Chirayu Tripathi

Santa Clara, CA | P: +1 352-740-6202 | chirayutripathi7@gmail.com | LinkedIn | GitHub | Portfolio | HuggingFace

WORK EXPERIENCE

Software Development Engineer   Dec 2024 – Present
Amazon   Santa Clara, CA, USA
- Spearheaded the desktop internationalization (I18n) of Rufus, Amazon's generative AI shopping assistant, successfully architecting and launching the service across 8 major global marketplaces.
- Scaled a generative AI pipeline for "Hear the Highlights," producing conversational audio summaries from product data and reviews across 6M+ products, cited in Amazon's 2025 Q3 earnings.
- Designed a multi-modal LLM evaluation framework to benchmark the quality of AI-generated audio and text, achieving a 90% agreement rate with human annotators and reducing content defects to less than 5%.

Data Scientist   July 2024 – Dec 2024
RevScale AI   Fort Lauderdale, FL, USA
- Architected an event-driven data pipeline using Azure Data Lake and Azure Functions to automate prospect file ingestion into MongoDB, achieving near real-time data synchronization and reducing processing latency by 40%.
- Fine-tuned a Small Language Model (SLM) reach-out agent that utilized optimal contact timing and conversation styles, resulting in a 20% increase in prospect response rates and automating over 1,000 weekly outreach interactions.

Machine Learning Researcher (RA)   Nov 2022 – June 2024
UF Health, PRISMAp Lab   Gainesville, FL, USA
- Employed a transformer (GatorTron) to extract social determinants of health (SDoH) from clinical notes, addressing the lack of SDoH data and resulting in a 4% AUROC increase for mechanical ventilation complication prediction.
- Leveraged the Fairness (AIF-360) and Interpretability (SHAP) libraries to address racial bias and lack of explainability in medical models, resulting in improved fairness and unveiling the factors influencing potential underlying conditions.
- Engineered predictive methodologies for arteriovenous fistula surgeries, employing time-series modeling techniques, to increase cost savings through the reduction of expensive scan procedures, validated by an impressive R2 score of 0.75.

Machine Learning Engineer   Jan 2021 – Jul 2022
ZkAGI (formerly Bitbaza.io)   Chennai, TN, India
- Led the real-time 'Social Listener' initiative using LLMs and Apache Kafka to identify potential clients on social media, replacing manual workflows and resulting in the onboarding of 2 new clients.
- Deployed a Deep Neural Network (DNN) on AWS to classify prospective customers from client lists, accelerating the identification of high-potential clients and reducing processing time by 50%.
- Engineered a cost-effective REST API by deploying a T5 transformer to replace GPT-3, maintaining performance efficiency while achieving a 15% reduction in operational costs.

PROJECTS

nl2query: Convert Text to Database Queries | Tech Stack: Text Generation, CodeT5+, Phi-2. LINK   Apr 2024
- Designed a framework to translate natural language texts into queries for the Pandas, MongoDB, Kusto, and Neo4j databases, enabling search across data sources using English, eliminating the need for specialized database languages.

ACHIEVEMENTS
- Co-authored the paper "Transparent AI: Developing an Explainable Interface for Predicting Postoperative Complications". LINK
- Co-authored ML-based AKI risk prediction research using clinical notes and SDOH, presented at AKI & CRRT Conference 2025. LINK
- Deployed a PyPI library for the nl2query project, garnering over 16,000 downloads and 91 stars on GitHub. [LINK]
- Contributed to open source by implementing CodeT 5 support for the SimpleT5 library. LINK

SKILLS

Technical Skills: Deep Learning, RPA, Statistics, Linux, DBMS, Algorithms, Data Visualization, MLOps, NLP, LLM.
Programming Languages & Databases: Python, C/C++, Java, R, Erlang, MySQL, MongoDB, Neo4j.
Frameworks and Applications: Hadoop, Tableau, NumPy, Pandas, Matplotlib, Scikit-learn, PyTorch, TensorFlow, LangChain, Git, AWS, Flask, REST API, Apache Spark, SAS, ETL, Docker, Kubernetes.

EDUCATION

University of Florida, MS in Computer Science (GPA: 3.8)   May 2024
Medi-Caps University, B.Tech in Computer Science (GPA: 3.8)   May 2021