Jiashu (Jerry) Huang

Evanston, IL | (+1) 773-410-6044 | zherrs102@gmail.com | Portfolio | GitHub | LinkedIn

Machine Learning and Data Science M.S. candidate focused on applied ML engineering. Built computer vision models and RAG systems, combining PyTorch model evaluation with AWS data pipelines and backend APIs.

EDUCATION

Northwestern University, *Evanston, IL* *Sep 2025 - Expected Dec 2026*
Master of Science | Major: Machine Learning and Data Science
University of California, Santa Barbara, *Santa Barbara, CA* *Sep 2021 - Jun 2025*
Bachelor of Science | Double Major: Statistics and Data Science; Applied Math GPA: 3.85/4.00

EXPERIENCE

Hax.ai (Startup), *Santa Barbara, CA | Full Stack Engineer* *Aug 2024 - Jan 2026*

- Built a workflow-scoped RAG system that decomposed user requests into research steps and retrieved evidence from user-selected webpages and PDFs using MiniLM embeddings, Pinecone vector search, PostgreSQL metadata, and LangGraph memory.
- Designed and deployed an asynchronous evidence-ingestion pipeline on AWS (API Gateway, Lambda, SQS, S3) with retries, dead-letter queues, and CloudWatch monitoring.
- Developed a Flask backend with 24 REST API endpoints integrating Amazon RDS through RDS Proxy and Pinecone; built React views to inspect recommendations, retrieved evidence, and memory traces.

Cheadle Center for Biodiversity and Ecological Restoration | *Research Assistant* *Jan 2024 - Jun 2025*

Advisor: *Professor Katja Seltmann, Cheadle Center for Biodiversity & Ecological Restoration*

- Processed 470+ bee-wing images across 20+ species; built a segmentation pipeline that removed ArUco-card backgrounds while preserving wing-scale information, achieving a 98% successful extraction rate.
- Trained and evaluated a PyTorch ResNet-50 species classifier, achieving 90% accuracy on 80 test images versus a 78% neural-network baseline.
- Extracted marginal-cell contour features using Fourier descriptors and PCA, then trained a gradient-boosting classifier that achieved 85% accuracy on 40 held-out images.

Xin Yi Tech, *Beijing, China* **|** *Machine Learning Engineer Intern* *Aug 2023 - Sep 2023*

- Built and deployed a face-detection and CNN-based classification pipeline for 1,000+ videos, improving video-origin identification accuracy from a 65% baseline to 78%.
- Integrated multiple Stable Diffusion models and a text classifier into a unified image-generation service, exposing model inference through REST APIs for product-team integration.

Correkt (Startup), *Santa Barbara, CA | Machine Learning Engineer* *Oct 2022 - July 2023*

- Built a ReWOO-style plan-then-execute fact-checking agent (GPT-3.5 planner and solver using Google, News and Wikipedia tools), with asynchronous tool calls to cut response time.
- Distilled GPT-3.5 judgments of whether a claim is worth fact-checking into a DistilBERT classifier, trained on 1,500 news headlines and reaching 80–84% agreement on a held-out set of about 225.

SKILLS

Languages: Python, C++, SQL, R, TypeScript, JavaScript
ML & AI: PyTorch, TensorFlow, scikit-learn, computer vision, NLP, RAG, LangGraph, LangChain, Airflow
Data & Statistics: NumPy, pandas, PostgreSQL, Pinecone, hypothesis testing, mixed-effects models
Cloud & Engineering: AWS (Lambda, SQS, S3, API Gateway, RDS), Flask, Docker, AWS CDK, REST APIs, React

PROJECTS

Structured Output Reliability Under Load

Built a reproducible Python CLI to evaluate JSON Schema validity, field correctness, latency, throughput, and quality goodput across LLM serving conditions, with regression gates and JSONL/Markdown reports. Ran 960 baseline and 768 stress-test attempts; traced 15 schema-valid but exact-value-incorrect outputs to synthetic padding, distinguishing formatting success from field-level correctness.

Dockside Evanston — Bike-Share Availability Risk

Built and deployed an AWS-hosted bike availability risk service with minute-level ingestion, daily model reassessment, and automatic retraining when quality gates fail; reduced Brier score by 7.4% versus a persistence baseline on a chronological holdout of 5,860 examples.

CDKCanvas: Visual AWS CDK Infrastructure Builder

Built a React and TypeScript visual editor for creating, connecting, validating, importing, and generating AWS CDK infrastructure through an interactive node-based canvas.