

Maryam Gohargani

Toronto, ON · mgohargani@gmail.com · mgohar.com · github.com/magnesium4 · linkedin.com/in/maryam-gohargani

Applying for graduate study in trustworthy ML, starting Fall 2027.

EDUCATION AND AWARDS

University of Toronto

September 2019 – April 2022

Honours Bachelor of Science, Specialist in Computer Science, Minor in Mathematical Sciences

Conferred with High Distinction · GPA 4.0/4.0 · 94.4% average

Dean's List Scholar (2020–2022) · Honour Roll (2020–2022) · Entrance Scholarship (2019)

Selected coursework:

Theory of Computation **100%**
Discrete Mathematics **100%**
Languages & Automata **100%**
Computer Organization **99%**
Software Engineering **98%**
Programming Languages **98%**

Linear Algebra I & II **96%, 98%**
Systems Programming **95%**
Algorithms **94%**
Databases **94%**
Information Security **93%**
Artificial Intelligence **92%**

NSERC Undergraduate Student Research Award (USRA)

2021

ACM-ICPC, team “U of T White”

2020 – 2021

East Central North America Regional Contest — 19th of 71 teams

Bronze Medal, Iranian National Olympiad in Informatics (INOI)

2016, 2017

RESEARCH

Label-Only Membership Inference — reproduction and analysis

July 2026 – Present

Reproducing Choquette-Choo, Tramèr, Carlini & Papernot, ICML 2021 (arXiv:2007.14321)

PyTorch, CIFAR-10, scikit-learn, GPU training

- Trained the target models and implemented the attacks from scratch. The gap-attack baseline reproduces the theoretical $\frac{1}{2} + (\text{acc}_{\text{train}} - \text{acc}_{\text{test}})/2$ exactly, at 0.718
- The data-augmentation attack (47 label-only queries per point) reaches 0.750 against that baseline, while a naïve count-threshold variant scores 0.683 — below the gap attack — so the gain comes from weighting the queries rather than from augmentation-robustness alone
- Established the margin with a paired McNemar test ($p = 2.8 \times 10^{-20}$) after finding that retraining moved the headline figure across three runs; report the spread rather than the best run
- Offline ablations overturned my own explanation of the result: the un-augmented query is redundant rather than dominant, and 74% of the gain comes from four small rotations. Also found that $\pm 1^\circ$ rotations are a no-op under nearest-neighbour interpolation, making the real query budget 45 rather than 47
- github.com/magnesium4/label-only-membership-inference · [three-part write-up](#)

Model-Written Evaluations — reproduction against current models

July 2026 – Present

Reproducing Perez et al. (2022) — Python, LLM APIs (Anthropic, Gemini), Ollama

- Built a multi-provider evaluation harness from scratch, saving raw per-question results so analyses re-run offline without further API spend
- Established each dataset's label orientation directly from the data before interpreting any results, since “% matching” is not “% concerning” uniformly across files
- Measured survival-instinct and power-seeking behaviours across three model tiers and found they *decrease* with capability, partly reversing the paper's original scaling trend
- github.com/magnesium4/model-written-evals

CSC492H5, Computer Science Implementation Project (two terms), supervised by Lisa Zhang

- Contributed to a research implementation of miniKanren representing its search space as a first-order data structure rather than as closures, making the search inspectable and transformable
- Implemented disequality and type constraints and their propagation through substitution extension; contributed rewrite rules, debugging tools, documentation and tests
- github.com/JonathanMMartin/first-order-miniKanren

Trajectory Labs

June 2026 – Present

Member — Toronto's in-person AI-safety community

RESEARCH INTERESTS

- Privacy and security of machine-learning systems
- Transparency and evaluability of intelligent systems
- Offensive cyber capability in AI systems, and keeping capability evaluations contained
- AI misuse and unintended consequences

TEACHING

Teaching Assistant — Computer Science & Mathematics

September 2020 – April 2022

University of Toronto — 15 course sections

- Designed and delivered tutorials and practicals for groups of up to 90 students; managed discussion boards for courses with 800+ enrolment; graded for 500+ students
- Sections included Theory of Computation (CSC240H1, enriched), Algorithm Design and Analysis, Principles of Programming Languages, Artificial Intelligence, Databases, Data Structures, Software Tools and Systems Programming, and five mathematics courses

TECHNICAL SKILLS

Machine learning: PyTorch, torchvision, scikit-learn, NumPy, training CNNs from scratch, GPU training on Colab

ML security & evaluation: membership-inference attacks, model-behaviour evaluation harnesses, LLM APIs, local models via Ollama

Languages: Python, Java, C/C++, JavaScript/TypeScript, SQL, Haskell, Racket, MIPS Assembly, R, Bash

Tools: Git/GitHub, Linux/Unix, Docker, ~~TeX~~TeX, CI-friendly workflows

INDUSTRY EXPERIENCE

Quantitative Research Analyst

July 2026 – August 2026

Stealth startup — AI-powered equity-research platform (contract). Further details available on request

- Validating an AI model's predictions against ground truth: designing a back-testing methodology to test whether the stock the model picks as the winner is the one that actually delivers higher returns

Software Engineer

April 2022 – July 2023

Avenu Insights & Analytics (formerly Harmari by LTAS Technologies)

- Backend software in Python, Java, PHP and MySQL for municipal and regulatory technology products; data pipelines for KML/XML and geospatial datasets; automated QA tooling; led the team's migration from SVN to Git

SELECTED PROJECTS AND SERVICE

Full-stack and backend projects (FastAPI, React, PostgreSQL, MERN) at mgohar.com/projects.

VP of External Affairs, Mathematical & Computational Sciences Society (2021–22). Mentor, Women in Science & Computing (2021). Mentor, International Student Mentorship Program (2020).

A fuller CV, including all fifteen teaching sections and the complete coursework record, is at mgohar.com/cv.