Sivakumar Godugu

6300257875 | gsivakumar7875@gmail.com | GitHub | LinkedIn | Portfolio

Professional Summary:

AI Engineer with 3.9 years of experience, blending deep expertise in Data Warehousing and ETL testing with hands-on development of Generative AI applications. Skilled in designing and deploying scalable, production-ready AI solutions using Python, SQL, GPT-4o, LangChain, LangGraph, Streamlit, and ChromaDB. Specialized in Retrieval-Augmented Generation (RAG), agentic reasoning, and multimodal systems. Proven ability to build AI-powered assistants, automate data workflows, and integrate LLMs like OpenAI and Google Gemini into real-world use cases. Strong background in validating enterprise data systems using tools like Informatica, IBM DataStage, Snowflake, Oracle, and Teradata. Experienced with Microsoft Azure, Docker, and GitHub for cloud deployment and version control. Passionate about building intelligent, user-centric AI systems that drive business outcomes.

Technical Skills:

Programming Languages & Scripting: Python, SQL
Machine Learning & AI: Generative AI, Agentic AI Systems, Large Language Models (LLMs), Prompt Engineering, Fine-tuning, NLP, Deep Learning, Regression Models, Data Mining, Data Analysis, Retrieval-Augmented Generation (RAG), Vector Databases (Chroma)
Frameworks & Libraries: PyTorch, TensorFlow, Transformers, Hugging Face, LangChain (including Agents/LangGraph), Microsoft Autogen, Pandas, NumPy
Generative AI Technologies & Tools: OpenAI, Google Gemini, Groq
Data Engineering & Reporting: ETL Process Automation, Data Warehousing, Data Integration, Data Quality Assurance, Data Cleaning, Microsoft Excel
Databases: MySQL, Snowflake, Oracle, Vertica, Teradata
Cloud & DevOps: Microsoft Azure, Docker, Linux, Git, Github
Application Development: Streamlit, Flask, FastAPI
Workflow & Project Management: Agile Methodologies, JIRA, Tivoli Workload Scheduler (TWS)

Professional Experience:

Generative AI Engineer (Independent / Self-Initiated Projects) **-** Amdocs Development Centre, Pune, India
Jul 2024 – Present

- Designed and developed AI-powered applications using GPT-4o, Google Gemini, LangChain, and Streamlit for real-world use cases such as document-based Q&A, YouTube summarization, and image analysis.
- Engineered Retrieval-Augmented Generation (RAG) workflows with ChromaDB to process and answer complex queries using vector search and LLM integration.
- Built agentic AI systems using LangGraph and Groq LLMs, enabling dynamic tool routing across Arxiv, Wikipedia, and the web.
- Integrated APIs for vision, document parsing, and external knowledge retrieval to create robust multimodal applications.
- Deployed production-grade GenAI apps with modular architecture, secure .env-based API key handling, and responsive Streamlit UIs.

- Focused on building scalable, user-centric, and secure AI systems aligned with best practices in prompt engineering and autonomous reasoning.

DWH/BI Test Engineer – Amdocs Development Centre, Pune, India
Oct 2021 – Jun 2024

- Led end-to-end validation of ETL pipelines and BI reporting solutions in the telecom domain for global client AT&T.
- Created complex SQL queries and data validations across Snowflake, Oracle, Vertica, and IBM DataStage environments.
- Automated test workflows to reduce manual effort and improve test coverage using Python and shell scripting.
- Performed regression, functional, and UAT testing for multiple releases in Agile delivery cycles.
- Collaborated with cross-functional teams including data engineers, analysts, and project managers to ensure timely and accurate delivery.
- Ensured compliance with enterprise data governance and quality standards while maintaining detailed test documentation and progress tracking.

Key Projects:

Generative AI Projects:

1. **NexaAI – All-in-One AI Assistant App (Git)**

   Technologies: Python, Streamlit, Google Gemini API, Hugging Face (Stable Diffusion), YouTube Transcript API, Pillow

   Developed a versatile AI assistant integrating multiple advanced capabilities in a single Streamlit interface. Powered by **Google Gemini**, **Stable Diffusion**, and **custom Python utilities**, NexaAI enables users to chat with an LLM, summarize YouTube videos, generate AI images from prompts, and analyze uploaded images with visual reasoning.

   Key Contributions:

   - Built a unified Streamlit UI for chat, video, and image tasks with real-time interactivity.
   - Integrated **Gemini 1.5 Pro** for multilingual chat and vision-based image captioning.
   - Implemented **YouTube summarizer** pipeline with transcript extraction and LLM refinement.
   - Connected Hugging Face's **Stable Diffusion API** for image generation from text.
   - Managed .env-based API key handling for secure deployment on **Streamlit Cloud**.

