Henry Liang

hanliang2023@u.northwestern.edu github.com/HenryLiang-123 310-853-9937

Education

Northwestern University – M.S. in Machine Learning and Data Science Dec 2023

University of California, Los Angeles – B.S. in Applied Mathematics, Statistics Mar 2022

Work Experience

Applied Scientist Intern, Amazon (Capstone) – Evanston, IL Sept 2023 – Dec 2023

- Developed a custom LLM application with LangChain to explain differences between knowledge graph snapshots and detect network routing anomalies and hijacking incidents
- Achieved 99% accuracy in summarizing path-related attributes by enhancing responses from GPT-3.5 Turbo through knowledge graph augmented generation (KGAG)
- Reduced erroneous Cypher queries by 95% via prompt engineering and continuous fine-tuning of the query generation process using a few-shot learning feedback loop
- Scraped Boarder Gateway Protocol data to determine peering relationships between autonomous systems and stored them in Neo4j graph database

Data Science Intern, Roblox – San Mateo, CA June 2023 – Sept 2023

- Architected a diagnostic framework for the data science and analytics team to address extreme impression changes, isolating monetization and update frequency as important drivers
- Created a 94% accurate early warning system for game developers using an XGBoost model to forecast probability of significant impression shifts
- Implemented an anomaly detection algorithm to identify and quantify extreme shifts in game impressions
- Designed an ETL pipeline with Spark SQL to extract and analyze game-level metrics across multiple sources with 1,000,000,000+ observations

Machine Learning Researcher, Northwestern University – Evanston, IL Oct 2022 – June 2023

- Led design and implementation of a scalable AWS Internet of Things (IoT) streaming pipeline for real-time time-series forecasting; achieved a throughput of up to 55 predictions per second
- Improved data processing efficiency by 75% and decreased cost by 50% through optimizing data ingestion and model retraining pipelines with serverless architectures
- Trained and deployed an LSTM model on AWS EC2 for hourly Divvy bike demand forecasting for each station; realized a mean absolute error of 9 bikes

Projects

Conversational AI for Cancer Drugs

- Created 96% accurate conversational AI to answer questions about clinical trials for cancer drugs by enhancing responses from GPT-4 through retrieval augmented generation (RAG)
- Prepared foundational data for RAG by chunking and transforming clinical documents into embeddings and storing them in a Pinecone vector database

Gender Biases in Text

- Applied Word2Vec and BERT to create custom word embeddings to detect gender biases in text by analyzing cosine similarities between gendered and gender-neutral words
- Implemented tokenization and text normalization techniques including stop-word removal and stemming to enhance the accuracy of word embeddings

Skills

Languages: Python, R, Java, JavaScript, C++

Tools: TensorFlow, PyTorch, OpenCV, NLTK, LangChain, Git, Docker, Apache Spark, Apache Hadoop, MapReduce, Cypher, Hive, Presto, PostgreSQL, MySQL, Tableau