Eugenie Y. Lai

Email: eylai@mit.edu
Google Scholar: scholar.google.com
Mobile: +1-617-460-0974
Website: eugenielai.github.io

Education

- **Massachusetts Institute of Technology** — Cambridge, USA
  Ph.D. student - Electrical Engineering and Computer Science (GPA: 4.6/5.0) — Sept 2021 - Present
  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***

Research Interests

The central goal of my research is to use algorithms and machine learning models to help users with the *data preparation* phase of the data-to-insight pipeline. Specifically, my research focuses on data preparation step prediction and explainable data preparation.

Research Experience

- **Research Assistant, MIT Data Systems Group** — September 2021-Present
  - Proposed a bottom-up script standardization framework that takes a user's data preparation script and transforms it into a simpler, more standardized, more boring version of itself.
  - Developed an algorithm to achieve optimization objectives subjective to constraints, and implemented a prototype system written in 3,000 lines of predominately Python source code.
  - Evaluated our approach against state-of-the-art methods on six real-world datasets and improved script standardization by 39.5%, while GPT-4 achieves 2.9%.
- **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).
  - Proposed a new algorithm to predict a sequence of structural transformations for a set of user input tables in Power BI so that the BI model is optimized, by using table-level features and join relations.
  - Leveraged a combination of machine learning models and graph search algorithm, conducted experiments on real-world user data, and showed a 30% F-score improvement over state-of-the-art methods.
- **Research Assistant, UBC Data Management and Mining Lab** — May 2019-August 2021
  - Participated in three research projects (1 first-author and 2 second-author) on topics including relational database provenance data usability and SQL query prediction using deep learning models.
  - Formulated query recommendation as a query prediction task and presented a new approach to recommend query information by learning from the sequential knowledge exploration patterns of historical users.
  - Trained and tuned RNN, convolutional sequence-to-sequence, and transformer models, extracted and cleaned human-generated query sessions from two real-world workloads, and made all the artifacts publically available.

Industry Experience

- **Data Scientist Intern, UBC Data Science Institute** — May 2019-August 2019
  - Partnered with the Environmental Sustainability Advisory Committee of the City of Surrey, BC to guide the development of the Surrey Electric Vehicle Transformation Strategy.
  - Designed and developed a web application as a data visualizer to give the city planners a user-friendly way to interact with the data, including the spatial and temporal distribution of vehicle stock, traffic flows, etc.
- **Software Developer Intern, Statistics Canada** — September 2017-April 2018
  - Implemented a web service application embedded in a toolbox using technologies such as C#, JavaScript, SQL, ASP .NET and exceeded clients' expectations by optimizing jQuery widgets.
  - Received a full-time return offer after the internship.

Selected Publications and On-Going Work

- **AutoPrep: Holistic Prediction of Data Preparation for Self-Service Business Intelligence** — Submitted
  Eugenie Lai, Yeye He, Surajit Chaudhuri
- **Bottom-Up Standardization For Data Preparation** — Submitted
  Eugenie Lai, Yuze Lou, Brit Youngmann, Michael Cafarella
- **LucidScript: Bottom-Up Standardization For Data Preparation** — VLDB '24 Demo
  Eugenie Lai, Yuze Lou, Brit Youngmann, Michael Cafarella
- **Extract-Transform-Load for Video Streams** — VLDB '23
  Ferdinand Kossmann, Ziniu Wu, *Eugenie Lai*, Nesime Tatbul, Lei Cao, Tim Kraska, Samuel Madden
- **Workload-Aware Query Recommendation Using Deep Learning** — EDBT '23
  Eugenie Lai, Zainab Zolaktaf, Mostafa Milani, Omar AlOmeir, Jianhao Cao, and Rachel Pottinger

Selected Awards

- 2022-2025 Natural Sciences and Engineering Research Council of Canada (NSERC) Scholarships – $126,000
- 2020 NSERC Undergraduate Student Research Award – $4,500
- 2020 University of British Columbia (UBC) Computer Science Rick Sample Memorial Research Award – $2,500