Akash Kumar

AI/ML Developer
+91-7010429267 | akash.kumar2113196@gmail.com

Education

Vel Tech University, Chennai, Tamil Nadu July 2018 – May 2022
Bachelors in Computer Science Engineering GPA: 8.03

KV AFS AVADI, Chennai, Tamil Nadu April 2017 – March 2018
XII CBSE 75%

Skills

Languages: Python, JavaScript, React Native, TypeScript, C, C++, HTML/CSS, Bash, XML, GraphQL
Databases: MySQL, SQLAlchemy, PostgreSQL, Argus, MongoDB
Libraries: Langchain, NumPy, Pandas, OpenCV, Transformers, PyTest, NLP
Frameworks: Taskweaver, FastAPI, PyTorch, Django, Flask, Node.js, Express,js, Bootstrap
Tools & Technologies: AWS, Azure, Large Language Model(LLM), Git, Docker, Heroku, Vercel, Apache Spark, Gitlab CI/CD, Kubernetes
Soft Skills: Problem-solving, Communication, Leadership, Adaptability

---

Experience (1 yrs 10 mos)

Cognizant, Software Developer (1 yrs 10 mos) Oct 2022 – July 2024

-gAICAW (Generative AI Content Authoring Workbench)

- Developed a PSUR Document Generation system using Gen AI, automating 95% of document sections, reducing manual efforts of medical writers by 90%.
- Extracted data from different sources i.e. word documents and excel files and put them together at desired places in final PSUR document.
- Executed cutting-edge Chain of Thoughts prompting techniques to extract structured data from Database systems resulting in a 40% decrease in data retrieval time and a 25% increase in data accuracy.
- Impacted hundreds of medical writers and content authors to focus on working on creative things.
- Implemented features to detect Adverse event, drugs interactions, safety issues by mining online medical articles about drug available in Pubmed/Embase significantly reducing data extraction time by 70%.
  Tools: Azure, AWS, Postgre, Python, gpt-4 turbo, Taskweaver, Angular, Fastapi

-CodeDocGen: Automated Documentation for Data Transformation

- Engineered and deployed a sophisticated LLM (Large Language Model) application leveraging Langchain and the Huggingface Mistral 7B instruct model on Azure.
- This innovative solution automates the generation of comprehensive documentation for dbt and PySpark files directly from GitHub repository with accuracy of 90%.
- Documented file can be downloaded as text file which can be further used as readme file for repositories, enhancing project documentation and accessibility by great extent.
  Tools: Python, Mistral, Langchain, Pinecone, Nextjs, Azure, Vercel

-Internship (4 mos) Jan 2022 – April 2022

- Worked with Apache Hadoop ecosystem tools, optimizing query workflows using PySpark. which led to a 30% improvement in query execution time and a 20% reduction in resource consumption for processing large-scale datasets
  Tools: Hadoop, Hive, Scala, Azure Databricks, PySpark

CipherSchools, Python Teacher Assisstant (2 mos) April 2021 – June 2021

-EdTech platform

- Managed EdTech platform having a batch of 30 students under me, creating educational content, handling assignment evaluations, and addressing community queries.
  Tools: Python, Excel, PPT

Projects

-LLM Question Answering App

- Engineered a web application using Streamlit for uploading and processing text files (docx, pdf, txt).
- The app converts text to embeddings and stores them in ChromaDB, enabling users to query documents for specific information with an average response time of 200 milliseconds.
  Tools: Python, Langchain, LLM, ChromaDB, Streamlit, Azure OpenAI

-Image Classification-Cats and Dogs

- Created a TensorFlow serving model for classifying images of cats and dogs, achieving an accuracy of 97% on the training set.
- Additionally, built a backend for user management using SQLAlchemy, supporting up to 1000 concurrent users and allowing them to upload images and receive immediate results.
  Tools: Python, Flask, SQLAlchemy, Docker, TensorFlow, HTML/CSS

-Student Performance Tracker (UG Major Project)

- Conducted data extraction and visualization of a student dataset with 31 attributes, aiming to understand and predict student exam performance.
- Utilized KNN and Logistic Regression algorithms to predict exam outcomes and identified key factors influencing student performance through data analysis.

  Tools: Python, Jupyter Notebook, Kaggle, Numpy, Pandas, Scikit-learn, matplotlib

-ReadHub (Library Management System)

- Designed a library management system with features for user authentication, book search, and issue management supporting capable of accommodating up to 500 simultaneous users.
- The system supports contacting administrators for queries and was deployed using Heroku and PostgreSQL.

  Tools: Python, Bootstrap, Flask, SQLAlchemy, PostgreSQL, HTML/CSS