Akhil Ghosh

Software Engineer @ AbbVie Inc. (U.S. Citizen)

Email : akhilghosh27@gmail.com
Mobile : 630 750 2260
linkedin.com/in/akhilghosh

Software Engineer with hands-on experience building NLP pipelines, fine-tuning large language models, and developing data-driven applications processing billions of rows in healthcare. Built and deployed analytics platforms serving 200+ monthly users at AbbVie.

Skills

- **Programming & ML**: Python, SQL, R, PyTorch, TensorFlow, Pandas, Scikit-Learn, Git, Linux, Docker, Claude Code
- **Data & Cloud**: Postgres, Snowflake, Azure certification, AWS, Hadoop, Impala, Trino, ETL Pipelines
- **Knowledge**: SAP BTP, SAP BDC, SAP Clean Core, SAP Gen AI, CI/CD, JSON, NLP, Machine Learning, Deep Learning, LLM Fine-tuning, Agile, Data Modeling, Microservices

Experience

- **AbbVie Inc. (Contractor via Kru-Marc Inc.)** — North Chicago, IL
  Software Engineer — *Feb. 2024 – Present*
  - **Gen-AI Data Pipeline & Analysis**: Built an end-to-end data pipeline that leveraged Claude Code and LLMs to define comorbidity cohorts and accelerate dataset construction for a 600K-patient migraine cohort from a 40M-patient Japanese medical claims database (JMDC). Involved translation of drug names & defining a Japanese market basket for migraine drugs. Implemented treatment-line classification logic and generated Kaplan-Meier survival models for adherence analysis. Reconciled hospital and payer claims for consistent cohort definitions. Helped researchers plan for their strategy within the migraine market in Japan.
  - **Application Development**: Led and engineered a large-scale data integration project, consolidating information from 57 healthcare databases and billions of rows into a single, high-performance application utilized by 200+ users monthly. This solution reduced server costs by replacing 13 standalone applications and drove innovation in the Real World Data team by providing researchers greater visibility into available patient-level demographic data.
  - **NLP Standardization Pipeline**: Built an NLP pipeline using fuzzy string matching (Levenshtein distance) and similarity scoring to standardize drug, condition, and procedure names across multiple healthcare databases and countries. Generated reference tables mapping naming conventions to canonical forms, improving data quality and cross-database comparability. Adopted across RWD teams for downstream analytics.
- **Loyola University Chicago** — Chicago, IL
  Graduate Research Assistant — *Aug. 2022 - Dec. 2023*
  - **Quantitative Modeling**: Applied analytical and problem-solving techniques in R leveraging the tidyverse package, to process and analyze stickleback fossil measurements. This prepared the dataset for future statistical analysis, and allowed for a deeper understanding of patterns and trends of trait evolution, ensuring data-driven conclusions in later research phases.
  - **Data Imputation**: Designed and implemented an imputation algorithm using the MICE package in R to predict the gender of Stickleback fossils. Validated the model's accuracy at 85% using K-Fold cross-validation, proving the viability of future analysis on the imputed fossil dataset.

Projects

- **Fake News Classifier**:
  - **AI Model Fine-tuning with Big Data**: Fine-tuned an AI LLAMA 7B model utilizing docker on a dataset of over 1.6 million news articles. Utilized data engineering techniques and the fine-tuned deep learning AI model to generate and preprocess a synthetic fake news dataset.
  - **Model Creation**: Developed an AI transformer classifier model in PyTorch, achieving 83% accuracy on an external fake news dataset. Feature engineering was employed to determine the best hyperparameters for this applied machine learning model. Solution helped tackle the issue of fake content being distributed on the internet.

Education

- **Georgia Institute of Technology (Part-time, Asynchronous)** — Atlanta, GA
  M.S. in Computer Science (Machine Learning Specialization) — *Aug. 2024 - Current*
- **Loyola University Chicago** — Chicago, IL
  M.S. in Data Science, ***GPA: 3.97/4.0*** — *Aug. 2022 - Dec. 2023*
- **Loyola University Chicago** — Chicago, IL
  B.S. in Statistics & B.S. in Biology (Double Major), ***GPA: 3.75/4.0*** — *Aug. 2018 - May 2022*