Yunyi Li

Ph.D. Candidate
McCombs School of Business
The University of Texas at Austin
Cell: (319) 519-3209
Email: Yunyi.Li@mccombs.utexas.edu
Web: yunyli.github.io

Education

- **Ph.D. in Information, Risk and Operations Management**
  The University of Texas at Austin, Expected May 2026
  - Thesis: Trustworthy AI and Data Quality: Understanding and Mitigating Label Errors
  - Advisors: Maria De-Arteaga and Maytal Saar-Tsechansky
- **M.S. in Informatics-Information Science**
  The University of Iowa, May 2020
- **B.B.A in Business Analytics with Honors (top 1%)**
  The University of Iowa, May 2018
  - Large Data Analysis Certificate
  - Study Abroad: National Taiwan University of Science and Technology, Taiwan, 2015

Research Interests

- Data Integrity, Data Quality, Label Bias, Mislabeling Detection, Human-AI Collaboration, Interpretable Machine Learning, Ethical AI, AI for Social Good.

Selected Publications

1. **Li, Y.**, De-Arteaga, M., and Saar-Tsechansky, M., *Bias-Aware Mislabeling Detection via Decoupled Confident Learning*. Under review at **Information Systems Research**. 2025
2. Wang, T., Yang, J., **Li, Y.**, and Wang, B., *Partially interpretable estimators (PIE): black-box-refined interpretable machine learning*. **Informs Journal of Computing (IJOC)**, 2025
3. **Li, Y.**, De-Arteaga, M., and Saar-Tsechansky, M., *Label Bias: A Pervasive and Invisibalized Problem*. **Notices of American Mathematical Society,** 2024
4. Srivastava, S., Xu, Z., **Li, Y.**, Street, W. N., and Gilbertson-White, S., *Gaussian process regression and classification using International Classification of Disease codes as covariates*. **Stat.,** 2023
5. **Li, Y.**, De-Arteaga, M., and Saar-Tsechansky, M., *More Data Can Lead Us Astray: Active Data Acquisition in the Presence of Label Bias*. **AAAI Conference on Human Computation and Crowdsourcing (HCOMP),** 2022
6. Gilbertson-White, S., Srivastava, S., **Li, Y.**, Laures, E., Saeidzadeh, S., Yeung, C., and Chae, S., *Multimorbidity and cancer: using electronic health record (EHR) data to cluster patients in multimorbidity phenotypes*. **Journal of Pain and Symptom Management (JPSM),** 2019
7. Wang, Y., Wang A., Liu, Z., Thurman, A., Powers, L., Zou M., Hefel, A., **Li, Y.**, Zabner, J., and Au, K.F., *Single-molecule long-read sequencing reveals the chromatin basis of gene expression*. **Genome Research,** 2019
8. **Li, Y.**, and Wang, T., *Next Hit Predictor - Self-exciting Risk Modeling for Predicting Next Locations of Serial Crimes*. **NeurIPS AI for Social Good,** 2018

Selected Awards

- **McCombs School of Business Dean's Fellowship** - The University of Texas at Austin (2020–2025).
- **UT Austin Machine Learning Laboratory Research Grant** - The University of Texas at Austin (2021).
- **UIowa Graduate College Presentation Fund** - The University of Iowa (2019).
- **Big Data Grant from National Science Foundation (NSF)** - The University of Iowa (2017–2018).
- **Carol Fethke Honors Scholarship** - The University of Iowa (2018).
- **Iowa Center for Research by Undergraduates Research Fellowship** - The University of Iowa (2017).
- **Dean's List** - The University of Iowa (2015–2018).

Selected Talks

1. **21st Big XII MIS Research Symposium (Invited Talk)**, April 2025
2. **2024 INFORMS Annual Meeting**, October 2024
3. **Symposium on Statistical Challenges in Electronic Commerce Research (SCECR)**, June 2024
4. **2024 Good Systems Symposium**, March 2024
5. **McCombs School of Business Research Spotlight (Invited Talk)**, November 2023
6. **2023 INFORMS Annual Meeting**, October 2023
7. **International Conference on Machine Learning (ICML)**, July 2023
8. **AAAI Conference on Human Computation and Crowdsourcing**, November 2022
9. **Annual Conference on Neural Information Processing Systems (NeurIPS)**, December 2018

Professional Service

- **Conference Session Chair:** INFORMS Annual Meeting 2024
- **Peer reviewer:**
  - International Conference on Information Systems (ICIS) 2023
  - Hawaii International Conference on System Sciences (HICSS) 2022 & 2023
- **Conference Assistant:** International World Wide Web Conference(WWW) 2023

Research Experiences

- **Graduate Researcher** The University of Texas at Austin, 08/2020–Present
  McCombs School of Business
  Topic: Trustworthy AI and Data Quality – Understanding and Mitigating Label Errors
  - Conducted a comprehensive literature review on the challenges, complexities, and solutions related to bias and noise in data annotations.
  - Investigated the origins and downstream effects of label bias in data powering information systems.
  - Led and published the project *More Data Can Lead Us Astray: Active Data Acquisition in the Presence of Label Bias*.

