Roksana Mirzaei

London,UK • LinkedIn • roksanamirzaei1999@gmail.com • +447301543422

Experience

Sainsbury's London, UK
Machine learning engineer September 23, 23 – Present

- Implemented **pre-run validation logic for daily Airflow DAGs** to catch failure-prone points, preventing unnecessary executions, reducing cross-team engineering effort by 33%, **saving ~~$~~30K/month** in compute, and improving pipeline reliability.
- **Optimized time- and memory-intensive module**, **achieving 2–10× speedups (50–90%)** and significantly improving scalability.
- **Improved AWS migration efficiency** with dependency-driven parallel plan, reducing effort from 5 to 2 months and informing key decisions.
- Improved accuracy of **10 KPIs** for **multi-touch attribution** pipeline by fixing subtle SQL and Python code errors, resolving cases of significant **over-reporting (>25x)** and ensuring reliable campaign performance reporting.
- **Improved pipeline reliability** by validating internal outputs against third-party platforms (Meta, DV360, Citrus-ad, etc), covering **60%** of channels; resolved accuracy gaps, **and stabilized downstream reporting**.
- Reduced technical debt and pipeline fragility by **re-architecting Airflow pipeline logic (SQL + PySpark)** to replace hard-coded conditional logic with ID-based joins against a Snowflake source-of-truth table, **decoupling pipelines from upstream schema** and naming changes and improving reliability of customer-facing analytics.
- **Owned E2E technical interface** for new team requirements, and aligning DS and engineering teams to increase delivery throughput and minimize blockers.
- **Demonstrated cross-team influence** by unblocking AI initiatives across multiple teams by sourcing critical datasets and implemented **ETL pipeline** to enable vision model experiments.
- Delivered GenAI PoCs addressing business challenges, influencing engineering roadmaps, some progressing into production.
  - fine-tuned **diffusion models** for virtual try-on using **FastAPI**, evaluated with **SSIM, PSNR, and MSE** metrics.
  - fine-tuned **LLMs** on internal data for incident triage and root-cause analysis using agent-based workflows and **RAG pipeline**.
  - GraphRAG prototype: brand-owner mapping using **knowledge graphs** and **Snowflake**-hosted data.

Amazon UK London, UK
AWS Cloud Women Mentorship Program September 2024 - June 2025

Research

- **Explainable AI Evaluation:** Mirzaei, S., Mao, H., Al-Nima, R.R.O. and Woo, W.L., 2023. Explainable AI Evaluation: A Top-Down Approach for Selecting Optimal Explanations for Black Box Models. Information, 15(1), p.4. **Supervisor:** Prof. Wai Lok Woo
- **Medium Article:** Roksana Mirzaei (2024). LLM Reasoning. [online] Medium. Available at: `https://medium.com/@roksanamirzaei/llm-reasoning-57bf6ac4630a`

Skills

ML Programming: Python, PySpark, TensorFlow, FastAPI
Machine Learning: NLP, LLMs, Agentic AI, Feature Engineering, Diffusion Models, Model Training, Deep Learning
Languages: Python, Java, TypeScript / JavaScript, Bash
DevOps: Docker, Terraform, CI/CD, GCP, Azure, AWS Services, MLflow, Grafana
MLOps: Airflow, dbt
Databases / Data Modeling: MongoDB, Snowflake, Data Vault methodology, Dimensional Modeling
Other: Git, Multi-Processing, Multi-Threading

Education

Northumbria University Newcastle upon Tyne, UK
M.Sc. Advanced Computer Science (Distinction) September 2022 –September Sep 2023

Massachusetts Institute of Technology (MIT)
Flow Matching and Diffusion Models (Online Course) Jun 2025

University of California, Berkeley
Advanced Large Language Model Agents (Online Course) April 2025
Large Language Model Agents (Online Course) September 2024