Eugenie Y. Lai

Email: eugenie.y.lai@gmail.com
Google Scholar: scholar.google.com
Mobile: +1-617-460-0974

Education

- **Massachusetts Institute of Technology** — Cambridge, USA
  Ph.D. student - Electrical Engineering and Computer Science (GPA: 4.6/5.0) — Sept 2021 - October 2026
  Advised by ***Prof. Michael Cafarella***
- **University of British Columbia** — Vancouver, Canada
  BComm. - Combined Major in Business and Computer Science (GPA: 4.0/4.0) — Sept 2015 - June 2021
  Advised by ***Prof. Rachel Pottinger***

Candidate Summary

MIT Computer Science PhD candidate working on machine learning for data systems; author of 12+ peer-reviewed publications at venues including SIGMOD, VLDB, ICLR; available to start full-time in October 2026.

Industry Internships

- **Student Researcher, Google** — May 2026-Present
  - Leading a project that improves scalability for AI operators in Google BigQuery, using model-as-a-cache, mentored by Dr. Fatma Ozcan in the Systems Research Group.
  - Designed a multi-layer caching algorithm to dynamically route LLM requests to the most suitable caching strategy and successfully reduced the runtime by at least 10x.
- **Research Intern, Microsoft Research** — May 2023-August 2023
  - Led a project that aims to predict data transformations for business intelligence (BI) applications such as Tableau and Power BI, mentored by Dr. Yeye He in the Data Systems Group.
  - Designed and implemented a new graph algorithm, in combination with machine learning models, to predict a sequence of structural transformations for a set of user input tables, by using table-level features and joins.
  - Conducted experiments on real-world user data, and showed a 30% F-score improvement over SoTA methods.
- **Software Developer Intern, Statistics Canada** — September 2017-April 2018
  - Implemented a web service application using technologies including C#, JavaScript, SQL, ASP .NET.
  - Received a full-time offer after internship from the Statistics Information System Division (SISD) executive team.

Ongoing Projects

- **Learning Data-Grounded Procedural Knowledge from Divergent Agent Trajectories**
  Eugenie Lai, Sushrut Borkar, Samuel Madden, Michael Cafarella
- **Don't Repeat Mistakes: A Lesson-Driven Agentic System for Data Science**
  Sushrut Borkar, Eugenie Lai, Samuel Madden

Selected Publications

- **Interpretable Attribute Discretization**
  Eugenie Lai, Inbal Croitoru, Brit Youngmann, Sainyam Galhotra, Elkindi Rezig, Michael Cafarella
  SIGMOD 2026
- **KramaBench: A Benchmark for AI Systems on Data-to-Insight Pipelines over Data Lakes**
  *Eugenie Lai**, Gerardo Vitagliano*, Ziyu Zhang*, Sivaprasad Sudhir, Om Chabra, Anna Zeng, Anton A Zabreyko, Chenning Li, Ferdi Kossmann, Jialin Ding, Jun Chen, Markos Markakis, Matthew Russo, Weiyang Wang, Ziniu Wu, Michael J Cafarella, Lei Cao, Samuel Madden, Tim Kraska
  ICLR 2026
- **Auto-Prep: Holistic Prediction of Data Preparation Steps for Self-Service Business Intelligence**
  Eugenie Lai, Yeye He, Surajit Chaudhuri
  Proceedings of the VLDB Endowment 18 (7), VLDB 2025
- **SeerCuts: Explainable Attribute Discretization**
  Eugenie Lai, Inbal Croitoru, Noam Bitton, Ariel Shalem, Brit Youngmann, Sainyam Galhotra, Elkindi Rezig, Michael Cafarella
  Companion of the 2025 International Conference on Management of Data, SIGMOD Demo 2025
- **Bottom-Up Standardization For Data Preparation**
  Eugenie Lai, Yuze Lou, Brit Youngmann, Michael Cafarella
  EDBT 2025
- **Workload-Aware Query Recommendation Using Deep Learning**
  Eugenie Lai, Zainab Zolaktaf, Mostafa Milani, Omar AlOmeir, Jianhao Cao, and Rachel Pottinger
  EDBT 2023

Selected Awards

- 2022-2025 Natural Sciences and Engineering Research Council of Canada (NSERC) Scholarships – $126,000
- 2020 NSERC Undergraduate Student Research Award – $4,500

Invited Talks

- Auto-Prep, hosted by Prof. Rachel Pottinger, University of British Columbia — July 2025
- Extracting Insights from Raw Data, hosted by Prof. Immanuel Trummer, Cornell University — October 2025
- Systems to Automate Data-to-Insight Pipelines, Google Research — February 2026