

EFTHIMIOS VLAHOS

(516) 606-0942 | [EfthimiosPortfolio.com](https://efthimios.com) | vlahos89@gmail.com | [linkedin.com/in/efthimios](https://www.linkedin.com/in/efthimios) | github.com/EfthimiosVlahos

PROFESSIONAL EXPERIENCE

CON EDISON

Manhattan NY [Hybrid]

Machine Learning Engineer

03/2025– Present

- Architected **agentic AI workflows** using **LangChain** and **LangGraph**, enabling cross-functional teams to query diverse databases, automating SQL query generation and report creation, **slashing report delivery time by 40% for business stakeholders**.
- Developed **propensity models with XGBoost** and scikit-learn to predict customer enrollment in Energy Efficiency (EE) programs, **achieving 85% precision and driving \$2M in program uptake** through targeted campaigns.
- Engineered **scalable data pipelines in Databricks using Python and Spark SQL**, integrating external datasets to identify eligible customers for clean energy initiatives, **reducing data processing time by 35%**.
- Built an **end-to-end MLOps framework for Databricks**, leveraging MLflow, Docker, and CI/CD pipelines to **standardize model training, deployment, and monitoring, cutting model onboarding time by 50%** and ensuring consistent performance **across business units**.

CORNERSTONE BUILDING BRANDS

East Meadow, NY [Remote]

Data Scientist

10/2024 – Present

- Built an **end-to-end ML pipeline in Python** using scikit-learn, NumPy, and Pandas, delivering personalized product recommendations that **generated \$3.5M in incremental revenue for sales teams**.
- Delivered a **production-grade web application** using **Flask** and **Docker** that allowed leadership to input product configurations and obtain real-time predictions on **quote-to-close outcomes** via **hyperparameter-tuned logistic regression** and **random forest models**, driving a **20% increase** in conversion rates.
- Developed scalable **ETL pipelines** in **Databricks** with **Python, SQL, and Bash** to extract and transform configuration metrics from **Azure Data Lake**, reducing pipeline runtimes by **35%** and lowering excess inventory costs by **15%**.
- Integrated **SQL-based Power BI dashboards** with real-time data feeds, empowering executives and plant managers to make **data-driven decisions** that align production outputs with dynamic market demand.

MICROSOFT

East Meadow, NY [Remote]

Data Analyst

10/2020 – 05/2024

- Analyzed enterprise-scale datasets with **Python, SQL, and R** to support an **LLM-based chatbot** solution, achieving a **30% reduction** in hallucination rates and a **25% boost** in user satisfaction through rigorous **anomaly detection** and data cleansing.
- Spearheaded **RAG** projects by fine-tuning **Phi model** using **TensorFlow** and **PyTorch**, resulting in enhanced **contextual accuracy** and improved revenue outcomes for key clients.
- Implemented iterative **A/B testing** and **hyperparameter tuning** with **scikit-learn**, decreasing manual escalations by **20%** and optimizing model performance across diverse client knowledge bases.
- Developed comprehensive **data visualizations** with **Seaborn** and **Matplotlib**, uncovering actionable features that drove targeted improvements in LLM functionality and overall operational performance.

SKILLS

Programming:

Python, SQL, Bash, C, C++, Java, R, Rust, TypeScript, JavaScript

Data Science:

TensorFlow, PyTorch, JAX, scikit-learn, PySpark, NumPy, Pandas, Statsmodels, RAG, LangChain, LangGraph, MCP, Agentic AI, Pydantic

Big Data:

AWS, GCP, Databricks, Snowflake, MySQL, MongoDB, Apache Spark, Chroma DB, Pinecone, FAISS

DevOps & MLOps:

Docker, Kubernetes, Terraform, Jenkins, GitHub Actions, Airflow, CI/CD, MLflow, DVC, Grafana, Azure ML, Vertex AI, AWS SageMaker

PROJECTS

AI-Driven Trading System (Agentic AI): Engineered a multi-agent trading simulation using LangChain and LangGraph, orchestrating 10 specialized AI agents (e.g., Sentiment, Risk Manager) to make collaborative stock trading decisions, achieving 10% portfolio growth in backtests. (see GitHub: [AI Hedge Fund](#)).

Generative Deep Learning for Alzheimer's Disease Drug Discovery: Designed and implemented a generative deep learning pipeline using LSTM and GRU networks in TensorFlow/Keras to process 300K generic and 1.5K targeted chemical compounds from ChEMBL, achieving 94.2% training accuracy and generating 18 novel molecular structures (see GitHub: [Generative-AD-DrugDiscovery](#)).

EDUCATION

SUNY Stony Brook University: Stony Brook, NY

GPA: 3.65 / 4.00

Master of Science: Applied Mathematics & Statistics

Graduation: August 2021 - May 2023

Coursework: Data Structures & Algorithms, Data Science, Machine Learning, Numerical Analysis I, II, and III, Big Data Analytics, Probability Theory, Stochastic Calculus, Linear Programming, Real Analysis

CUNY Hunter College: Manhattan, NY

GPA: 3.8 / 4.00

Bachelor of Science: Mathematics & Physics (Minor: Computer Science)

Graduated: August 2016 - May 2020

Coursework: Linear Algebra, Vector Calculus, Ordinary Differential Equations, Partial Differential Equations, Probability & Statistics, Quantum Mechanics, Classical Mechanics, Complex Analysis

CERTIFICATIONS

[AWS Certified Cloud Practitioner](#), AWS (Apr 24)

[Databricks Certified Data Engineer Associate](#), Coursera (Jun 25)

[Large Language Model Operations \(LLMOps\)](#), Coursera (Jun 24)

[MLOps | Machine Learning Operations](#), Coursera (May 24)