

Sravan Kumar Yadav Gangi

AI/ML Engineer

gangisravankumaryadav02@gmail.com | LinkedIn | GitHub

Summary

AI/ML Engineer with a demonstrated track record of delivering high-impact AI solutions, specializing in designing and deploying **production-grade AI systems** for enterprise clients, spanning intelligent document automation to multi-agent workflow platforms. Adept at steering initiatives from early proof-of-concept through to full-scale deployment, with a keen focus on **automation, precision, and measurable business impact**. Proficient in modern AI frameworks (**LangChain, LangGraph**) and cloud infrastructure (**Microsoft Azure**).

Experience

AI/ML Engineer

Jul 2024 – Present

Accellor Software Pvt. Ltd., Hyderabad

- Designed and deployed an enterprise-scale **Agentic RAG platform** processing **17K+ enterprise documents**, using **LangChain, Azure OpenAI, PostgreSQL/PGVector**, and multi-agent orchestration.
- Architected an automated **document ingestion pipeline** integrating **SharePoint, ADI, and Blob Storage**, with scheduled **Container App Jobs** for embedding generation and delta/retry-based processing.
- Built and deployed **Python/FastAPI microservices** using **Docker, Azure Container Apps, and GitHub Actions CI/CD**, secured via **Azure Private Endpoints and VNET-based networking**.
- Evaluated the RAG Response using **industry-standard evaluation metrics** for relevance and ranking quality, optimizing inference through **advanced chunking and caching** to cut response time by over **90%**.
- Architected an enterprise **Multi-Agentic AI platform** that converts **RFP (Request for Proposal)** documents into AI-generated proposal drafts, built on **FastAPI** with **Service Bus** job orchestration.
- Built a **Hybrid RAG pipeline** unifying **SharePoint, Salesforce, and internal knowledge bases**, ensuring every compliance requirement is grounded in verified source evidence for accuracy and auditability.
- Designed a **Human-in-the-Loop review workflow** that flags ambiguous requirements, routes them to the right subject-matter expert, and resumes automated draft generation reducing proposal turnaround time.

Personal Projects

DeckPilot | AI-Powered Deck Generator

2026

- Built an end-to-end pipeline that takes **text or PDF** as input and converts it into a fully formatted **PowerPoint presentation**, auto-populating slides using the company's branded template.
- Leveraged **LLM-based content structuring, layout generation, and dynamic slide design** to summarize source material into slide-ready narratives, eliminating manual formatting work.
- Reduced manual deck-creation effort by **80%**, cutting presentation turnaround from hours to a few minutes per deck.

Skills

Languages & Frameworks: Python, SQL, Bash, FastAPI, REST APIs, Machine Learning, NLP

Cloud Platforms: Microsoft Azure, Azure OpenAI, Azure AI Services, Azure Blob Storage, AWS

Databases: PostgreSQL, PGvector, MongoDB, DynamoDB, Vector Databases, Pinecone, FAISS, ChromaDB

AI/ML & Automation: RAG, Multi-Agentic RAG, LLMs, SLMs, LangChain, LangGraph, LangSmith, LangFuse, Agentic AI(Agno, Crew AI), Prompt Engineering, Model Evaluation, Semantic Search, Hybrid Search, Hugging Face, Tokenization

DevOps & Best Practices: Git, GitHub, GitHub Actions, Docker, CI/CD, JWT Authentication, Code Reviews, High Availability System Design, Secure Coding, Async Processing

Certifications

- Databricks Certified Data Engineer Associate (2025)**
- Udacity – Foundations of Generative AI (2025)**



Education

Sastra University, Thanjavur – B.Tech in Mechatronics Engineering | **CGPA: 7.9**

2019 – 2023