- First-authored and published the article *Label Bias: A Pervasive and Invisibilized Problem*.

- **Research Assistant** The University of Iowa, 12/2018–05/2020
  Department of Statistics
  Topic: Statistical Learning of Electronic Health Records (EHR) to Cluster Cancer Patients by Multimorbidity Profiles
  - Identified common clusters of chronic conditions co-occurring in patients with advanced solid tumor cancer and described differences across these clusters.
  - Designed a kernel-based distance function for cancer patient data using electronic health records.
  - Employed advanced statistical learning methods, including Gaussian Processes and Gibbs Sampling, to solve regression and classification problems involving cancer patient data.

- **Research Intern** The University of Iowa, 05/2018–09/2018
  Carver College of Medicine
  Topic: Single-Molecule Long-Read Sequencing Reveals the Chromatin Basis of Gene Expression
  - Created visualizations using R for chromatin accessibility and nucleosome positioning.
  - Generated an Rcircos landscape plot to display multiple features at the whole-genome scale.
  - Applied a Hidden Markov Model and dynamic programming to analyze signal distribution.
  - Participated in team meetings to brainstorm solutions and perform quality control tasks.

- **Research Lead** The University of Iowa, 12/2016–07/2018
  Tippie College of Business
  Topic: Next Hit Predictor – Self-Exciting Risk Modeling for Predicting Serial Crime Locations
  - Designed a novel spatio-temporal crime prediction model for the Cambridge, MA Police Department.
  - Formulated a convex learning objective based on pairwise rankings and trained the model using stochastic gradient descent (SGD).
  - Presented research at NeurIPS 2018, Iowa Research Open House, Data Mining Iowa Group, and the Iowa Dare to Discover Campaign Showcase.

Teaching Experiences

*__Completed Teaching Training and Supervised Teaching__ at McCombs School of Business

1. **Artificial Intelligence & Machine Learning for Executives: MIS 281N**, Teaching Assistant
   The University of Texas at Austin (Spring 2025)
2. **Intro to Information Technology Management: MIS 301**, Teaching Assistant
   The University of Texas at Austin (Spring 2025)
3. **Ethics of AI: BAX 372**, Teaching Assistant and Course Co-designer
   The University of Texas at Austin (Spring 2024)
4. **Intro to Problem Solving and Programming: MIS 304**, Teaching Assistant
   The University of Texas at Austin (Fall 2020 & Fall 2022)
5. **Data Science: MSCI 6070**, Teaching Assistant
   The University of Iowa (Spring 2020)

6. **Advanced Database Management & Big Data: MSCI 4220**, Teaching Assistant
   The University of Iowa (Spring 2020)
7. **Data Management and Visual Analytics: MSCI 6050**, Teaching Assistant
   The University of Iowa (Fall 2018)
8. **Big Data Summer School**, Student Mentor
   The University of Iowa (Summer 2017 & Summer 2018)

Leadership Experiences

- **McCombs Ph.D. Council** McCombs School of Business, 09/2022–07/2023
  - Assisted in planning and organizing the McCombs School of Business Centennial Celebration.
  - Represented the interests and concerns of Ph.D. students to the administration to ensure their voices were heard.
- **Graduate Student Assembly** The University of Texas at Austin, 09/2021–07/2022
  - Participated in university-wide policy setting for the Graduate School and reviewed academic programs.
  - Served as a voting member in Graduate School decision-making processes.
- **2nd Place, USITCC Business Analytics Competition** San Antonio, TX, 04/2018 *U.S. Information Technology Collegiate Conference (USITCC)*
  - National undergraduate information technology competition held annually.
  - Competed in two events: awarded 2nd place in Business Analytics and received Honorable Mention in Database Management.
- **Team Leader, MinneMUDAC Competition** The University of Iowa, 08/2017–05/2018
  Annual Midwest Undergraduate Data Analytics Competition (MinneAnalytics) — Eden Prairie, MN
  - Led a team of four business students to develop data-driven solutions for real-world problems.
  - Applied machine learning techniques to analyze the relationship between water quality and housing prices.

Skills

Programming/Software:

- Python, R, C++, SQL, LaTeX
- Microsoft Certified: Azure AI Fundamentals (November 2023)

Natural Languages:
Chinese (native), English (fluent, lived in U.S. for 10 years)