

Hanson Li

971-275-7450 | hanson.l@wustl.edu | linkedin.com/in/hanson-li1 | [GitHub](https://github.com) | San Jose, CA | U.S. Citizen

EDUCATION

Washington University in St. Louis St. Louis, MO
M.S. in Computer Science – GPA: 3.9/4.0 May 2026
B.S. in Computer Science and Economics – GPA: 3.6/4.0 Dec 2024

EXPERIENCE

Software Engineer Intern June 2025 – Aug 2025
CarMax Dallas, TX

- Developed a full stack Python application to configure A/B tests of vehicle recommendation models; utilized Azure App Service, CosmosDB, and integrated SSO authentication with Azure Entra ID
- Wrote unit tests for project reaching 80%+ code coverage; automated sandbox deployments with Azure CLI and Bicep
- Built embedding-based vehicle recommendation and semantic search proof of concepts leveraging a vector database paired with a vision-language model (CLIP) finetuned on 5k+ vehicle image-text pairs for vehicle embedding generation

Machine Learning Intern Jan 2025 – Apr 2025
Norfolk Southern Railway Atlanta, GA

- Trained a YOLO model using real and synthetic data for rail-car wheel defects, achieving 0.95+ mAP; detected two previously unknown defect classes via parallelized offline inference over 18M+ images
- Prototyped an object detection-based rail-car sill-step defect detector that was adopted over an under-performing classification baseline; achieved 0.8+ mAP by down-sampling no-defect images to address class imbalance
- Explored synthetic data generation by training FastGAN with 30 images to generate a novel case of rail-car wheel defect
- Built an active-learning data pipeline that identifies and uploads low confidence inference images to CVAT for labeling with SQLite, CVAT API, and Cron orchestration; achieved automation of 300+ image uploads to CVAT daily

Machine Learning Research Assistant May 2024 – Feb 2025
Washington University School of Medicine St. Louis, MO

- Trained 3D-ResNets and shallow ML models for radiology brain scan quality classification, achieving AUC of 0.92 and recall of 0.94. Incorporated radiomics and motion-artifact related features
- Built containerized OCR extractor for biomarkers across 800+ CTs; accelerated inference by 4× via thread pools

Software Engineer Intern May 2024 – July 2024
Alpha and Omega Semiconductor Limited San Jose, CA

- Developed and shipped 3 Java-based logistical dashboard web apps handling analytics of 100k+ rows of tabular data
- Built a reusable fuzzy string-matching utility for vendor name matching, enhancing downstream analytics accuracy and addressed lack of text-processing libraries in company's legacy Java environment

PROJECTS

Computer Vision Wood Staves Counter App | [Live Site](#) | [GitHub](#) | *PyTorch, FastAPI, ONNX, MLflow, Github Actions, OCI*

- Trained YOLO-based dense staves counting model, achieving 0.8+ mAP with domain-relevant pretraining; tracked experiments and registered models using MLflow hosted on Dagshub
- Developed a model inference web app hosted on OCI, achieving 2.5× inference speedup on CPU using ONNX Runtime; automated build and deployment with CI/CD workflow built with Github Actions

iOS Motion Analytics App | [GitHub](#) | *Swift, Python, PyTorch, PySpark, Airflow, Kafka, Azure EventHub, Snowflake*

- Developed an iOS App with CoreMotion and CoreLocation libraries to stream device motion and location data to Snowflake data warehouse via Azure Eventhub for data ingestion and Spark for data processing
- Trained a transformer encoder on user-collected data to detect risky driving behavior, achieving 97%+ validation accuracy; integrated the model into an Airflow-orchestrated email analytics pipeline

LoRA-tuned LLM with RAG for Rust Q&A | [GitHub](#) | *Python, LangChain, Hugging Face, ChromaDB, Llama.cpp*

- Fine-tuned Qwen3 0.6B with LoRA on 5k synthetic Rust Q&A pairs, increasing model conciseness. Developed RAG system with quantized model for grounded Q&A; achieved 88% win rate vs. baseline using a judge LLM for evaluation

TECHNICAL SKILLS

Languages: Python, Java, SQL, JavaScript, Bash, Rust

ML Tools: PyTorch, TensorFlow, LangChain, Hugging Face, Scikit-Learn, OpenCV, ChromaDB, ONNX, Docker, MLflow

Full Stack/Data Tools: Node.js, Flask, FastAPI, Django, MongoDB, MySQL, Pandas, Spark, Airflow, Kafka, AWS, Azure