Sudarshan Dudhe

857-352-9843 — sdudhe92@gmail.com — linkedin.com/in/sudarshandudhe —
github.com/SudarshanDudhe-NEU
Boston, MA

Summary

Senior software developer with 5+ years building scalable SaaS platforms and AI-powered features. Full-stack expertise in Angular/TypeScript and Python/FastAPI, with proven ability to integrate LLMs (OpenAI, LangChain) and design AI validation frameworks for enterprise production systems.

Skills

- **AI/ML:** LLMs (OpenAI GPT-4), LangChain, AI Agents (tool-calling, multi-agent orchestration), RAG, Prompt Engineering, LangSmith observability, NLP, Generative AI
- **Frontend & Web:** Angular, TypeScript, React, AngularJS, JavaScript, responsive web design
- **Backend & APIs:** Python, FastAPI, Django, Flask, REST/JSON APIs, Webhooks, SQL, Bash
- **Data & Databases:** ORM frameworks (SQLAlchemy, JPA/Hibernate), PostgreSQL, SQL Server, MongoDB, Vector DBs (Pinecone, FAISS), Snowflake
- **Cloud & DevOps:** Azure (Blob Storage), AWS (S3), Google Cloud Platform (GCP), Docker, Docker Compose, Git, GitHub Actions, Jenkins, CI/CD pipelines
- **Additional:** Algorithms & Data Structures, Design Patterns, Apache Airflow, MLflow, GitHub Copilot, Performance Monitoring

Professional Experience

AI Engineer Intern   Jan 2025 – May 2025
FounderWay, Boston, MA (Remote)

- Designed and deployed production AI agents with tool-calling and multi-agent orchestration on Azure, integrating OpenAI APIs and Pinecone vector database for a curriculum delivery platform serving 100+ concurrent users with 99.9% uptime
- Built multi-step LLM prompt chains with dynamic templates and versioning; architected parent-child agent workflows for complex, multi-step user interactions with seamless tool integration
- Implemented end-to-end observability using LangSmith, custom Python logging, and error tracking; designed AI validation metrics and human-in-the-loop feedback systems to measure response quality and drive continuous improvement
- Executed systematic prompt testing for AI agents; documented evaluation checklists to ensure production reliability and quality standards

Full Stack Developer   Jan 2021 – Aug 2023
Arges Global Limited, Mumbai, India

- Built responsive full-stack web applications using Angular, TypeScript frontend and Python/FastAPI backend; delivered scalable data management platform serving enterprise customers with high availability requirements
- Owned end-to-end feature delivery from component design through deployment; architected efficient algorithms and data structures, reducing support overhead by 60% through rigorous testing and production-quality code
- Designed and implemented ORM-based data access patterns using SQLAlchemy and JPA/Hibernate; built robust REST APIs and webhook integrations, independently solving complex integration

challenges
- Implemented comprehensive testing, code reviews, and DevOps practices using Git and GitHub; ensured software met high standards for quality, performance, and security

Solutions Engineer Nov 2019 – May 2020

Deqode Solutions, Indore, India

- Developed distributed full-stack web applications using Angular and TypeScript frontends with Node.js/JavaScript backends; collaborated with customers on-site to identify pain points and deliver tailored technical solutions
- Built and enhanced responsive web applications with comprehensive REST/JSON API integrations; implemented production automation features with rigorous testing using Git and modern DevOps tooling

Full-Stack Developer Apr 2018 – Nov 2019

Simplify Reality, Inc, Gurugram, India

- Architected full-stack web applications with Angular/TypeScript frontends and Python backends; integrated AI and automation features for production environments serving global enterprise clients
- Implemented ORM-based data persistence using SQLAlchemy with robust API design and testing; delivered production systems handling complex business logic for healthcare and automation domains

Project Experince

Enterprise AI Systems on Azure Jan 2025 – May 2025

FastAPI, LangChain, RAG, FAISS, Pinecone, Azure Blob Storage, Docker, GitHub Actions CI/CD

- Built enterprise-grade multi-tenant RAG systems with FAISS vector indexing, data isolation, and 30% faster query response times; deployed on Azure with FastAPI backend handling 150+ requests/second and 99.9% uptime
- Designed parent-child multi-agent orchestration architecture enabling complex, multi-step workflows with dynamic tool integration; adopted across 5+ enterprise projects
- Reduced storage costs by 25% through optimized vector indexing and Azure Blob Storage configuration; created comprehensive deployment documentation reducing engineer onboarding by 60%

JobFit AI – LLM-Powered Resume Platform Feb 2024

Python, OpenAI GPT-4, LangChain, RAG, FastAPI, Pinecone, AWS S3, Snowflake, Apache Airflow, Docker

- Developed production LLM application using OpenAI GPT-4 with RAG architecture; integrated Pinecone vector database for semantic retrieval and FastAPI backend with Docker containerization
- Built multi-step prompt chains with dynamic templates, versioning, and parameter tuning; orchestrated end-to-end data pipelines from AWS S3 and Snowflake using Apache Airflow

Education

Northeastern University, Boston, MA Sep 2023 – Aug 2025

Master of Science in Information Systems

- Coursework: AI Agent Infrastructure, Neural Modeling, Data Science Engineering, Big Data Systems

Yeshwantrao Chavan College of Engineering, Nagpur, India Sep 2012 – May 2017

Bachelor of Technology in Information Technology