

PROFILE

AAKASH YADAV

 7506444919

 aakash27.2000@gmail.com  Mumbai

SUMMARY

Innovative and results-driven AI Engineer with around 4 years of experience in developing and deploying machine learning models, particularly in natural language processing and large language models (LLMs). Proven track record of improving operational efficiency and predictive accuracy through advanced AI solutions. Adept at coordinating with stakeholders to deliver customized AI solutions and continuously advancing skills in the latest AI technologies.

SKILLS

LLM Models - Claude 3.5 Sonnet, FLAN-T5, LLAMA 2, Mistral, Chat-GPT-3.5, Chat-GPT-4

Natural Language Processing (NLP): NLTK, SpaCy, Text Analysis, Sentiment Analysis, LLM, Langchain, Huggingface, Gen AI, Prompt Engineering, LLM Fine tuning, LangGraph Workflow, Agentic Workflow and Agent2Agent Protocol.

Data Science: Pandas, NumPy, Data Visualization (Matplotlib, Seaborn), Statistical Analysis

Programming Languages: Python

Cloud Platforms: AWS (Sagemaker, Textract, Kendra, Lambda, Glue), Azure(Functions, OpenAI, Cognitive services, LU portal)

Machine Learning & Deep Learning: PyTorch, Scikit-learn

Web-Framework: Flask

Model Deployment: Serverless Architecture, Docker

Version Control & Collaboration Tools: Git, Jira

AutoML Data Science Platform: Dataiku

Coding Tools: Github Copilot, Claude Code and Cursor.

Other Tools: Selenium, UI Path, Swagger API

Professional Experience

AI Consultant

Dec 2024 – Present

Deloitte US

Client -Pfizer

Knowledge Mining Platform (Chatbot + Data as Service Platform)

Role and Responsibilities:

- Designed and developed a unified platform to deliver data from multiple clinical data sources.
- Built an MCP Server and Agent to consume live data streams and implemented a vector database to store static/unlive data for efficient retrieval.
- Developed an orchestrator agent capable of connecting to and querying across diverse data sources, enabling seamless information access for users.
- Deployed the orchestrator agent as a FastAPI service on AWS for scalable and secure consumption.

Tech Stack Used:

- LangGraph Workflow, React Agent, LLM Model (Open AI and Anthropic)

ACCT Chatbot (Text-to-SQL chatbot)

Role & Responsibilities:

- Designed and developed a Text-to-SQL chatbot capable of converting natural language queries into SQL and retrieving insights from the AACT clinical trials database.
- Built an MCP Server and Agent to enable secure, real-time connectivity with the AACT database and streamline data access.
- Created an intelligent agent layer on top of the MCP setup to interpret user queries, generate optimized SQL, and return accurate results.
- Integrated the chatbot into a scalable application workflow, ensuring smooth interaction between the LLM, MCP server, and the AACT database.

Tech Stack Used:

- LangGraph Workflow, React Agent, LLM Model (Open AI and Anthropic), MCP Server

AI & ML Engineer

Feb 2021 – Dec 2024

Company Name-i2e Consulting

Client -Pfizer, Alexion

Clinical Trials Chatbot (LLM Bot)

Role and Responsibilities:

- Designed and deployed an end-to-end chatbot to assist clinical researchers by integrating unstructured data sources, significantly reducing manual query resolution time.
- Leveraged Retrieval-Augmented Generation (RAG) technique, powered by AWS, to enhance query handling and

information retrieval.

- Implemented a CI/CD pipeline using Terraform scripts to automate deployment and updates.

Tech Stack Used:

- AWS Kendra, Textract, SageMaker, FastAPI Framework, LLM Models (Flan T5 XL, Llama 2 Chat Completion), Terraform.

Support Chatbot

Role and Responsibilities:

- Served as an AI Architect, collaborating with the development team during the initial stages to design the chatbot's technical architecture.
- Addressed the challenge of accessing infrastructure data from a large database by creating a chatbot supported by a knowledge base.
- Provided users with seamless information access, eliminating the need for database expertise.

Tech Stack Used:

- LangChain, Vector DB, Flask Framework, LLM Models (Claude 3 Sonnet, GPT-4).

Predictive Quality Event Model

Role and Responsibilities:

- Developed a predictive binary classification model on historical data to identify protocols likely to encounter significant quality events (SQE).
- Achieved a 30% increase in protocol design success rates by flagging potential quality issues.
- Built an ML-powered alert system to mitigate the impact of global events on clinical trials.

Tech Stack Used:

- Python, Scikit-learn, Pandas, NumPy, ML Deployment Tools.

ADDITIONAL RESPONSIBILITY

- Mentored and guided junior developers by conducting weekly code reviews, providing constructive feedback, and offering tailored learning resources.
- Involved in various Pre-sale activities and POC delivery.

EDUCATION

- Master of Science in Data Science and Artificial Intelligence | Mumbai University
- Bachelor of Mathematics | Mumbai University

AWARDS

- Employee of the Month -July 2021
- Employee of the Month -Nov 2022

IMPORTANT LINKS

- [GITHUB](#)
- [LINKEDIN](#)

SOFT SKILLS

- Problem-Solving: Complex AI-related Problem Resolution
- Analytical Thinking: Data Interpretation, Decision Making
- Communication Skills: Technical Concepts for Non-Technical Audiences
- Teamwork & Collaboration: Interdisciplinary Team Projects