Cynthia Zhang

cczhang@uwaterloo.ca linkedin.com github.com cynthiacxzhang.me

EDUCATION

University of Waterloo Sep. 2023 – Present

Bachelor of Computer Engineering, Honours | Minors: Cognitive Science, Psychology *Waterloo, ON*

- **Relevant Coursework:** Data Structures & Algorithms, Object Oriented Programming, Real-Time Operating Systems

EXPERIENCE

RBC Borealis Sept 2025 – Dec 2025

Software Engineering Intern *Toronto, ON*

- Shipped the first agent-to-data integration service to **17,000+** users across RBC, exposing core financial assets to executive-level business initiatives generating **$6.8B+** in net revenue.
- Designed and deployed end-to-end real-time inference pipelines via RAG-based LLM grounding (FastAPI + LangChain), reducing hallucination rate by **99.9%** and improving autonomous reasoning accuracy by **90%** in production.
- Deployed Kafka-ingested, Spark-based anomaly detection (Scala/SQL on Hadoop/YARN) in Lumina Platform ETL pipelines, boosting dataset coverage by **40%** and reducing data loss by **65%**.

University of California, Berkeley May 2025 – Present

Machine Learning Engineer *San Francisco, CA*

- Designed and deployed a content understanding pipeline (BERT fine-tuning, PyTorch + TensorFlow) for multi-class text classification with domain adaptation via masked language modeling, achieving **89%** accuracy.
- Optimized distributed training workloads on HPC infrastructure (SLURM-scheduled GPU cluster, parallel filesystem I/O) via batching, vectorization, and concurrent processing, reducing end-to-end pipeline runtime by **60%**.
- Built an unsupervised textual categorization system, applying latent-dirichlet-allocation (Gensim + NLTK) to extract **12+** document-topic distributions from **30K+** articles across **50+ years** of textual corpora.

Environment and Climate Change Canada Jan 2025 – Apr 2025

Software Developer Intern *Toronto, ON*

- Designed and implemented an unsupervised anomaly detection pipeline using Isolation Forest (Scikit-learn, NumPy, Pandas) over **2TB+** of METAR aviation data reported globally by **720+** commercial airlines.
- Designed a parameterized batch data retrieval pipeline for Canada's national climate reporting network, enabling downstream risk modeling across **25+** CMIP-6 global multi-model ensemble datasets.
- Engineered an automated validation system (pytest, Selenium) for early-stage data ingestion, applying schema constraints and integrity checks across **250K+** storm instances, improving data quality by **75%**.

PROJECTS

TRuST Network Research Assistant | *University of Waterloo* 2025 - Present

- Investigated zero-shot generalization failure and subgroup robustness in facial recognition models, analyzing how inductive reasoning and over-generalization produce systematic performance degradation on statistical minority groups.
- Designed a knowledge-grounded inference system using retrieval-augmented generation (Claude, MCP, SSE Starlette) to support AI literacy benchmarking and model behavior evaluation for policymakers.

Range Equity | *Cloudflare, Web Workers, Express, Vanilla JS, Claude API* 2026

- Built a Texas Hold'em equity calculator running **50K** Monte Carlo simulations in a web worker, with a hand evaluator using integer card encoding and precomputed combination indices for sub-second 7-card evaluation.
- Integrated Claude API for plain-English range translation and GTO-aware analysis on hand, board, and equity state.

cynthiacxzhang.me | *Vercel, Supabase, PostgreSQL, Notion API, JavaScript, HTML/CSS* 2025

- Built a personal portfolio site on a three-layer backend architecture – Supabase (PostgreSQL + RLS) REST APIs for structured data access, Vercel serverless functions proxying protected API requests, and Notion as a headless CMS.
- Enforced zero-trust client architecture – RLS, JWT for access control, and serverless reverse proxies for token injection.

SKILLS

Languages: Python, C/C++, SQL, Scala, Java, Javascript, Verilog, VHDL, Bash/Shell
Frameworks: Spark, Hadoop, Hive, Kafka, PyTorch, TensorFlow, Scikit-learn, LangChain, FastAPI, NLTK, Gensim
Tools: Docker, Kubernetes, AWS, GCP, Snowflake, PostgreSQL, NoSQL, Cloudflare, NumPy, Pandas, Linux, Git
Concepts: supervised/unsupervised learning, RAG, LLM infrastructure, distributed computing, transformer fine-tuning, feature engineering, batch/real-time inference, zero-shot/few-shot generalization, multimodal content evaluation