Rishi Kumar

📞+1 5513446747 — ✉ rishi.k.st24@gmail.com — linkedin.com/rishi-kumar-06 — github.com/rishi-kumar0612

Experience

StableCoupons Inc, NY, *Gen AI Engineer*   **Sep 2024 - Present**

- Designed, developed, and deployed an AI-powered chatbot solution for the BulkMagic e-commerce platform using AWS Lex. The chatbot enables seamless user interaction by assisting customers with order tracking, and inquiries.
- Implemented AWS Lambda functions to handle user queries within the AWS Lex chatbot. The Lambda functions efficiently retrieved real-time inventory and pricing data from AWS RDS (PostgreSQL), ensuring accurate responses.
- Developed comprehensive functional workflows to compare LLM performance across multiple platforms, including AWS Bedrock, OpenAI Playground, and Hugging Face Hub. Evaluated key factors such as response accuracy and cost efficiency to determine optimal model selection for various context driven NLP tasks.
- Enhanced chatbot intelligence by leveraging Amazon Bedrock with LLM models to build a Retrieval-Augmented Generation (RAG) application, enabling personalized and context-aware user interactions for general inquiries.

Radical AI, NY, *AI Engineer*   **Sep 2024 - Dec 2024**

- Functioned as an Applied AI Engineer, collaborating with teams to design and build a Quiz application using Gemini.
- Optimized data pipelines by utilizing PyPDFLoader, LangChain, and ChromaDB for data processing and retrieval.
- Integrated Vertex AI with Google Gemini to generate context-driven quiz questions, enhancing the user engagement.

Teledyne LeCroy, NY, *Software Engineer Co-Op*   **Sep 2023 - May 2024**

- Functioned as a Software Engineer Co-Op, upgrading QPHY solutions with Python-based automated testing protocols and conducting rigorous pre-release testing of oscilloscope firmware to identify and resolve bugs.
- Created the PCIe Fixture kit characterization for Gen 4 & Gen 5, developing and integrating automated scripts to validate performance and compliance guidelines provided by Intel using Time Domain Reflectometry.
- Enhanced USB 3 RX/TX scripts with Host Program Control (HPC), implementing specialized test scripts to add HPC functionality within QPHY, also created python script for waveform analysis of power devices.

Paras Defence and Space Technology Ltd, INDIA, *Software Developer*   **Nov 2019 - Jun 2022**

- Functioned as a Software Developer, creating GUIs for aerospace applications in Python, programming Embedded-C firmware with IoT integration, and constructing analytics databases using SQL (SQLite) and NoSQL (MongoDB).
- Contributed to an AI-powered periscope pan-tilt unit, enabling faster object detection and improved system response.
- Implemented CI/CD pipelines using GitHub Actions, to automate, optimize, and seamlessly streamline continuous integration, deployment, and testing processes, paired with backend development in Python/Flask.
- Applied Machine Learning algorithms such as gradient descent and xgboost to optimize PID parameters in motorized systems, bolstering motion control and system precision which lead to reduced power consumption by the motor.

Education

Stevens Institute of Technology, NJ, MS in Applied Artificial Intelligence	**Sep 2022 - May 2024**
Mumbai University, INDIA, Bachelor of Engineering in Computer Engineering	**Sep 2015 - May 2019**

Skills

Technical Skills: Python, SQL, R, C++
Cloud Skills: AWS(S3, EC2, Lex, Lambda, RDS, QuickSight), GCP(VertexAI, BigQuery, Looker, Dialogflow)
AI Skills: Machine Learning, Regression, TensorFlow, PyTorch, Keras, LLM Finetuning, HuggingFace, LangChain
Data Skills: Vector DBs, Tableau, Power BI, Alteryx, Snowflake, PySpark, Airflow, Databricks, PostgreSQL
Additional Skills: Git, Perforce, Data Structures, Algorithms, Natural Language Processing, CI/CD, Docker

Projects

Serverless AI Chatbot with AWS *AWS, DynamoDB, API Gateway, FastAPI, OpenAI, Terraform, GitHub Actions*
Developed a fully serverless chatbot using AWS Lex and OpenAI's GPT-4, integrated with DynamoDB for session management. Exposed the chatbot API via AWS API Gateway and FastAPI, leveraging Lambda for scalability. Automated infrastructure deployment with Terraform.

Certifications

- **Docker Foundations Professional Certificate**   Issued May 2024
- **AWS Essential Training for Developers**   Issued Feb 2024