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

Professional Summary:

Results-driven **AI enthusiast** and **Data Professional** with 3.7 years of experience in **Generative AI development** and **DWH/BI testing**. Built and deployed AI applications using **GPT-4o**, **LangChain** and **Streamlit**, focusing on **RAG workflows**, **intelligent agents**, and **multi-modal interfaces**. Previously worked as a **DWH/BI Test Engineer** for 2.7 years, validating **ETL pipelines**, **data integration**, and **BI systems** using tools like **Oracle**, **Snowflake**, **Vertica**, **Informatica**, and **IBM DataStage**. Strong in **SQL testing**, **automation**, and working within **Agile environments**. Skilled in **Python**, **SQL**, **Microsoft Azure**, **Linux**, and **open-source ML/LLM frameworks**. Known for delivering **scalable**, **automation-driven solutions** that bridge data integrity with modern AI capabilities.

SKILLS:

Programming & Query Languages: Python, SQL
AI/ML & Data Science: Generative AI, LLMs, NLP, Deep Learning, Regression Models, RAG, Vector Databases, Pandas, NumPy, Data Mining, Data Cleaning, Data Analysis
Frameworks & Libraries: LangChain, LangGraph, Transformers, Flask, Streamlit, FastAPI
Cloud & Infrastructure: Microsoft Azure, Linux, Git, Docker
Databases: MySQL, Snowflake, Oracle, Vertica, Teradata
Data Warehousing & ETL Tools: Informatica, IBM DataStage, Infoworks
Data Reporting: Microsoft Excel
Project & Workflow Management: JIRA, TWS (Tivoli Workload Scheduler)

WORK EXPERIENCE:

Generative AI Enthusiast (Independent Projects)- Amdocs Development Centre, Pune, India
Duration: 1 year
- Built AI-powered applications using GPT-4o, Gemini, LangChain, and Streamlit
- Engineered workflows supporting document-based RAG search, image generation, and YouTube summarization
- Designed agent routing systems for knowledge retrieval via web, Arxiv, and Wikipedia
- Deployed real-world solutions with multi-modal capabilities, modular architecture, and secure API key management
- Focused on building scalable, user-friendly, and production-ready Gen AI apps.

DWH/BI Test Engineer – Amdocs Development Centre, Pune, India
Duration: 2.7 years
- Validated ETL workflows and BI reporting systems in telecom domain for global client AT&T
- Performed end-to-end testing of data ingestion, transformation, and warehouse integration
- Created test cases and SQL queries to verify data accuracy across Snowflake, Oracle, and DataStage systems
- Led QA activities for multiple releases including regression, functional, and UAT testing
- Automated test workflows to reduce manual validation efforts and increase test coverage
- Collaborated with business analysts, developers, and project managers in Agile environments
- Ensured data integrity, compliance, and consistent delivery of enterprise-grade analytics platforms

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

   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.

- **Other- projects: Data Warehouse Testing**
  Led ETL and database testing for a large-scale data warehouse project, ensuring data accuracy and performance. Collaborated with cross-functional teams to define test strategies and resolve issues. Created and maintained test cases for continuous data validation, and provided clear documentation and progress updates to stakeholders

---

EDUCATION:

Bachelor of Technology in Computer Science and Engineering (CSE)
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.