

---

# HIROAKI OSHIMA

## Portfolio Website Link

[LinkedIn Link](#)

[Github Link](#)

Email: [kifa0422@gmail.com](mailto:kifa0422@gmail.com)

Phone: 253-275-8522

## Language

- **English:** Native/Bilingual
- **Japanese:** Native/Bilingual

## Skills:

- **Languages & Tools:**  
Python, SQL, Spark,  
Terraform, React.js, REST  
API
- **Cloud & Infrastructure:**  
AWS (Glue, ECS, EKS, etc),  
Docker, Kubernetes,  
Databricks
- **Big Data:** Data Lake, ETL,  
ELT, Data Pipeline, Data  
Warehouse, Scalability  
Programming
- **DevOps:** CI/CD, Monitoring  
(CloudWatch), Infrastructure  
as Code, Github Action
- **Machine Learning:**  
PyTorch, ML Development,  
MLOps LLMs, RAG, Data  
Visualization, GPU/CUDA

## Work Experience

### Data Engineer, Blue River Technology; Santa Clara, CA — July 2021-November 2023

- Built a scalable Data Lakehouse on AWS and Databricks to improve data pipeline replication time from 24 hours to 1 hour, enabling real-time analytics.
- Developed secure, high-volume ETL pipelines (>1TB/day) using AWS and Spark to integrate data across multiple organizations.
- Provided consultation and office hours to data scientists and robotics engineers, optimizing their ML data pipelines and unblocking critical workflows.
- Created a metadata catalog and storage layer for high-resolution imagery, accelerating data query speeds by up to 200x.
- Designed, deployed, and maintained dozens of Kubernetes-based data pipelines that migrated and processed terabyte-scale datasets across our internal data lake.

### Software Engineer Intern: FriendlyRobots.co — November 2019– February 2020

- Automated CI/CD pipelines for a self-driving vacuum robot, accelerating deployment cycles by 60%.
- Implemented CloudWatch alarms for performance, memory, and CPU usage monitoring.

## Certification

### AWS Certified Machine Learning – Associate

## Volunteer

### Machine Learning Engineer, DataKind, San Francisco CA, — September 2024 - Current

- Developed ML applications for nonprofits and county governments in CA and FL to support data-driven policymaking
- Built an interactive map-based application visualizing housing affordability and renters' burden across 3,000+ counties.
- Designed and deployed a RAG-enabled LLM chatbot with REST APIs to automate data visualization generation, reducing analysis time by 90%
- Deployed the entire application to AWS with scalable infrastructure.

## Education

### University of California Berkeley, Berkeley, CA — Bachelor's in Data Science 2019

## Projects

- MLOps-Driven ML Pipeline with Live Inference & Orchestration: Developed a production-ready stock price prediction system using Temporal Fusion Transformers, orchestrated with MLOps tools like MLflow, Airflow, Docker, and Kubernetes
  - Delta Lakehouse Architecture for Scalable Data Pipelines: Designed and implemented a Delta Lake-based data lakehouse to streamline scalable, cost-efficient ML data pipelines with enhanced reliability and performance.
-