Swathika .K

Pleasanton, CA · Swathika.kalyan0488@gmail.com · (925) 579-1278 ·
https://www.linkedin.com/in/swathika-kalyanasundaram-2461a2346/

Education

Madras University Chennai,India
Master of Science in Information Technology

Experience

Lucid Motors Newark, CA
Senior Machine Learning Engineer Apr 2022 - Present

- Created AI assistants leveraging Generative AI (GenAI) using Large Language Models (LLMs) to deliver personalized and context-aware responses.
- Leveraged Retrieval-Augmented Generation (RAG) techniques and LangChain to integrate multiple knowledge sources, optimizing the efficiency of information retrieval and enhancing response quality.
- Developed and implemented evaluation pipelines to assess the effectiveness of RAG workflows, focusing on the accuracy, relevance, and speed of information retrieval and response generation.
- Used Hugging Face Transformers to connect models and FAISS to quickly and accurately retrieve information in RAG systems.
- Utilized LangChain for advanced conversational capabilities and workflow automation, enhancing the chatbot's ability to manage complex dialogues and tasks.
- Utilized Vector Databases to store and retrieve embeddings efficiently, facilitating quick and accurate similarity searches for improved user query handling.
- Developed and deployed scalable embedding pipelines and similarity search systems using Hugging Face, BERT, and Scikit-learn, enabling efficient text representation, real-time content retrieval, and seamless integration into large-scale workflows.
- Developed a multi-agent system using LangGraph with Agentic AI, to collaborate for complex task automation, improving decision-making and optimizing workflows.
- Developed FastAPI's for deploying machine learning models, ensuring low latency and high availability in production environments.
- Designed and implemented prompt engineering architectures for various applications, creating scalable and reusable templates such as zero-shot, few-shot and instruction-based prompts, while iteratively refining them to ensure consistent and high-quality outcomes
- Fine-tuned LLMs using domain-specific datasets to optimize the chatbot's performance, ensuring high accuracy and relevance in conversations
- Deployed the chatbot using AWS Bedrock for seamless integration with foundation models, enabling scalable model inference and customization for diverse use cases.
- Utilized AWS ECR for container image storage and AWS EKS for scalable deployment, ensuring efficient model serving and high availability across various environments.
- Developed and executed robust model evaluation protocols, including A/B testing and user feedback integration, to continuously refine and improve the chatbot's performance
- Utilized metrics like BLEU, ROUGE, perplexity, and BERT Score to assess both retrieval accuracy and the quality of generated text, ensuring contextually relevant and high-quality output tailored to diverse use cases.

Talent Screen Dublin, CA
Machine Learning Engineer May 2020 - Apr 2022

- Developed and deployed adaptive machine learning models using PyTorch, implementing real-time classification systems with dynamic difficulty adjustment and ensuring scalability in evaluation workflows.
- Designed and implemented machine learning pipelines, automating model training, evaluation, and deployment, using Scikit-learn for seamless integration and scalability.
- Utilized distributed training frameworks to accelerate the training of large-scale deep learning models, ensuring efficiency..

- Evaluated model performance rigorously, employing metrics such as precision, recall, F1-score, and mean squared error, and applying cross-validation for robustness and reliability.
- Deployed machine learning models in production environments using Docker, Kubernetes, and TorchServe, ensuring scalable and efficient model serving.
- Implemented monitoring and logging solutions, leveraging Prometheus to track system performance and troubleshoot machine learning workflows effectively.
- Built and optimized recommendation systems using collaborative filtering, content-based filtering, and hybrid approaches to personalize user experiences and improve interaction efficiency.
- Utilized AWS SageMaker to streamline the end-to-end machine learning lifecycle, from model training to deployment and monitoring, enabling scalable, cost-efficient, and automated workflows for model development and inference.
- Implemented MLOps practices to automate and streamline the deployment, monitoring, and management of machine learning models, ensuring continuous integration and delivery (CI/CD), model versioning, and performance tracking in production environments.

United HealthCare Services, Inc Elgin, IL
Senior Data Engineer Dec 2018 - Apr 2020

- Designed and implemented data ingestion pipelines using Python ensuring efficient extraction, transformation, and loading (ETL) of data from diverse sources.
- Built and managed real-time data streaming applications with Python using Apache Kafka and Apache Flink enabling real-time data processing and analysis for timely insights.
- Optimized data storage solutions using Python with PostgreSQL, balancing structured and unstructured data storage requirements for high performance and scalability.
- Developed real-time data streaming solutions with Apache Kafka enabling timely insights and decision-making.
- Utilized scalable data warehouses like Google BigQuery to store, query, and analyze large datasets.
- Processed and analyzed large-scale datasets using Apache Spark and Hadoop, achieving efficiency improvements in data workflows.
- Automated and orchestrated data workflows using Apache Airflow, ensuring consistent and reliable data availability.
- Deployed containerized data applications using Docker and Kubernetes, enhancing the scalability and reliability of analytical environments.

E-Trade San Francisco, CA
Data Analyst Apr 2016 - Nov 2018

- Led end-to-end data analysis projects, managing tasks from data collection and preprocessing to analysis and reporting, ensuring that actionable insights were delivered to support business decision-making.
- Developed custom Python scripts for web scraping and API integration, for collection of data from various web services and websites.
- Developed efficient data extraction pipelines to gather and process large datasets, ensuring timely and accurate data for further analysis.
- Implemented data preprocessing workflows to clean, transform, and format data, maintaining consistency and preparing datasets for deeper analysis.
- Conducted exploratory data analysis (EDA) to uncover trends, patterns, and relationships, enabling data-driven business strategies and improving model accuracy.
- Applied data transformation and visualization techniques to enhance the clarity and interpretability of complex datasets, supporting key business decisions.
- Built statistical models to address business challenges such as forecasting trends, customer segmentation, and sentiment analysis, ensuring reliable and scalable results.
- Automated reporting processes and created interactive dashboards to provide real-time insights and enhance decision-making across teams.

- Collaborated with business teams to define objectives, gather requirements, and translate them into actionable data-driven solutions.
- Implemented monitoring systems to track and analyze performance metrics, ensuring consistent and reliable data flows throughout the analysis pipeline.
- Developed user-friendly data interfaces and dashboards, allowing stakeholders to easily explore and interpret key insights for strategic use.

Skills

Languages:	Python, JavaScript
Frameworks:	TensorFlow, PyTorch, Keras, MLflow, Reactjs, Nodejs,Fastapi
Libraries:	Langchain, Llamaindex, NLTK, HuggingFace Transformers, Scikit-learn, NumPy, Pandas
LLMs:	LLaMA 2, Anthropic Claude, BERT, LLaMA 3.1, BART
Database:	SQL, PostgreSQL, Faiss, AWS EKS, Bedrock, RDS, CloudWatch, Lambda, S3 bucket, EC2