

Venkata Sainath Ganta

+91-8978060476 | sainathg1002@gmail.com | portfolio | git-hub | linked-in

PROFILE

Postgraduate AI Engineer with strong hands-on experience in building end-to-end AI applications using Generative AI, machine learning, and computer vision. Skilled in Python, LLM integration, and deploying AI-powered solutions with realworld use cases.

EDUCATION

Madanapalle Institute of technology and sciences

Madanapalle, Andhra Pradesh, India

Post Graduate in Computer Applications

2023 – 2025

- Relevant coursework in [AI, Data Science, ML, etc.]

Sri Venkateswara Degree College

Anantapur, Andhra Pradesh, India

Under Graduate in B.sc (Computer Science)

2019-2023

SKILLS

Technical: Generative AI, Large Language Models (LLMs), Prompt Engineering, Retrieval-Augmented Generation (RAG), NLP, Embeddings, Model Fine-Tuning Basics, API Integration, Model Evaluation, Python Programming.

Tools: Python, Hugging Face, LangChain, OpenAI/Groq/Gemini APIs, FAISS, ChromaDB, FastAPI, Streamlit, Git, GitHub, Docker (Basics), Linux.

PROJECTS

Go Fetch - RAG-Based News Research System ([Link](#))

Tech Stack : Python, FAISS, Sentence Transformers, Groq API, LLaMA-3.1, Streamlit, BeautifulSoup, REST APIs.

- Built a multi-source RAG pipeline ingesting up to 3 news URLs, performing FAISS-based semantic search, and generating source-attributed answers using Groq LLaMA-3.1.
- Implemented intelligent web scraping and overlapping text chunking (500-char chunks) to improve retrieval relevance and minimize context loss in long-form articles.
- Achieved low-latency inference (<1s/query) using normalized 384-dim embeddings with balanced top-K retrieval to prevent single-source dominance.

Coder app - AI Agent-Based Application Generator ([Link](#))

Tech Stack : Python, LangGraph, LLM API(Groq), Pydantic, Prompt Engineering, File System Tooling, Git, Uv .

- Built multi-agent GenAI pipeline (LangGraph + Groq) that auto-generates 6+ complete HTML/CSS/JS apps from prompts with Planner–Architect–Coder orchestration and dependency handling.
- Created FastAPI-powered agent service with rate limiting (5/min), caching, secure workspaces, ZIP downloads, and live previews, achieving reliable 180-second generation for isolated projects.

CERTIFICATIONS

- **IBM-RAG** :Completed Retrieval-Augmented Generation (RAG) certification IBM-Coursera, mastering advanced techniques including vector database integration FAISS, document chunking, hybrid search, reranking, and prompt optimization to build production-grade RAG pipelines with significantly reduced hallucination rates. ([Link](#))
- **Qualcomm-AI Upskilling** :Earned AI certification (Qualcomm Academy), gaining hands-on expertise in large language models (LLMs), prompt engineering, agentic workflows, fine-tuning basics, evaluation metrics, and responsible AI practices to design and deploy intelligent, context-aware AI applications. ([Link](#))