KD Bartholomew

(408) 499-6130 bartholomewkd@gmail.com LinkedIn GitHub

AI/ML Engineer specializing in designing and deploying scalable machine learning systems. Experienced in end-to-end model development, MLOps, and cloud infrastructure to deliver production-ready, data-driven solutions focused on efficiency, interpretability, and impact.

Education

Master of Science in Data Science — **San Francisco, CA**
University of San Francisco — *Jul 2024 – Jun 2025*

Bachelor of Science in Data Science — **San Francisco, CA**
University of San Francisco — *Aug 2022 – Jun 2024*

Experience

Full Stack Data Scientist — **San Francisco, CA**
Queer Life Space — *Feb 2025 – Present*

- Built a custom **A/B testing framework** using **JavaScript** and **GCP**, applying sequential statistical experiments optimized for low-traffic, high-variance settings, **increasing donation conversions by 15%**.
- Engineered a **HIPAA-compliant ETL pipeline** in **Python** and **GCP**, anonymizing client data **across five heterogeneous sources** and **architecting a PostgreSQL database** with role-based access control and encryption.
- Built a **pricing sensitivity model** using regression techniques to identify the nonprofit's **minimum sustainable fee**, emphasizing **interpretability** through coefficient analysis and feature importance visualization for **20+ non-technical stakeholders**.
- Developed a containerized **React + FastAPI** dashboard on **GCP**, featuring statistical visualizations and model outputs enabling real-time pricing adjustments, **improving client accessibility by 35%** while maintaining financial stability. Implemented **CI/CD pipelines** with **GitHub Actions**.

Machine Learning Engineer — **San Francisco, CA**
The Nature Conservancy — *Sep 2024 – Present*

- Built a **globally scalable, distributed surface water detection pipeline** using NDWI, Otsu thresholding, and Canny edge detection, processing **3TB+ of satellite imagery weekly** across 5 pilot sites. Deployed on **AWS (S3, EC2, Lambda)** with autoscaling infrastructure and orchestrated via **Apache Airflow**.
- Replaced physical stream gauges with an automated satellite-based system, delivering **$225K in first-year savings** across 5 test sites with projected exponential savings as coverage scales.
- Developed **XGBoost forecasting models** with engineered seasonal and hydrological features to predict stream drying, **boosting accuracy from 82% to 93%** and outperforming SARIMAX baselines across multiple watersheds.

Machine Learning Engineer — **San Francisco, CA**
Schroeder Lab, University of San Francisco — *Aug 2022 – Jun 2024*

- Developed and deployed custom U-Net architecture using **PyTorch** with transfer learning from ResNet backbone, achieving **98% pixel-level accuracy** for automated cell segmentation across 50,000+ microscopy images.
- Optimized **multi-GPU training pipeline** with **PyTorch DDP**, **CUDA**, and **NCCL**, cutting training time by **60%** and enabling **on-the-fly data augmentation** that expanded dataset size **20x**.
- Implemented **comprehensive MLOps pipeline** featuring **Docker containerization**, **MLflow experiment tracking**, and **CI/CD deployment**, **reducing model inference time by 45%** via **CUDA optimization** and **TensorRT acceleration**.
- Designed **robust data preprocessing pipeline** applying **advanced augmentation techniques** to boost model robustness, **improving generalization by 15%** on unseen microscopy datasets and **saving $36K annually** in manual annotation costs.

Math Instructor — **San Jose, CA**
Independent — *Jan 2017 – Aug 2022*

- Taught AP and **college-level calculus, differential equations, linear algebra, probability, and statistics to 100+ students,** adapting lessons to diverse learning styles and emphasizing real-world applications to deepen understanding and critical thinking.

Projects

Multi-Agent RAG Weight Loss AI App

- Developed a **multi-agent system** for automated food logging and personalized coaching by integrating **CV and LLM APIs** with **RAG** using **FAISS** and **Supabase** for efficient **vector-based context retrieval** by specialist agents.
- Built a scalable, production-ready pipeline on **GCP** with **Docker**, **LangChain**, and a **PostgreSQL + Supabase** backend to support real-time agent orchestration and contextual knowledge integration.

Distributed Sentiment Analysis Pipeline

- Built a **distributed Apache Spark pipeline** on **GCP** and **MongoDB Atlas** processing **500K+ daily rows** of financial news and YouTube comments, applying statistical analysis to compare AlphaVantage sentiment scores with public opinion for "Magnificent 7" stocks using **Hugging Face transformer models**.

Metropolis-Hastings for Cryptography

- Engineered and optimized Metropolis-Hastings, a **probabilistic MCMC sampling algorithm**, for cryptographic applications **in a novel way, enabling convergence on large-scale cryptographic inputs that traditional methods couldn't handle.**

Publications & Awards

Founder Award – USF Innovation Summit — **Apr 2025**

- Built **AI Hire**, a full-stack, privacy-first AI job platform with **fine-tuned Sentence Transformers** and **edge-deployed** inbox scraping.
- Won **Founder Award** for innovation in **LLM tuning, edge AI, cloud architecture**, and product strategy.

Best Overall Nationally – American Statistical Association — **Dec 2017**

- Awarded Best Overall Nationally and **published** in AMSTATNEWS for statistical analysis and actionable insights on Seattle police data.

Technical Skills

ML Frameworks: PyTorch, TensorFlow, Keras, Hugging Face, CUDA
Programming: Python, SQL, NoSQL, Bash, Zsh
MLOps & DevOps: Docker, Kubernetes, CI/CD (GitHub Actions), MLflow, MetaFlow
Cloud Services: AWS (S3, EC2, Lambda), GCP (GKE)
Data Engineering: Apache Spark (PySpark), Airflow, PostgreSQL, MongoDB
Development: FastAPI, React, Flask, GitHub