Varun G

GenAI Developer

Bengaluru +91-9731416185 varungangu1@gmail.com LinkedIn Medium

Summary

GenAI Developer specializing in enterprise AI solutions with 4+ years of Python development experience. Proven track record of reducing operational time by 95% through AI-powered automation (GRD project). Expert in building RAG-based chatbots, document processing pipelines, and data extraction systems using Langchain, Langgraph, and GPT-4o. Awarded for innovative AI solutions at MITC Hackathon 2024.

Experience

Merck Group **2024-Present**
GenAI developer Bengaluru

- Proficient in developing AI-powered applications with a focus on Generative AI.
- Expertise in building conversational AI systems, document processing pipelines, and data extraction solutions.
- Hands-on experience with Langchain, Langgraph and vector databases like Qdrant for efficient document retrieval and multi-modal data access.
- Proven ability to optimize complex processes, significantly improving operational efficiency and reducing task completion times.
- Award-winning innovator, recognized for delivering cutting-edge AI solutions in hackathons and internal competitions.

Merck Group **2021-2024**
Python developer Bengaluru

- Engineered and implemented a comprehensive end-to-end data transformation pipeline, revolutionising the company's data processing capabilities.
- Developed robust Python scripts to extract and parse complex, multi-sheet Excel files, handling various data formats and inconsistencies
- Implemented data cleaning and validation procedures, including handling of missing values and data type standardization, ensuring data integrity
- Integrated automated data quality checks and error logging mechanisms, significantly reducing manual intervention and improving data reliabilit
- Enabled the creation of interactive Tableau dashboards and reports, empowering stakeholders with real-time, data-driven insights for strategic decision-making

Merck Group **2019-2021**
Application support engineer Bengaluru

- Started professional career as an Application Support Engineer, demonstrating strong technical foundation and problem-solving skills
- Provided comprehensive support for both GxP (Good Practice) and Non-GxP applications, ensuring regulatory compliance and operational efficiency
- Developed expertise in critical process workflows, including Request for Change (RFC) and change management procedures, contributing to streamlined operations and risk mitigation

Projects

GRD (Global Response Document) **2024-Present**

GenAI Summarisation App

Developed an AI-powered application to generate Global Response Documents (GRDs) for healthcare professionals, reducing the GRD creation time from 3–4 hours to just 10–20 minutes. The solution included a document preprocessing pipeline using PyMuPDF for text extraction and AWS Textract for table extraction. Tailored system prompts were created based on user-selected Therapeutic Area (TA), product, and GRD type, while a batch processing system enabled efficient multi-page document handling. Using a single-shot prompting approach, the application maintained a scientific tone and format. Integrated with the GPT-4o model via Langchain for both data extraction and summarization, the app featured a Streamlit-based user interface for easy interaction and document uploading. It also provided a citation system for page-level references and generated output in downloadable Word format for seamless distribution and editing.

Python, Langchain, Streamlit, PyMuPDF, AWS Textract, GPT-4o

BrAIn (Breeze AI Navigator) **2023-2024**

GenAI Chatbot

Developed an AI chatbot with access to Breeze SOPs and training documents, supported by an automated data ingestion pipeline using AWS Lambda, S3, and SharePoint. Leveraging Azure OpenAI's text-embedding-3-large model and Qdrant DB, we built a vector store for efficient document retrieval. The chatbot’s interface was created using Streamlit with a Python and Langchain backend, incorporating multi-vector retrieval for parent-child document relationships and context-aware conversations using router and question reframe chains. Powered by the GPT-4o model, the chatbot delivers accurate responses with a citation system providing page-level references from source documents.

Python, Azure OpenAI, Qdrant, Streamlit, Langchain, SharePoint, AWS, Lambda, S3

PACO (Protocol Analysis for Complexity Optimisation) **2023-2024**

GenAI Data Extraction App

Developed an AI-powered application to extract structured data from unstructured Word and PDF documents, featuring a document processing pipeline using PyMuPDF and the Unstructured library. The application included a user-friendly Streamlit interface for document uploading and result downloading. Custom system prompts were engineered for various extraction tasks, including JSON schema specifications, and Langchain's structured output function ensured consistent JSON-formatted results. Powered by the GPT-4o model for high-accuracy information extraction, the app implemented a batch processing system to parallelize tasks and improve efficiency. Additionally, we developed a conversion process to transform extracted JSON data into a downloadable Excel format for easy user review.

Python, Langchain, Streamlit, PyMuPDF, Unstructured, GPT-4o, JSON, Excel

MiNE MCP **Aug-2025 - Present**

MCP setup for MiNE (exploration)

Takenup the MCP server setup for MiNE project as exploration/POC task.

MCP, LLM Agents, AWS

Projects

Text to SQL Conversational App **2023-2024**

POC

Developed a proof-of-concept application enabling natural language queries on Foundry data tables, featuring a conversational AI system using the GPT-4o model and Langgraph for agentic workflows. A SQL query agent was designed to convert natural language inputs into accurate SQL queries, supported by a custom tool to execute these queries on Foundry tables and retrieve relevant data. An answer agent analyzed query results and generated human-readable responses. The application included a Streamlit-based interface for seamless user interaction, integrating table schema information and custom system prompts to improve query accuracy and enhance data accessibility through natural language interaction.

Python, Langgraph, Streamlit, GPT-4o, SQL, Foundry

MiNE **Oct 2025 - Present**

Report generation tool

Built an AI-powered tool that automatically creates reports by gathering information from multiple sources (social media, scientific articles, and medical data). Designed a two-step process: first, the system extracts key information from user questions to quickly filter through millions of database records; second, it uses smart retrieval methods to find the most relevant data and generate accurate reports.

Foundry, Qdrant, Langchain, Streamlit, RAG

Education

UAS, Dharwad **2014 - 2016**
Agriculture B.Sc.
7.25

G.K.V.K Bengaluru **2016-2018**
Agriculture Statistics M.Sc.
8.2

Skills

Programming language
●●●●○
Python

Generative AI
●●●●○
Large Language Models (LLM), Prompt engineering, Langchain, Langraph

Cloud & DevOps
●●○○○
AWS, Git, Docker

Web Development
●●●○○
Streamlit

Project Management
●●●●○
Agile methodologies

Awards

Be Curious and Innovate Boldly Award **2024**

Runner-up

MITC Hackathon **2024**

Runner-up

Languages

Kannada
●●●●●

English
●●●●●

Hindi
●●○○○