David Mingfei Liu

liu.davi@northeastern.edu
571-528-3466
dliu18.github.io

Education

- **Northeastern University** (2020 - Present)
  Ph.D. in Computer Science
  Focus ares: Fairness in Machine Learning, Computational Social Science, and Computational Reproducibility.

- **Princeton University** (2018)
  B.S.E in Computer Science with Certificate in Statistics and Machine Learning
  GPA: 3.73 — *summa cum laude*
  Coursework includes: Probability and Stochastic Systems, Machine Learning and AI, Neural Networks, Analysis of Big Data, Social Networks, Fairness in Machine Learning, Applied Social Statistics, Applications in Econometrics.

Publications

- Liu, David M., and Matthew J. Salganik. "Successes and Struggles with Computational Reproducibility: Lessons from the Fragile Families Challenge." *Socius*, (January 2019). doi:10.1177/2378023119849803

Projects

- **Analyzing language trends in post-secondary FERPA notices**
  Advisor: Arvind Narayanan Feb 2017 - Jun 2017
  Scraped education policy documents and used SVM classification to isolate FERPA notices. Data used to survey the state of education-policy compliance.
  `https://github.com/dliu18/papers/blob/master/legal_nlp.pdf`

- **Discovering Princeton's History: A textual analysis of collegiate newspaper headlines**
  Advisor: Brian Kernighan Sep 2016 - Jan 2017
  Created an interactive, public n-gram viewer based on the electronic archives of Princeton's student-run campus newspaper.
  `https://github.com/dliu18/papers/blob/master/digital_humanities.pdf`

Industry Experience

- **Bloomberg LP - *Software Engineer***   July 2018 - Present
  Implementing scalable, reliable C++ services on a derivatives market data team with a focus on message queueing, automated testing, and interface design.

- **The New York Times - *Software Intern***   Summer 2017
  Architected backend supporting the renovation of the "My Account" page of the New York Times, increasing site usability. Renovated site launched in October 2017. Initiated side project with the NYT Data Science team, improving visualization of readership data from embedded static files to native, interactive Google Charts.

- **HookLogic Inc. - *Data Science Intern***   Summer 2016
  Reduced data analysis overhead time for team by initiating Hive queries on AWS clusters to analyze performance of machine learning models.

Awards and Honors

- **Sigma Xi Nomination:** Princeton chapter   **Jun 2018**
- **Daily Princetonian Award for Excellence in Sports Writing:** Annual award given to staff member demonstrating superior writing.   **Jun 2016**

- **Residential College Advisor Scholarship:** One of 30 campus advisers chosen from class on the basis of leadership and character. Responsibilities include building community among 20 freshmen for one year. Monetary compensation of $20k, covering room and board. **Jun 2016**

Activities

- **ML Reproducibility Challenge:** Participant in Papers with Code's challenge to reproduce machine learning results from top computer science conferences. **2020**
- **Peer Academic Advisor, Princeton University:** Provide academic assistance to freshmen and sophomores with choosing courses, majors and independent projects at Princeton University **2016 - 2017**
- **Undergraduate Lab Teaching Assistant, Princeton University:** Provide debugging and conceptual assistance on programming assignments for introductory Computer Science courses; assist the head Lab TA in the smooth functioning of the Lab TA program **2015 - 2016**

Talks

- **RabbitMQ Summit:** Co-delivered the Bloomberg keynote on scaling distributed message queueing. Recording is at `rabbitmqsummit.com`. **Nov 2019**
- **Fragile Families Challenge Scientific Workshop:** Led a breakout session on using Docker to improve the reproducibility of crowdsourced machine learning prediction models. **Nov 2017**