Yuejia (Roger) Guo

Waterloo, ON — (437) 818-9423 — y52guo@uwaterloo.ca — yuejiaguo.com

SUMMARY OF QUALIFICATIONS

- Computer Science and Data Science student focused on backend systems and scalable software engineering
- Experienced in building data pipelines, backend services, and API-driven applications using Python, C++, SQL, and JavaScript
- Familiar with distributed system concepts including caching (Redis), concurrency control, and client-server architecture
- Strong foundation in algorithms, operating systems, and database systems with hands-on project experience

EXPERIENCE

Software Engineering Intern *May 2026 – Aug 2026, Remote*

Capsule AI

- Engineered a multi-layer NMPA/CDE regulatory ingestion pipeline across four NestJS modules and nine Playwright routes, resolving anti-bot defenses, JavaScript-rendered content, nested result pages, popup and PDF documents, inconsistent metadata, bilingual entity matching, and extraction fallback behavior
- Developed Publication Watch, a seven-source scientific-literature tracking system with fault-isolated parallel retrieval, cross-source deduplication, and a Gemini-powered natural-language query compiler with deterministic fallback behavior
- Refactored company onboarding from one-off, company-specific scraper development into an AI-assisted batch workflow, increasing effective throughput by 20x+ and enabling the validation and onboarding of 50+ new companies per day
- Applied and hardened reusable scraper architectures across 500+ distinct companies, adapting extraction, pagination, validation, and failure-handling strategies to heterogeneous websites and data platforms spanning press releases, investor events, and regulatory records

Data Analytics Coop (Portfolio+) *Jan 2025 – Aug 2025, Toronto, ON*

Sanofi Digital Data & AI

- Rebuilt a fragmented reporting system into a unified data platform by redesigning data models and backend KPI computation logic, improving system consistency across multiple business units
- Developed scalable ETL pipelines using Snowflake and SQL to automate data ingestion and transformation, reducing manual processing effort by 40%
- Optimized query performance and data access patterns, reducing dashboard load time from ~5–10 minutes to under 2 minutes (70%+ improvement)
- Implemented data validation and processing logic to ensure consistency and reliability across large-scale datasets (100k+ records)
- Prototyped LLM-powered querying interface to enable natural language access to structured data, improving user accessibility for non-technical stakeholders

Business Planning and Operations Intern *Jun 2022 – Aug 2022, Remote*

Visa Inc.

- Built a Python-based data processing pipeline integrating Google Forms API and automated preprocessing scripts to handle survey data collection
- Processed and analyzed 2,000+ records, identifying key behavioral patterns and improving reporting accuracy by 25%
- Designed interactive Tableau dashboards to support business decision-making and improve insight delivery efficiency

PROJECTS

CoopPilot — AI Job Matching Platform *Jan 2026 – Jun 2026*

- Designed and built a full-stack web application to process and analyze job postings with over 1,000+ listings
- Implemented backend pipelines to parse unstructured job descriptions and extract structured features such as skills, requirements, and experience levels
- Developed scoring algorithm to evaluate job relevance using contextual signals beyond keyword matching, improving matching precision by 30%
- Built API endpoints and integrated frontend (Next.js) to support real-time filtering and recommendation workflows

Backend Service and System Design Practice *Jan 2026 – Jun 2026*

- Implemented RESTful API services and simulated high-concurrency scenarios to understand backend system behavior under load
- Explored caching strategies using Redis (cache-aside pattern) to reduce database access latency and improve response time
- Studied distributed locking mechanisms and rate limiting (token bucket) to handle concurrent requests safely
- Built simple TCP-based client-server programs in C++ to understand connection handling and data transmission

LRU Cache Implementation (C++ / System Design) *Jan 2024 – Apr 2024*

- Implemented an LRU cache with O(1) time complexity using hash map and doubly linked list
- Simulated cache eviction policies similar to Redis memory management strategies
- Improved cache hit rate by optimizing access patterns in test scenarios

URL Shortener Service (Backend System Design) *Feb 2025 – May 2025*

- Designed and implemented a URL shortening service supporting high read/write traffic with unique ID generation
- Built RESTful APIs for URL encoding/decoding with persistent storage using MySQL
- Optimized read performance using Redis caching, reducing average response latency by 50%
- Handled collision and consistency issues through hash-based ID generation and validation logic

Sanofi Internal Hackathon — 1st Place *Sanofi Digital Data & AI*

- Developed a prototype system to identify and classify enterprise data quality issues using rule-based logic
- Generated structured outputs including issue categorization and recommended resolution workflows

Tetris Game (C++, ADT-Based) *Sep 2024 – Dec 2024*

- Built a complete Tetris game in C++ implementing rotation logic, collision detection, and scoring system
- Debugged memory and pointer issues using GDB and Valgrind, improving program stability and preventing memory leaks

EDUCATION

Bachelor of Computer Science, Data Science — University of Waterloo *Sep 2023 – Present*

- GPA: 4.0 / 4.0

SKILLS

- **Programming:** Python, TypeScript, JavaScript, C++, SQL
- **Backend and Web:** Node.js, NestJS, Next.js, React, REST APIs, Data Pipelines
- **Automation and Systems:** Playwright, Browser Automation, Web Scraping, Git, Linux, Redis, Docker
- **Data:** Snowflake, MySQL, Power BI, Tableau, Pandas, ETL, Data Modeling
- **AI:** LLM Applications, Gemini, Prompt Engineering, Structured Outputs, AI-Agent Workflows