

Yichen Wu

+1 778-938-2527 | yichenwujob@gmail.com | [LinkedIn](#) | Vancouver, BC

EXPERIENCE

DaoAI Robotics Inc.

Vancouver, BC

Software Engineer — Production AI Platform

May 2025 – Present

Video Analytics Platform — Python · FastAPI · OpenCV · PostgreSQL · Redis · Docker · Kubernetes

- Designed and shipped **8+ reusable AI workflow modules** for detection, segmentation, OCR, and visual analysis — Pydantic configuration schemas, REST API integration, and automated worker deployment on Kubernetes.
- Built a **workflow validation framework** with FastAPI and Pydantic, implementing dependency-graph traversal and standardized error handling across model, logic, visualization, and alert nodes.
- Built and productionized a **real-time multi-object tracking service** in Python with NumPy and SciPy for dozens of concurrent live video streams — Redis-backed track state, configurable lifecycle, and confidence thresholds.
- **Optimized real-time pipeline performance** by profiling hot paths and replacing per-frame Redis and shared-memory access with **in-memory TTL caches**, eliminating network round trips and serialization overhead.
- Developed **15+ stateful real-time analytics features** with OpenCV and Redis — speed estimation, vehicle counting, congestion metrics, incident detection, and alerting — running 24/7 in production on Kubernetes.
- Built an **automated data pipeline** in Python that filters and samples production detection results and uploads annotated images via REST and FTP to the training platform, enabling iterative model retraining.

Generative AI Production Platform — Python · TypeScript · Next.js · FastAPI · PostgreSQL · Celery · Redis

- Co-built a **full-stack generative AI platform from the ground up** with Next.js, TypeScript, FastAPI, and PostgreSQL, integrating third-party REST APIs from OpenAI, Gemini, Kling, Seedance, Suno, and Topaz.
- Built **resilient asynchronous job processing** with Celery and Redis — WebSocket status updates, polling recovery, jobs persisting across page refreshes, automatic failure refunds, and centralized error handling.
- Integrated **third-party video-generation APIs** (Kling, Seedance) end to end — request validation, content-compliance pre-checks in Celery workers, automatic retry and failover, and provider-specific payload adapters.
- Designed and built **REST APIs with FastAPI, SQLAlchemy, and PostgreSQL** powering multimodal chat, storyboarding, media editing, and asset management, connected to React/Next.js frontends, background jobs, and object storage.
- Reduced LLM agent prompt overhead **from ~3,500 to ~800 tokens** by modularizing system prompts, filtering tools by entity type, and scoping context to the active entity and adjacent panels.
- Established **automated testing and CI/CD quality practices** — pytest integration tests against real databases, Playwright end-to-end suites, and AI-agent code review enforcing testing and security standards in the pipeline.

SageArk LLC

Chicago, IL

Full-Stack Developer

Feb 2024 – Apr 2025

- Built and deployed the company website with Next.js and AWS S3, CloudFront, and Route 53, including custom-domain setup and CI/CD.
- Developed a job-recommendation model on **AWS SageMaker** ranking roles from student coursework and experience — feature engineering, model training, and an end-to-end demo application.

TECHNICAL SKILLS

AI / ML / CV: LLM/VLM API integration, agent & tool orchestration, prompt engineering, PyTorch, OpenCV, multi-object tracking, OCR

Backend & Distributed Systems: FastAPI, REST APIs, asyncio, Celery, SQLAlchemy, Alembic, Pydantic, Redis, PostgreSQL, WebSockets, MinIO, pytest

Infrastructure & DevOps: Docker, Kubernetes, Helm, AWS (S3, CloudFront, Route 53, SageMaker), CI/CD

Languages: Python, TypeScript, JavaScript, SQL, C/C++, Java

Frontend & Media: React, Next.js, TypeScript, React Query, Tailwind CSS, Playwright, FFmpeg

EDUCATION

Northwestern University, MS in Computer Engineering

Aug 2022 – Dec 2023

University of Illinois Urbana-Champaign, BS in Computer Engineering · *Dean's List*

Aug 2018 – May 2022

Certifications: AWS Certified Solutions Architect — Associate; AWS Certified Cloud Practitioner.