

Farah Mohamed Abdallah

Junior Data Engineer

fxrxhmohamed@gmail.com | (+20) 1003293502 | Giza, Egypt

linkedin.com/in/farah-m-abdallah | github.com/fxrxhmohamed

Summary

Junior Data Engineer with hands-on experience building scalable ETL/ELT pipelines, transforming and integrating data, and developing production-ready datasets for machine learning and analytics using Python, SQL, and Apache Spark. Skilled in relational and NoSQL databases, DataOps, and MLOps, with a strong foundation in delivering reliable datasets for analytics and machine learning.

Technical Skills

- **Programming Languages:** Python, Java, SQL, JavaScript, HTML, CSS
- **Data Engineering & Big Data:** Apache Spark, Apache Hadoop, ETL, Data Warehouse, dbt, Web Scraping, PostgreSQL, MySQL, Microsoft SQL Server, MongoDB, NoSQL, GIS
- **ML & Data Science:** Machine Learning, Deep Learning, NLP, Feature Engineering, Hypothesis Testing, Scikit-learn, NumPy, Pandas, Matplotlib
- **MLOps & Deployment:** MLOps, Docker, MLflow, FastAPI, Flask, REST APIs, Postman, Git/GitHub
- **BI & Analytics:** Power BI, DAX, Advanced Excel
- **Software Engineering:** OOP & Data Structures, Django, Laravel, JUnit Testing, UI/UX Fundamentals

Education

Bachelor's Degree in Computer Science and Artificial Intelligence

Faculty of Computer Science and Artificial Intelligence, Cairo University, Egypt, 2026

Grade: Excellent (Honors)

Work Experience

DataOps Engineer Trainee

National Telecommunication Institute (NTI) HireReady, Cairo | July 2026 - Present

- Built and automated ETL/ELT pipelines for extracting, transforming, validating, and loading data from multiple sources.

- Designed analytical data models using Star Schema, Snowflake Schema, and Fact/Dimension tables for data warehousing.
- Developed, tested, and maintained scalable ETL/ELT workflows using Python, SQL, Pandas, PySpark, and Spark SQL to support analytics and machine learning applications.
- Orchestrated batch and event-driven data pipelines with workflow automation, scheduling, monitoring, and data quality monitoring.
- Integrated DataOps practices including CI/CD concepts, automated pipeline monitoring, and reliability improvements.
- Implemented real-time and distributed data processing concepts using Apache Spark and Apache Kafka as part of hands-on projects.

IBM Data Scientist Course Trainee

Digital Egypt Pioneers Initiative, Cairo (Hybrid) | July – December 2025

- Utilized Python, Pandas, and SQL to clean and analyze 3+ real-world datasets, identifying key patterns to support model training pipelines.
- Cleaned, transformed, integrated, and analyzed real-world datasets to prepare production-ready data for analytics and machine learning models.
- Built and deployed supervised ML models integrated into an MLOps pipeline using MLflow for experiment tracking, versioning, and reproducible delivery.

Machine Learning Trainee

Orange Digital Center, Cairo (Hybrid) | September – October 2024

- Mastered intensive training covering Linear Algebra, Probability, Supervised & Unsupervised Learning, Neural Networks, and Intro to DNN.
- Applied ML concepts through hands-on projects, achieving 100% final score.

Network Security Intern

Fixed Solutions Company, Cairo | July – November 2023

- Configured 20+ firewalls, routers, and switches, improving network security reliability.
- Implemented firewall policies, VLAN configurations, and basic IDS/IPS setups through hands-on training

Projects

Talk2System | AI-Powered Requirements Engineering Platform (Graduation Project)

AI-powered platform that transforms stakeholder meetings into structured software requirements, automatically generating SRS documents and UML diagrams from voice, audio, or text using NLP and LLMs. It streamlines requirements engineering by reducing manual effort and accelerating software planning.

- Built an end-to-end AI-powered requirements engineering platform.
- Automated speech-to-text transcription, requirement extraction, and SRS/UML generation.

- Integrated Qwen, Google Gemini, AssemblyAI, and Kroki APIs.
- Optimized AI workflows with asynchronous processing, transcript chunking, retry mechanisms, and robust output validation and testing.
- Developed a scalable full-stack application using React, FastAPI, Python, and PostgreSQL.

Sales Forecasting & Optimization System

Designed and deployed an end-to-end sales forecasting system with automated ML pipelines, FastAPI, MLflow, and Streamlit, achieving $R^2 = 0.99996$ using LightGBM.

- Designed and deployed an end-to-end sales forecasting system using advanced feature engineering (lags, seasonality, promotions, holidays).
- Trained and benchmarked XGBoost, LightGBM, and LSTM, achieving $R^2 = 0.99996$ with LightGBM and low RMSE/MAE.
- Productionized models via FastAPI with MLflow for experiment tracking, versioning, and real-time inference.
- Built an interactive Streamlit dashboard enabling live sales predictions through API integration.
- Implemented model monitoring (latency, logging, and data drift detection) to ensure production reliability

E-Commerce Order Management System (SOA & Microservices)

Built a scalable SOA e-commerce platform using Flask microservices and a Java JSP API gateway to orchestrate end-to-end order processing.

- Designed a scalable Service-Oriented Architecture (SOA) using 5 independent Flask microservices for core e-commerce operations.
- Built a Java JSP (Tomcat) API gateway to orchestrate end-to-end order workflows across distributed backend services.
- Implemented RESTful HTTP/JSON communication with MySQL-backed services for order processing, inventory, pricing, customer management, and notifications.

Handwritten Digits Recognition

Containerized a handwritten digit recognition system using Random Forest, achieving 96.7% accuracy on the MNIST dataset with automated testing and code quality enforcement.

- Trained a Random Forest Classifier on MNIST (70K images) achieving 96.7% test accuracy with full model persistence via pickle.
- Containerized with Docker, enforced code quality via Ruff, and ensured reliability through unit tests and Git LFS versioning

Wuzzuf Jobs Web Scraper

Engineered an automated job data extraction pipeline using Python, Selenium, and BeautifulSoup for real-time market insights and structured data export.

- Developed automated pipelines to scrape job listings, cutting manual collection by 90%.

- Enabled real-time job market insights with structured CSV/JSON outputs.

Courses and Certifications

- AI and Machine Learning, Sprints x Microsoft | Sep 2025
- Machine Learning in Python, Udemy | Sep 2025
- Introduction to Generative AI Studio, Google Cloud | Sep 2025
- Sentiment Analysis in Python, DataCamp | Sep 2025
- Natural Language Processing with spaCy , DataCamp | Oct 2025
- CS50: Introduction to Computer Science, Harvard University | Nov 2025

Languages

Arabic (Native) | **English** (Fluent) | **Spanish** (Intermediate)