Alireza Mohaghegh Dolatabadi

+1 (647) 967-7328 | a3mohagh@uwaterloo.ca | alireza-mohaghegh-dolat | amohagh.com

Education

University of Waterloo — Waterloo, Ontario
Honours Computer Science with Combinatorics & Optimization Minor; — *Sep 2023 - Dec 2027*
GPA: 4.0/4.0

Experience

HeyGen — Toronto, Ontario
Software Engineer – Internship — *May 2026 – August 2026*

- Ported H100 DiT video-generation inference to TPUs via **torchax**, reusing PyTorch model code and APIs
- Benchmarked Megatron TP vs Ulysses CP on TPU v6e/v7x, tuning sharding for HBM and collective costs
- Tuned sharding layouts, attention blocks, and custom/sparse Pallas kernels; validated GPU/TPU numerical parity

TalkingComputers (YC W26) — Remote
Researcher – Part-time Contract — *February 2026 – April 2026*

- Built synthetic data & task-generation pipeline and GRPO-based RL post-training stack for fine-tuning agents

Groq — Toronto, Ontario
Compiler Software Engineer – Internship — *September 2025 – December 2025*

- Added multimodal architecture support to Groq **inference engine**, extending model coverage on LPU backend
- Built **PyTorch** AOT CPU-vs-LPU validation to catch compiler, runtime, and numerical correctness bugs
- Implemented MLIR scheduling & optimization passes for Groq **LPU** compiler targeting inference latency

TC43 ULC — Toronto, Ontario
Quantitative Developer – Internship — *January 2025 – April 2025*

- Migrated feature generation from **Dask** to **Snowflake**'s UDTF, reducing runtime by **90%** and costs by **98%**
- Architected **500+** beta features, developing framework for table generation, regression logic, and dynamic updates
- Created templates to align feature spec with implementation, resolving consistency issues across **10,000+** features

Cohere — Remote
Data Engineer – Part-time Contract — *October 2024 – August 2025*

- Evaluated AI models' problem-solving skills, solving and correcting advanced math and programming questions

Projects

SpecLoRA |

- Engineered adaptive-**LoRA** speculative decoding, achieving **1.5-2x** speedup vs autoregressive decoding
- Built failure-mining loop to retrain draft models, improving acceptance rate and speculative decode throughput

PQTree library |

- Optimized SageMath's PQTree using **Rust** with multi-threading, increasing performance by **200%** in benchmarks
- Implemented **planar graph** extensions from published PQ-tree algorithms in SageMath-compatible library

Awards & Achievements

National Mathematics Olympiad - Silver Medalist — *September 2021*

- Ranked **12th** out of **20,000+** in National Mathematics Olympiad, an **IMO** qualifier

National Entrance Exam - Ranked 56th — *September 2022*

- Ranked **56th** out of **145,000+** in National Entrance Exam across math, physics, and reasoning

Combinatorics Olympiad (ICO) - Silver Medalist — *June 2022*

- Won **silver medal** at **Combinatorics** Olympiad, competing with international math olympiadists

Skills

Systems/ML: PyTorch, torchax, XLA, Pallas, MLIR, TPU, Slurm, Docker
Languages: Python, C/C++, Rust, Go, Mojo, Java
Data/Infra: Snowflake, PostgreSQL, Airflow, PyTest, Git, Nginx