2. **Agentic RAG Chatbot (Git)**

   Technologies: Python, Streamlit, OpenAI GPT-4o, LangChain, ChromaDB, PyPDF2, python-docx

   Developed an advanced Retrieval-Augmented Generation (RAG) chatbot that leverages OpenAI's GPT-4o and LangChain to process and answer complex user queries based on uploaded documents. Implemented agentic behavior by decomposing queries into subtasks, retrieving relevant context using ChromaDB embeddings, and dynamically constructing context-aware answers. Built an interactive Streamlit UI supporting multi-format document upload (.pdf, .docx, .txt, .json) with persistent chat history and transparent result tracing.

   Key Contributions:

   - Designed and implemented multi-document ingestion and chunk-based indexing pipeline.
   - Integrated OpenAI embeddings (text-embedding-3-large) and ChromaDB for semantic search.
   - Enabled agentic reasoning by combining LangChain tool chains with custom prompt logic.
   - Developed a responsive Streamlit UI with real-time query handling and subtask breakdowns.

3. **Smart Chain Agents – Tool-Routing AI Assistant (Git)**

Technologies: Python, LangChain, LangGraph, Groq LLM (qwen-qwq-32b), Streamlit, Arxiv API, Wikipedia API, Tavily(for internet search)

Built an intelligent AI agent that dynamically routes user queries to the most appropriate source — Arxiv, Wikipedia, or the Web — using a **LangGraph-powered decision graph**. Integrated tool-augmented reasoning with **Groq's high-speed LLM** and external APIs to deliver contextual, source-specific answers. Developed an interactive Streamlit interface and designed a modular agent architecture for flexible tool invocation and real-time response generation.

Key Contributions:

- Implemented agentic logic using LangGraph for state-based, conditional tool selection.
- Integrated Arxiv, Wikipedia, and Tavily search tools for accurate knowledge routing.
- Built and deployed a responsive UI using Streamlit for seamless user interaction.
- Combined LLM reasoning with external data retrieval to enhance relevance and context.

DWH/BI Projects:

1. Modernization of Aux Credit & Promotion Process (Client- AT&T)

Summary: Spearheaded the modernization of credit and promotional processes for mobile devices, developing scalable datasets and self-service reporting capabilities that replaced ad-hoc processes, boosting operational efficiency.

Key Contributions:

- Designed data ingestion pipelines with ADLS Gen2 and Snowflake for seamless promotional data integration.
- Automated workflows using custom engines (Promo Crediting, Compliance, Reward Card) to improve efficiency.
- Developed attribute rule sets for enrollment, eligibility, and crediting, ensuring smooth operations.
- Collaborated with vendors to ensure accurate trade-in and reward card reconciliations.
- Deployed production-grade solutions, reducing turnaround times and meeting data governance standards.
- Enabled real-time reporting insights through MDE data templates for informed decision-making.

Technologies/Tools: Snowflake, Python, ADLS Gen2, DIY Ni-Fi, SharePoint, Excel Power Query, Promotion Device Manager.

2. Data Transformation - Migration from Teradata to Snowflake (Client- AT&T)

Summary: Led the end-to-end migration of 500 tables from Teradata to Snowflake, delivering the project with a 95% success rate. As a key individual contributor, I ensured smooth deployment and optimized data flow, enhancing system performance and reliability.

3. CIMCDR (Client- AT&T)

Summary: Contributed to the Customer Information Management project by testing and validating a data standardization and matching mechanism, ensuring accurate assignment of unique enterprise addresses and customer keys to data from multiple sources (e.g., tables, files, DB2). This process enhanced data consistency and integration across systems.

Technologies/Tools Used: IBM DataStage, Pre-processors, Post-processors, Oracle DB.

Enterprise Data Warehouse Testing

- Led system-wide ETL and database testing for a large-scale data warehouse implementation.
- Created reusable SQL-based test cases, collaborated with developers for defect resolution, and maintained test documentation in Agile sprints.

Education:

Bachelor of Technology in Computer Science and Engineering (CSE) - CGPA 7.5

Sir CV Raman Institute of Technology and Science, Tadipatri, Andhra Pradesh | 2021

Extracurricular Activities:

- Actively participated in internal corporate sports events and contributed to team success by winning multiple contests, demonstrating teamwork, leadership, and a competitive spirit.