

Shazan Ansar Mohammed

shazanansar@gmail.com

| +1 7473529855. | CA, US

| [LinkedIn](#)

| [Github](#)

SUMMARY

Data and Machine Learning Engineer with hands-on experience building end-to-end ML and data pipelines, from large-scale data preparation to model training, evaluation, and analysis. Worked on multilingual NLP systems in clinical domains, delivering measurable improvements through model fine-tuning, structured evaluation, and data quality engineering. Proficient in Python, SQL, PyTorch, and AWS, with experience designing reproducible workflows that support research, analytics, and ML production use cases.

EXPERIENCE

California State University-Dominguez Hills, *Research Assistant*

Feb 2025 – Aug 2025

- Designed and implemented a medical-domain translation pipeline using mBART and NLLB-200, achieving a 32% BLEU score improvement.
- Curated and cleaned a domain-specific corpus (70k+ entries) and optimized fine-tuning with PyTorch.
- Conducted error profiling and ethical assessment for low-resource NMT systems in clinical contexts.
- Created visualization dashboards for model evaluation (BLEU, METEOR, BERTScore, COMET).

PROJECTS

Real-Time Data Lakehouse for Analytics and ML Features

- Designed a medallion-style lakehouse (raw, cleaned, curated) to address inconsistent analytics data, enabling near real-time processing and reducing data latency by 40%.
- Implemented schema enforcement, late-arrival handling, and partitioning strategies, reducing data quality issues and downstream pipeline failures by ~30%.
- Delivered reusable, feature-ready datasets that shortened ML experiment setup time by 25%.

LLM-Powered Analytics Assistant (RAG, Evaluation, Cost Optimization)

- Built a retrieval-augmented generation pipeline using embeddings and grounded retrieval to improve accuracy of analytics question answering, increasing answer relevance across repeated queries.
- Implemented automated evaluation and inference optimizations (caching, retrieval filtering, adaptive routing), reducing response latency by 35% and controlling per-query inference cost.
- Standardized prompt and model evaluation workflows, enabling safe iteration and preventing quality regressions across releases.

End-to-End ML Experimentation and Evaluation Framework

- Created a reproducible ML experimentation framework with data versioning, configuration management, and metric tracking, improving experiment comparability across training runs.
- Automated evaluation and regression testing to detect performance drift, reducing manual validation effort and accelerating iteration cycles.

EDUCATION

California State University Dominguez Hills (CSUDH)

M.S. in Computer Science | GPA: 4.0 |

2024 – Present

CERTIFICATIONS & PUBLICATIONS

- RANLP 2025: *Advancing Clinical Translation in Nepali through Fine-Tuned Multilingual Models* (Co-author) ([Link](#))
- AWS Certified Solutions Architect – Associate, Google Data Analytics

SKILLS

Programming and Data: Python, SQL, Java, PySpark, Bash

Data Engineering and Platforms: Apache Spark (batch concepts), Airflow concepts, Delta Lake concepts, medallion architecture, ETL/ELT pipelines, lakehouse design, data modeling, feature engineering, data quality and reliability

Machine Learning and AI: PyTorch, TensorFlow, scikit-learn, applied machine learning, NLP, multilingual models, embeddings, LLM fine-tuning concepts, RAG, model evaluation and error analysis

Cloud and MLOps: AWS (EC2, S3, Lambda), GCP and Vertex AI, containerized workflows, Docker, scalable training and inference pipelines, latency and cost-aware deployment

Analytics and Visualization: Pandas, NumPy, Tableau, metric design, evaluation dashboards

Software and Professional Practices: Git, Linux/Unix, reproducible workflows, experiment tracking concepts, technical documentation, compliance-aware development, cross-functional collaboration