# Leo Liu

**Canadian Citizen**

(778)990-6015 | leo@liuwork.ca | linkedin.com/in/leolinked | github.com/leowrites | cs.toronto.edu/~lliu

## Education

**BSc Computer Science, University of Toronto** Sep 2021 – 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
- **Interests**: High Performance Computing, Computer Systems, Backend
- **Activities**: President (2023) at UofT Blueprint, Student Developer (2022, 2024) at UofT Department of CS

## Work Experience

**ParaMathics Lab (Compiler for Tiled-Level Sparse Matrix Multiply)** Toronto, ON

Undergraduate Kernel Engineer Researcher May 2025 – Present

- Contributed to supporting 2:4 sparsity in OpenAI's Triton compiler for NVIDIA tensor cores by developing mixed-sparsity kernels, achieving up to a 30% increase in throughput compared to cuSPARSELt
- Enabled optimal tile size discovery across varying sparsity levels and matrix dimensions by extending Triton's autotuner to support matrix preprocessing and tensor decomposition
- Optimized kernels to achieve up to 8% performance gains with autotuned configurations, addressing issues related to L2 cache utilization, tail effects, and load imbalance

**Mozilla (Firefox Privacy)** Toronto, ON

Software Engineer Intern May 2025 – Present

- Delivered an Enhanced Tracking Protection (ETP) customization feature to over 1.5M+ users, giving users the option to mitigate over 1000 site-breaking issues in ETP-Strict mode, while preserving privacy protections
- Led the design and rollout of an interactive infobar, driving high user engagement and adoption through cross-functional collaboration with Messaging Systems and UX teams
- Boosted bug diagnostic efficiency by building a custom developer panel for QA teams, streamlining debugging process

**Konrad** Toronto, ON

Software Development Intern May 2024 – Aug 2024

- Owned the backend development of an internal employee gaming leaderboard using Express, GraphQL and PostgreSQL, designed and optimized the SQL database schema and implemented several efficient GraphQL resolvers
- Developed a daily data ingestion pipeline that integrates four gaming platform APIs to automatically fetch and aggregate game statistics for over 500 employees, thereby enabling real-time leaderboard functionality

## Relevant Projects

**MiniC Compiler Development** | *C++, LLVM*

- Developed compiler for a C subset 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
- Achieved a 1.7x speedup on benchmarking programs by implementing an alloca2reg promotion optimization pass, reducing memory allocations, loads, and stores
- Achieved up to 5x speedup in sample programs by designing an optimization pass to eliminate redundant global loads utilizing the LLVM framework to perform loop analysis and code re-write

**Particle Simulation Acceleration** | *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, JavaScript, TypeScript, C, C++, SQL

**Systems & Tools**: Git, Linux/Unix, Slurm, Vim, LLVM/MLIR, Nsight Compute