

Lawrence Lai

(626)-202-2052, Cambridge, MA

laitcl@gmail.com

github.com/laitcl | [linkedin.com/in/laitcl/](https://www.linkedin.com/in/laitcl/) | laitcl.com

Skills

Programming Languages

Python, Ruby, Java

Data Engineering Specific Tools

MYSQL, PostgreSQL, Kafka, Airflow, AWS services (EC2, S3, RDS)

Tools and Packages

Git, AWS, Ruby on Rails, React, Dash, Numpy, Pandas, Keras

Experience

Indigo Ag.

Senior Data Engineer, Boston, MA

March 2021 – Present

- Automated and optimized high volume ETL and data analysis processes utilizing Python, R, Amazon Managed Workflows for Apache Airflow (MWAA), incorporating statistical methods including linear mixed models and Fishers exact test to provide intelligence and ensure scalability for biological research.
- Designed and maintained over 100 Indigo Ag's schema tables and/or views in MySQL to ensure data consistency, full normalization, and data transparency.
- Designed and maintained two data governance tools using plotly Dash, used for data upload, processing, and QC, maintaining data integrity in Indigo Ag's biological research
- Migrated database infrastructure hosted on AWS aurora to utilize MySQL8.0 in serverless infrastructure

PatientsLikeMe

Data Engineer, Cambridge, MA

September 2019 – March 2021

- Maintained and designed over 100 extract/transform/load (ETL) tasks written in Ruby, Java, and Python
- Provided stakeholders with streamlined, appropriately denormalized, and accurate data free of patient identifying or otherwise sensitive information
- Improved database infrastructure using AWS tools EC2, S3, Lambda, RDS, Aurora, and Parameter Store, storing over 700 GB of historical data
- Developed web apps for data collection, leveraging Ruby on Rails, Devise with SSO, and React

Insight Data Science

Data Engineering Fellow, Boston, MA

June 2019 – August 2019

- Developed data pipeline to analyze real time chat room traffic and sentiment for over 100 channels, highlighting media content with high audience participation, visualized using chrome extensions
Deployed chat bot written in NodeJS to ingest messages and save sentiments for data pipeline usage
- Optimized pipeline to process over 2000 messages per minute with cloud computing using AWS EC2: data ingestion and allocation by Kafka, database management by PostgreSQL
- Built classification model for chat room reaction utilizing Python packages Pandas, Numpy, and Keras, capable of identifying disappointment, laughter, and questions from chat room messages

Massachusetts Institute of Technology

Graduate Research Assistant, Cambridge, MA

September 2013 – June 2019

- Characterized chemical details of reactive systems through computational generation of chemical models containing over 200 species and 4000 reactions
- Developed open-source freeware Reaction Mechanism Generator by contributing > 3000 parameters as training data for convolutional neural network prediction algorithm for chemical characteristics
- Developed efficient estimation method of unknown chemical parameters using group additivity methods and decision trees to achieve thermochemistry accuracy of < 3 kcal/mol, using Numpy and Pandas

Education

PhD in Chemical Engineering, Massachusetts Institute of Technology,
Cambridge, MA

June 2019

BSE in Chemical Engineering, University of Michigan, Ann Arbor, MI

December 2012