Ish Kumar

Email : ishkumar@buffalo.edu
Linkedin: linkedin.com/in/ish-kumar
Noida, India

Education

- **University at Buffalo, SUNY, New York, USA -** Masters in Engineering Science (Data Science)
  Courses: Intro to Probability Theory, Machine Learning, Computer Vision and Image Processing, Data Model and Query Language
- **APJAK Technical University, India -** Bachelors in Electronics and Communication Engineering

Skills Summary

- **Languages and Tools:** Python; SQL(PostgreSQL, NoSQL); MATLAB; R
- **ML & AI:** Tensorflow; Pytorch; Transformers; LLMs; Prompt Engineering; Hugging Face
- **Frameworks and Tools:** LangChain; FAISS; Django; OpenAI Embeddings; Retrieval Augmented Generation(RAG); Docker; AWS (Bedrock, Lambda, S3); Streamlit; Git
- **Data Visualization** Tableau; Seaborn; Plotly; Matplotlib

Experience

- **Outlier.AI - Prompt Engineer (Part-Time)** Aug 2024
  - Developed and optimized prompts for evaluating and benchmarking multi-modal AI model coding abilities across key data science libraries (NumPy, pandas, matplotlib, scikit-learn), including tests for correctness, efficiency, and adherence to Pythonic style.
  - Implemented LLM-specific prompt tuning (few-shot learning and chain-of-thought prompting), enhancing understanding of model behavior across transformer architectures (GPT, Gemini, Sora).
  - Conducted **100+ evaluations** of AI-generated code, information, and video outputs, including text-to-video alignment tasks; identified and documented 25+ edge cases and areas for improvement in video generation, resulting in enhanced feedback precision.
  - Introduced adversarial and stress-testing prompts to evaluate model resilience, contributing to safer and more reliable multi-modal AI systems.
- **Airdrive Smart Solutions PVT LTD - Associate Data Scientist** Jun 2021 - Jul 2022
  - Optimized **customer transaction data** preprocessing pipelines using data automation techniques, reducing manual data cleaning efforts by **20%** and improving data availability for analytics applications.
  - Evaluated **classification models for fraud detection** using K-Fold Cross Validation and Hyperparameter Tuning in **TensorFlow** and **PyTorch**, contributing to an improvement in the **F1-score** from **72%** to **81%** while mitigating overfitting
  - Developed NER and text classification models to extract product categories from reviews using deep learning (**PyTorch**), reaching **78%** accuracy, enabling enhanced sentiment analysis for potential application in understanding demand drivers.
  - **Containerized** model evaluation scripts using **Docker** to ensure consistent runtime environments across development and testing.
- **Value Creation - Associate Data Scientist** Aug 2020 - Apr 2021
  - Collaborated with business stakeholders to engineer **time series forecasting models (ARIMA, Prophet)** to predict temporary workforce demand, contributing to a reduction in projected overstaffing.
  - Worked with senior data scientists to build a **random forest and XGBoost** model for employee churn prediction, incorporating tenure, performance ratings, and engagement metrics. This project contributed to improved prediction accuracy from **74%** to **86%**
  - Designed Tableau dashboards with PostgreSQL to track attrition trends, cutting reporting lag by **24%**.

Academic Projects

- **Automated SQL Query Generator using Google Gemini**
  - Created a Streamlit application using Google Gemini on GCP & Vertex AI APIs to generate SQL queries from natural language, achieving **95%** accuracy across 30+ test cases covering diverse query types (joins, aggregations, subqueries).
  - Implemented dynamic SQL validation and optimization mechanisms, reducing syntax errors across the test suite and improving query execution time by an average of **15%** on sample database schemas.
  - Leveraged multi-modal AI to handle domain-specific queries, enhancing the system's ability to generate optimized SQL for various data types, resulting in a **40%** improvement in contextual accuracy.
- **Multiagent Retrieval-Augmented Generation (RAG) System**
  - Constructed a modular architecture for multi-agent RAG system using LangChain and FAISS enabling plug-and-play document retrievers and real-time query handling.
  - Implemented query routing to switch between Wikipedia and vector store, boosting efficiency by **25%**
- **Hybrid Prompt Injection Detection System**
  - Developed a hybrid prompt injection detection system integrating regular expressions, compound pattern analysis, and machine learning classifiers, adaptable to various LLMs.
  - Implemented enhanced regex patterns categorized and weighted by attack type (direct instructions, harmful content, probing, etc.). Designed compound pattern detection to identify combinations of suspicious phrases.
  - The system integrates rule-based and machine learning techniques for improved accuracy and robustness
  - **Code available on GitHub: Source Code (Active development to refine outputs, and enhance performance)**