMMARY

- **Agentic AI Engineer with 3+ years of experience** in architecting autonomous multi-agent workflow, RAG pipelines, and enterprise-grade AI systems optimized for reliability, observability, and high-volume real-time inference.
- Proven expertise in building **LangGraph- and LangChain-powered autonomous agents**, LLM tool-calling frameworks, and context-driven decision automation using **MCP** and custom evaluation loops.
- Proven experience in designing **multi-tenant RAG architectures**, vector-search systems (pgvector, Chroma), and latency-optimized retrieval layers engineered for production workloads on AWS & Kubernetes.
- Recognized for delivering **impactful AI automation** such as radiology report validation agents, Text-to-SQL reasoning engines, and autonomous SEO content generation systems that drive measurable business ROI.
- Skilled in implementing **LLM evaluation frameworks** (Ragas, custom metrics) to ensure output faithfulness, reduce hallucinations, and operationalize trust & safety guardrails across AI systems.
- Strong command over **Python + Java/Spring Boot hybrid engineering**, enabling end-to-end delivery of intelligent microservices, high-throughput APIs, and secure back-end integrations.
- Hands-on experience in deploying **scalable cloud-native infrastructure** using Terraform, Docker, AWS ECS/Lambda, and CI/CD pipelines—drastically improving deployment efficiency and operational resilience.

WORK EXPERIENCE

Senior Software Engineer - AI | 5C Network | Sept 2025 – Present | Bengaluru, India
- **Designed and implemented** an autonomous Multi-Agent Supervisor System using LangGraph to audit **5,000+ daily radiology reports**, incorporating **human-in-the-loop workflows** for cases with low confidence.
- **Integrated Ragas evaluation pipelines** to continuously assess agent **accuracy and answer relevancy**, reducing manual QA efforts by **30%** through automated scoring.
- Deployed a **context-aware RAG system for internal technical support**, leveraging hybrid retrieval that combines lexical keyword search (BM25-style) with semantic vector similarity over proprietary medical datasets to improve retrieval precision and reduce latency.

AI & Backend Consultant | Freelance | Nov 2024 – Jun 2025 | Remote
- **Developed and deployed** a Text-to-SQL Agent integrated into a legacy Spring Boot microservice, empowering non-technical teams to query complex databases using **natural language**.
- **Implemented Model Context Protocol (MCP) patterns** to standardize tool-calling interfaces between LLMs and internal APIs, improving **system modularity and maintainability**.
- **Modernized core infrastructure on AWS** with Terraform and containerized microservices, establishing a **CI/CD pipeline** that reduced deployment times to under **10 minutes**.

Software Engineer | Guenstiger | Apr 2023 – Oct 2024 | Delhi, India
- **Designed and implemented an early Generative AI pipeline** leveraging OpenAI APIs to automatically generate hundreds of **SEO-optimized product descriptions**, enhancing **organic search visibility and user engagement**.
- **Optimized high-throughput backend services** using **Java/Spring Boot & Python**, increasing system throughput by **15%** to support user growth without additional infrastructure costs.

Software Engineer | Edifition | Aug 2021 – Apr 2022 | Remote
- Developed and maintained backend modules for client-facing applications, primarily utilizing **Java**, ensuring alignment with service communication standards.
- Enhanced system stability and minimized development rework by rigorously adhering to **API contracts and technical specifications** across all assigned projects.

PROJECT

Enterprise RAG Orchestration Platform
- Architected a scalable **multi-format RAG pipeline** enabling enterprise-grade conversational intelligence with robust ingestion, indexing, and retrieval workflow.
- Designed a secure **Multi-Tenant Architecture** on AWS using schema-per-tenant PostgreSQL and RBAC, ensuring complete context isolation and preventing cross-tenant information leakage.
- Built a **high-performance agent orchestration system using LangGraph and LangChain**, integrated with pgvector-based hybrid vector search to reduce retrieval latency and enable real-time, context-aware agent reasoning.