

Yucheng Yang

Vancouver, BC | ☎ 672-339-6253 | ✉ ycyang.dev@gmail.com | [in](#) LinkedIn

Education

Master of Science in Information Systems (GPA: 3.8/4.0)

Sep. 2024 - Dec. 2025

Northeastern University, Vancouver, BC

Relevant Courses: Self-improving AI, Data Science, Application Engineering

Bachelor's Degree in Digital Media Technology, School of Software

Sep. 2014 - Jun. 2018

Yunnan University, China

Relevant Courses: Data Structures, Networking, Database, Software Engineering

Technical Skills

Machine Learning:	PyTorch, TensorFlow, Scikit-learn, Hugging Face (Transformers, Datasets), NumPy, Pandas
Tools & Infra:	WandB, Git, Docker, AWS (EC2, S3), Linux, CI/CD Pipelines
Programming:	Python, SQL (PostgreSQL, DuckDB, SQLite), Java, Kotlin, C#, TypeScript
Web/Backend:	Spring Boot, ASP.NET, REST APIs, Thrift, React, TailwindCSS

Work and Projects Experience

Open-Source Community Toxicity Analysis

Mar. 2025 - Oct. 2025

Developed a large-scale NLP pipeline to model toxicity and patterns in open-source ecosystems

- Built a high-performance data pipeline using **DuckDB** to ingest and transform **2TB+** of raw GitHub Archive data, extracting a curated **300M-comment** dataset
- Fine-tuned and evaluated LLMs (**BERT**, **RoBERTa**, **DistilBERT**) via **Hugging Face**; improved **AUC-ROC from 0.76 to 0.93** through domain-specific adaptation on GitHub code review datasets
- Leveraged **WandB** for experiment tracking and hyperparameter optimization, ensuring model reproducibility
- Analyzed 6-year toxicity dynamics within domains (ML, DevOps, Game, etc.) using time-series analysis, identifying the impact of external events (e.g., COVID-19) on community toxicity

Autonomous Vehicle Behavior Clustering

Sep. 2024 - Feb. 2025

Engineered an unsupervised pipeline on Argoverse 2 to model multi-agent interaction patterns

- Mined **1,322 scenarios** from a 250k-sequence Argoverse 2 database using SQLite and Python, retaining key interactions between autonomous vehicles and motorcycles
- Developed a feature extraction module using **NumPy** and **Pandas** to derive spatial-temporal metrics (e.g., yaw rate, lateral/longitudinal distance, velocity diff), quantifying the dynamic relative motion of interacting agents
- Categorized autonomous vehicle behaviors by implementing **K-Means** and **DBSCAN** on Savitzky-Golay smoothed trajectories, utilizing Silhouette Scores to ensure optimal cluster separation
- Visualized behavioral clusters using **static trajectory maps** and **animated GIFs**, illustrating temporal cluster transitions to identify key interaction patterns

Quantum Key Issuance Project (at Reborn Quantum Technology Co., Ltd.)

Jul. 2019 - Oct. 2021

Architected a hardware–software system for secure quantum random number distribution

- Developed a data ingestion layer to interface with Quantum Random Number Generators (QRNG), managing secure data streams for centralized pools
- Optimized a custom ECC-based cryptographic model to secure random number storage and ensure data integrity throughout the processing pipeline
- Implemented a low-latency service layer using **Apache Thrift (RPC)** to facilitate efficient communication between centralized QRNG and remote issuance terminals
- Engineered a JavaFX-based management system integrating **JNI (C/Java)** for hardware initialization and lifecycle management

Quantum-Secured Office Network Project (at Reborn Quantum Technology Co., Ltd.)

Jul. 2018 - Mar. 2019

Engineered an enterprise communication system with quantum-encrypted networking and conferencing

- Integrated **quantum encryption protocols** into OpenVPN and VoIP architectures, enabling hardware-level security for private networking and voice communication
- Optimized **Nginx reverse proxy** and server-side configurations for the BigBlueButton stack, ensuring secure, low-latency distribution of real-time media streams
- Designed a cross-platform identity **authentication** module using quantum USB keys to enforce secure user access for Android and web-based enterprise systems