

Ken Wu

kenwu.is-a.dev | GitHub: [KenWuqianghao](#) | LinkedIn: [kenwu](#) | ken.wu@uwaterloo.ca | 437-599-3179

SKILLS

Languages: Python, SQL, C++, C, TypeScript, JavaScript, R, Racket, Bash

ML & Data: PyTorch, CatBoost, TensorFlow, PySpark, Pandas, NumPy, GeoPandas, spaCy, MLflow, Hive, CUDA

Frameworks: FastAPI, Flask, LangChain, LangGraph, React, Next.js, Tailwind

Cloud & Databases: AWS, Azure, Docker, Git, PostgreSQL, MongoDB, Supabase, ChromaDB

EXPERIENCE

Mercor

Member of Technical Staff

Incoming
Enterprise AI | San Francisco, United States

Lyft

May 2026 – Aug 2026

Machine Learning Software Engineer Intern | Python, CatBoost, SQL, GeoPandas Bailout Model | Toronto, Canada

- Doubled the bailout model's cancel-detection recall (+**34.2pp** at 2.5% FPR) via **17** new route/GPS features
- Grew Canada bailout coverage from **1.4%** to **100%** of rides via census backfill for **2,964** geohashes
- Root-caused a defect in Lyft's bailout model flagging **76.6%** of rides; shipped automated per-cluster thresholds
- Shipped shadow inference and holdout capture into the bailout service with **zero** live decision impact

Nokia

Jul 2025 – Dec 2025

Software Engineer Intern | Python, PyTorch, Unsloth, Transformer

5G Network Agent | Ottawa, Canada

- Built a LoRA fine-tuning pipeline turning Qwen into a Camunda BPMN XML generator, with GGUF quantization
- Exposed overfitting in NCSC-only training data with an evaluation framework benchmarking vs. base models
- Unblocked training runs by resolving GPU memory and Triton kernel failures; shipped reusable training tooling

TD Bank

Jun 2025 – Aug 2025

Data Scientist Intern | Python, SQL, PySpark, Pandas, Databricks

Insurance Analysis | Toronto, Canada

- Replatformed pipelines for **1.5M+** rows of data and reproduced a 154,340 row deliverable with **100%** parity
- Automated **30+ minutes** of manual ingestion per cycle, cutting QA time by **80%** via a row-level parity harness
- Saved **2–4 hours/week** of manual execution by migrating the on-prem pipeline to Azure with automated runs

August

Sep 2024 – Dec 2024

Software Engineer Intern | Python, LangGraph, FastAPI, AWS

LLM Agent | New York, United States

- Handled **2000+ requests/min** by deploying **10+** live **API endpoints** using FastAPI, AWS, and Supabase
- Streamlined evaluation cycles for **15+ LLM Agents** via a round-robin multi-agent and ELO scoring framework
- Architected a multi-agent RAG pipeline end-to-end, powered by LLM-as-Judge evaluation strategies

Health Canada

Apr 2024 – Aug 2024

Machine Learning Engineer Intern | Python, Azure, LangChain, Streamlit

Document QA | Ottawa, Canada

- Built a document QA system using **Llama3 7B** and **ChromaDB** for OECD report search and summarization
- Increased response and semantic accuracy by **~20%** using **query transformation** and **contextual memory**

Respan AI (Y-Combinator W2024)

Mar 2023 – Jun 2023

Software Engineer Intern | Python, spaCy, SQLite

Resume Parsing | New York, United States

- Parsed **1000+ resumes** with a spaCy-based NER pipeline to extract structured and relevant recruiter data
- Reduced response delay by **98%** through integration of SQLite-based result caching into the parsing engine

Intapp (Formerly delphai)

Jul 2022 – Sep 2022

Machine Learning Engineer Intern | Python, spaCy, W&B, BS4

Entity Recognition | Berlin, Germany

- Boosted recall by **20%** through improved entity labeling workflows and language-specific model training sets
- Achieved **70% recall** by fine-tuning spaCy models and optimizing hyperparameters via W&B on Azure

EDUCATION

University of Waterloo

Sep 2022 – Apr 2026

Bachelor of Computer Science, Business Specialization, Philosophy Minor

Waterloo, Canada

- Neo Scholar Finalist | **5x** Hackathon Winner | URA @ UWaterloo: Symbolic Regression GPT | Researcher @ Lancaster University: Clustering Optimization | Open Source Dev @ Fermilab: Agentic Lagrangian Extraction