

Fred Liu

<https://fredliu.netlify.app>

Email : fredliu876@gmail.com

Github: [FredLiu876](#)

Mobile : 613-698-4894

LinkedIn: [fredliu1](#)

EDUCATION

University of Waterloo

Ontario, Canada

- Candidate for Bachelors in Computer Science; GPA: 3.98/4.00

2020 - 2025

Relevant Coursework: Data Structures & Algorithms, Operating Systems, Databases, Compilers, ML

EXPERIENCE

Cohere

Toronto, ON

- *Software Engineer Intern*

Jan 2024 - Apr 2024

- Processed 278TB of big data using distributed computing with Spark to create a pre-training code dataset that outperforms Starcoder and improved the LLM's coding capabilities by 4% on average across all code benchmarks
- Researched minhash LSH algorithm for near deduplication of text and designed new sub-quadratic time algorithm within Spark to reduce its false positive rate within clusters from 45% to 1%
- Developed data visualization internal tool integrated with pipelines to automatically build relevant visualizations at each data processing step, becoming the most contributed repo on the team for catching mistakes at every step

Super.com

Toronto, ON

- *Data Engineer Intern*

May 2023 - Aug 2023

- Orchestrated seamless, safe upgrade of highly integrated tool DBT from 1.0 to 1.6 by coordinating the data team of 6 and devising a quick failsafe rollback plan, successfully modernizing data processing options.
- Architected and implemented scalable data pipeline factory integrated with SmartSheet, improving collaboration with suppliers and reducing time spent updating sheet by 20 hours per week.
- Researched and executed cost-saving initiatives in the data stack, achieving a reduction of over \$50,000 annually
 - * Developed in-house data syncs connecting several data sources to Snowflake, saving \$16000 on SAAS costs
 - * Implemented lifecycle rules safely to reduce storage by 50%

Titanium Agency

Waterloo, ON

- *Full Stack Developer Intern*

Sep 2022 - Dec 2022

- Minimized data fetch size from database by translating complex backend data transformation logic to SQL and limiting results by implementing pagination, accelerating frontend dashboard operations by 5 times
- Developed loading bar component using WebSocket connection for realtime progress between frontend and backend, enhancing user experience for timely operations
- Built new webpage using Facebook Business API for bulk ad cloning, reducing cloning time by up to 70%

Wish

San Francisco, CA

- *Data Scientist Intern*

Jan 2022 - Apr 2022

- Engineered end-to-end ETL data pipeline for automated data visualization of app notification metrics on Tableau, utilizing attribution data models to measure engagement, increasing engagement by 2 times
- Conducted analysis of anomalies in user retention, discovering a crucial bias under advertising dimensions. This proved there was an actual 7% increase in the north star metric, improving employee morale

PROJECTS

- **Personal AI Agent:** Developed 2 different AI agents that are integrated to the shell, local file system, and the web to automate tasks such as a typical git workflow or contribute to a local code repo
- **FredZip:** 55% compression ratio on text files by implementing and combining Huffman and LZ77 algorithms in C++

AWARDS

- **Best-In-Class at Super.com:** \$2000 awarded for exceptional performance among cohort of 66 interns

SKILLS

- **Languages:** Python, SQL, C++, C, Javascript, HTML & CSS, C#, Java, R
- **Technologies:** Snowflake, Spark, Airflow, DBT, Flask, NodeJS, Vue, ASP.NET, MongoDB, MySQL
- **Tools:** GCP, AWS, Github Actions, Docker, Git, Linux