Moiz Ahmed

Iselin, NJ, 08830 • moizahmed.swe@gmail.com

Professional Summary

Senior MLOps Engineer with 8+ years of hands-on experience designing ML systems, building scalable Python APIs, and deploying LLM and CV models with robust DevOps pipelines. Combines deep AI infrastructure knowledge with full-stack capabilities to support cross-functional AI teams. Skilled in backend systems using FastAPI and Flask, and frontend dashboards using React and Tailwind for real-time model feedback. Proven success in leading model lifecycle automation, edge deployment, and cloud-native GenAI infrastructure.

Core Skills

- Languages & Frameworks: Python, JavaScript, TypeScript, FastAPI, Flask, React, Tailwind, Spark, TensorFlow, PyTorch
- MLOps & Cloud: Docker, Kubernetes, MLflow, Airflow, Kubeflow, GitHub Actions, AWS (SageMaker, Lambda, etc.), GCP, Runpod, Terraform
- Frontend/Backend Tools: FastAPI, Flask, React.js, Next.js, Node.js, Express, Nest, REST APIs, WebSockets, Chart.js, GraphQL, Socket.io
- AI & GenAI Tools: LangChain, HuggingFace, LlamaIndex, LoRA, Ollama, AutoGen
- Data & Vector Stores: PostgreSQL, MySQL, Redis, Pinecone, Milvus
- Expertise: LLM Infrastructure, Backend API Development, Hybrid Cloud-Edge MLOps, DRL, CV/NLP Systems, GenAI App Delivery

Professional Experience

Lead AI/MLOps & Platform Engineer

GridWise AI Solutions — Boston, MA | Mar 2025 – Present

- Built and maintained FastAPI-based ML model APIs with JWT authentication and integrated caching (Redis).
- Developed an internal web dashboard (React + Chart.js) for monitoring model drift, version metrics, and retraining triggers.
- Reduced deployment time by 30% by templating Helm/Kubernetes stacks for GenAI services.

Senior MLOps Engineer / Full-stack Contributor

GridWise AI Solutions — Boston, MA | Mar 2024 – Feb 2025

- Delivered DRL-based optimization pipelines as modular Python services using Flask and Docker.
- Created an admin portal in React.js to visualize HVAC system performance, energy impact, and feedback logs.
- Standardized the API delivery layer for LLM inference endpoints (LangChain + FastAPI) across projects.

AI Product Developer (Infra + UI)

Sunlytix Energy Intelligence Inc. — Austin, TX | May 2022 – Feb 2024

- Built MLOps backend in Python with support for multiple training schedules and real-time model triggers.
- Designed a React interface for solar farm managers to view GHI forecasts and panel cleaning schedules.
- Integrated Airflow pipelines with API hooks to refresh predictions and log usage analytics.

AI Systems Engineer

VoltLogic Technologies — Denver, CO | Mar 2021 – Apr 2022

- Developed RESTful APIs using FastAPI to expose models to frontend control systems.
- Partnered with the frontend team to create interactive Grafana dashboards with backend model confidence hooks.
- Maintained a versioned model registry with auto-triggered webhooks for pipeline deployment.

ML Engineer / Full-stack ML Tools

DataStream AI Labs — San Francisco, CA | Jan 2020 – Feb 2021

- Built full-stack analytics tooling to monitor energy consumption anomalies using React, Flask, and D3.js.
- Developed an API layer for fraud detection scoring, linked to the smart meter analytics frontend.

Data Analyst → Junior Backend Engineer

Solara Energy Consulting Group — San Jose, CA | Sep 2018 – Dec 2019

- Created internal admin tools using Flask + jQuery to monitor energy audit data and visualization.
- Developed internal libraries for data parsing and visualization embedded into Flask templates.

Data Engineering Intern

PowerNet Solutions — Dallas, TX | May 2017 – Aug 2018

- Built custom dashboards with HTML/JS for telemetry monitoring, connected via REST endpoints.
- Developed back-end ETL functions using Python for AWS Lambda with API Gateway triggers.

Education

B.S. in Computer Science

Rutgers University–New Brunswick | 2013 – 2017 | New Brunswick, NJ

Certifications

- AWS Certified Machine Learning Specialty — Udemy
- Generative AI with Vertex AI — Google
- TensorFlow Developer Certificate — ZTM
- Full-Stack React & FastAPI Bootcamp — Udemy
- YOLOv8 & Object Detection — Udemy
- MLOps with Kubernetes & MLflow — Udemy

Publications

"Zigbee-Based Wireless Mesh Network for Smart Building Control" — Presented at 36th IEEEP Annual Conference – Gold Category Winner