

HAMZA FAHEEM MIR

hamzafaheem996@gmail.com • [linkedin.com/in/HamzaFaheem](https://www.linkedin.com/in/HamzaFaheem) • Islamabad, Pakistan

PROFESSIONAL SUMMARY

Senior ML Engineer (5+ yrs) specializing in reproducible benchmarking and end-to-end ML/data pipelines. Strong Python and Docker; Airflow + PySpark on AWS (S3, Athena, EMR, SageMaker). Deployed GPT-powered analytics via Kubernetes, Flask, Streamlit and LangChain; drove 90%+ IVT loss reduction (~\$5M/yr) and 18% forecasting MAPE gain. Skilled at bridging engineering and data-science roadmaps, validating AI-generated implementations, and mentoring teams to raise code-quality standards.

CORE SKILLS

- ML & Benchmarking: Reproducible experiment design, MLflow, model/perf benchmarking, training-speed profiling
- Data & Orchestration: PySpark, Airflow, Databricks, SQL
- MLOps & Cloud: Docker, Kubernetes, AWS (S3, Athena, EMR, SageMaker)
- Languages & Tools: Python, Git; LangChain, Prophet, scikit-learn, statsmodels

EXPERIENCE

MonetizeMore (Canada-based AdTech company, Google Certified Publishing Partner) Remote | Jan 2022 – Present

Senior Machine Learning Engineer

- Led “Traffic Cop” IVT program: Airflow + PySpark on AWS; reduced clawbacks 5% → 0.52% (~\$5M/yr).
- Benchmarked fraud models & training throughput; standardized MLflow-tracked runs and Dockerized environments for reproducibility.
- Built GPT-powered analytics (LangChain + Flask + Streamlit on Kubernetes); cut case-resolution time ≈80%.
- Engineered Prophet forecasting workflow; improved MAPE 18% via variant benchmarking & scheduled retrains in Airflow.
- Authored step-by-step experiment plans & validation checklists to align implementation with specs; enforced code-review standards.

Baltoro (ML/AI data platform, US-based)

Remote | Apr 2021 – Dec 2021

Data Scientist

- Delivered geospatial ETA predictions in PySpark MLlib; cut compute time 50× (75h → 1.5h) via parallelization.
- Prototyped trigger-based ML execution enabling on-demand benchmarking & reproducible reruns.
- Optimized probabilistic models with Spark/MapReduce; improved training throughput and reliability.

PROJECTS

- **Website Analysis Agent** — Architected a Kubernetes-native Flask API that runs Playwright scrapers and pipes JavaScript signals to a LangChain LLM, rendered via Streamlit for on-demand site-health and ad-compliance insights under 60s latency.
- **IVT Dashboard with LangChain** — Designed a GPT-powered Ad Ops dashboard (Flask + Streamlit + LangChain) deployed on Kubernetes with AWS Cognito auth and Route 53; cut validation time by 80%.
- **Risk-Level Assessment Pipeline** — Built PySpark + Airflow pipeline scoring invalid traffic across Internet Network Hierarchy; persisted to S3 and auto-hydrated Athena tables for real-time Metabase monitoring.
- **Revenue Forecasting with Prophet** — Improved MAPE by 18 % across two products via logistic-growth modelling and Airflow orchestration.
- **MLOps Rules-Generation Pipeline** — Created a SageMaker + Airflow pipeline that derives decision-tree rules, registers models, and ships them to S3/Athena via CI/CD.

EDUCATION

B.S. Computer Science — National University of Computer & Emerging Sciences (Islamabad), 2019