

Zixuan Kang

778-682-7953 | zixuank0287@outlook.com | [LinkedIn](#)

EDUCATION

Columbia

Masters of Computer Science

New York, United States

Expected to start Sept 2026

University of British Columbia

Bachelor of Computer Science

Vancouver, BC

Sep. 2020 – Expected May 2026

- GPA: 3.76 / 4.0
- Relevant Coursework: Software Engineering, Computer Hardware and Operating Systems, Intermediate Algorithm Design, Advanced Algorithms Design, Introduction to Relational Databases, Advanced Relational Databases, Machine Learning and Data Mining, Advanced Machine Learning, Parallel Computation

WORK EXPERIENCE

Data Science Intern, Siemens, Beijing China | *Python, PySpark, Pandas, Numpy, ML, AWS* May - September 2025

- Implemented business project name de-duplication algorithm leveraging **cosine similarity**, **fuzzy matching techniques**, and domain-specific knowledge integration, eliminating manual work completely
- Improved F1 score by **15%** and accuracy by **12%** by building a business classification system using **transformer fine-tuning**, **LightGBM**, **focal loss**, and **synthetic sampling** on business scope and industry data
- Built **scalable** data processing pipeline with **Pyspark** and **NumPy** to consolidate **50M+** customer pricing records from service requests and compute annual product- and customer-level price change metrics
- Developed anomaly detection system using **Variational Autoencoders** in Keras for customer pricing records, executing comprehensive data **preprocessing** and **sampling-based validation** to assess model performance
- Built automated PDF processing pipeline using **Tesseract OCR** with **text-position extraction**, **spatial alignment** and **binary search** to locate contract numbers and tabular content, eliminating manual data entry

Co-Founder, Ace Automations, Remote | *Rust, parallelization, TCP/UDP, security, SQL* Dec 2023 - Jan 2025

- Spearheaded development of high-performance trading software on decentralized exchanges
- Leveraged thread-pool to develop a **multithreaded transactional handler**, processing buy/sell requests in **under 100ms**, achieving up to **20x** faster performance compared to manual executions
- Utilized **Tokio Console** to identify and debug **software hangs**, **deadlocks**, and **performance bottlenecks** by monitoring threads, task lifecycles, and slow execution speeds
- Designed scalable messaging protocol using **gRPC** and **custom protocol buffers** to stream across distributed clients
- Implemented statistical detection algorithm to identify anomalies in asset launch patterns, enhancing clients' safety
- Designed AWS workflow utilizing **Lambda**, **ECS**, and **VPC** to handle high loads and ensure seamless performance

Software Developer Intern, LMI Technologies, Burnaby | *C++/C, Python* Jan – September 2023

- Developed **C++ multi-threaded** solution reducing sensor start time by **40%**
- Reduced binary size of build artifacts by **20MB** through resolution of technical debt
- Maintained build server stability by eliminating **intermittent** build failures in multi-threaded environments
- Improved **Python** scripts to eliminate upgrade hangs, enhancing software reliability
- Automated testing in **C++**, creating mocks that increased extensibility and confidence in code

RESEARCH EXPERIENCE

Railway AI Video Surveillance Researcher | *Supervised by Prof. Zhang, Peking* June - November 2025

- Fine-tuned **multimodal LLMs** with **LoRA adapters** and **sliding-window temporal sampling**, detecting mudslides from real-time railway video and sensor streams
- Engineered **environmental condition augmentations**(rain intensity, fog density, debris), cutting false positives and raising model recall under adverse weather

Algoverse Researcher | *Mentored by Shreyas from Amazon, Kevin Zhu from UC-Berkeley* June - November 2025

- Extended **Uncertainty of Thoughts (UoT)** with **Bayesian prior estimation** and **multibranching**, generalizing past binary entropy-based splitting
- Implemented **likelihood caching** over sentence-clustered questions, cutting redundant model calls
- Accepted to **Agentic AI in the Wild** workshop, **ICLR 2026**

SKILLS

Languages: C++, Rust, Python, TypeScript/JavaScript, Java, SQL/NoSQL, x86-64, CUDA, Erlang, PHP
Machine Learning: Transformers, LoRA fine-tuning, VAEs, CNNs, SGD, MLE/MAP estimation, PCA/t-SNE
Frameworks: PyTorch, TensorFlow, Scikit-learn, Pandas, NumPy, PySpark, OpenCV, React Native, Next.js
Infrastructure: Kafka, Docker, Terraform, AWS (Lambda, ECS, VPC), PostgreSQL, Redis, gRPC/protobuf
AI Tooling: Claude Code, Codex, Cursor, MCP servers, agent skills, prompt & context engineering
Interests: LLM reasoning & alignment, blockchain & DeFi, fintech, vision research, data pipelines