Leo Liu

(778)990-6015 | leo@liuwork.ca | linkedin.com/in/siqiliu- | github.com/leowrites | liuwork.ca

Education

BSc Computer Science, University of Toronto Expected May 2026
GPA: 3.95/4.0 Dean's List

- **Courses**: Compilers & Interpreters, Research in ML Compiler with MLIR, Computer Architecture, Operating Systems, Parallel Programming, Data Structures and Algorithms, Computer Networks, Intro to ML, Databases

Work Experience

ParaMathics Toronto, ON
Undergraduate Researcher May 2025 – Present

- Contributed to 2:4 sparsity integration for OpenAI's Triton compiler, developing mixed-sparsity kernels and achieving up to 30% throughput at various sparsity levels compared to cuSPARSELt
- Extended Triton's autotuner to support preprocessing by modifying kernel argument handling and caching mechanisms, enabling optimal tile size discovery across varying sparsity levels and matrix dimensions
- Achieved up to 8% performance gains by optimizing kernels with autotuned configuration to address issues related to L2 cache utilization, investigated tail effect and load imbalance

Mozilla Toronto, ON
Software Engineer Intern May 2025 – Present

- Engineered an Enhanced Tracking Protection flexibility option impacting over 1.5 million users, adding an reduce production feature to mitigate over 1,000+ site-breaking issues while maintaining privacy protection
- Responsible for notifying existing users of the new felxibility option by working with cross-functional teams(Messaging Systems, UX and content) and implementing an infobar prompt
- Accelerated diagnositc of web compatibility related issues by developing a devtool panel for QAs, introducing a new workflow increasing efficiency

Konrad Toronto, ON
Software Development Intern May 2024 – Aug 2024

- Owned the backend development of an internal employee gaming leaderboard, designing and implementing the core system architecture and database schema
- Developed a daily data ingestion pipeline integrating 4 gaming platform APIs to automatically fetch and aggregate game statistics for over 500 employees, enabling real-time leaderboard functionality
- Optimized API efficiency by implementing GraphQL resolvers with Apollo Express, reducing complex nested query latency by 75% by eliminating request waterfalls.

Relevant Projects

MiniC Compiler Development | *C++, LLVM, Compiler Design, Software Engineering*

- Developed a complete C compiler from scratch using C++ and LLVM framework, implementing all phases of compilation including lexical analysis, parsing, semantic analysis, and code generation for primitive types and arrays
- Built and tested an alloca2reg promotion optimization pass, reducing memory allocations and achieving a 1.7x performance improvement on benchmarking programs
- Designed an optimization passe to eliminate redundant global loads by utilizing the LLVM framework to perform loop analysis and code re-write, achieving up to 5x speedup in sample programs

Particle Simulation Parallelization & Optimization | *C, OpenMP, OpenMPI*

- Experimented with different optimization techniques such as binning, static arrays, and bucket sort to decrease sequential runtime for simulating 160,000 particles on SciNet, achieving a 2.94x increase relative to baseline
- Leveraged OpenMP to parallize 90% of the sequential code using static decomposition and uniform partitioning of data, further reducing the runtime to 2 seconds and achieving a performance improvement of 9x

Technical Skills

Languages: Python, TypeScript, C, C++, SQL
Systems & Tools: Git, Linux/Unix, Slurm, Vim, LLVM/MLIR, Nsight Compute