

Xinyi (Jessy) Han | Curriculum Vitae

77 Massachusetts Avenue, 32-D666 – Cambridge, MA 02139

☎ (617) 945 3358 • ✉ xyhan@mit.edu • 🌐 web.mit.edu/xyhan/www

Research interest: fair causal inference for decision-making in social science and life science

Education

Massachusetts Institute of Technology, IDSS/LIDS

Cambridge, MA

Ph.D. in Social & Engineering Systems, GPA – 5.0/5.0

Sep 2020–Jun 2025 (expected)

Advisors: Dr. Fotini Christia and Dr. Devavrat Shah

Columbia University

New York, NY

B.S. in Computer Science, GPA – 4.10/4.0

Sep 2016–May 2020

Double minor in Applied Mathematics and Economics

Graduation honors: Summa Cum Laude, Russell C. Mills Award

Advisors: Dr. Augustin Chaintreau and Dr. Ana-Andreea Stoica

Research and Work Experiences

Regularized Causal Inference Method for Survival Probability

Cambridge, MA

Lead Researcher

May 2022–Now

- Develop a novel regularized synthetic intervention method for survival probability analysis
- Apply the regularized synthetic intervention to understand the relative effectiveness of two different treatments for T-cell lymphoma (TCL) patients
- Produce one of the largest datasets on TCLs patients to serve as a benchmark dataset in the field

Impact of Police Surveillance on Policing Racial Bias

Cambridge, MA

Lead Researcher

Jan 2023–Now

- Conduct extensive analysis on adoption of policing technologies, such as face recognition, body-worn cameras, and predictive policing, among more than 2000 police departments from 2000 to 2022
- Utilize high-resolution data on individual traffic stops, arrests, and other law enforcement actions to estimate shifts in racial biases resulting from the implementation of new tools
- Promote a greater understanding of the impact of new policing technologies on communities of color and raise awareness of the potential for such technologies to exacerbate existing inequalities

Data Generative Approach to Record linkage

Cambridge, MA

Lead Researcher

Sep 2022–Now

- Propose a novel data generative process for multi-stage data to enable more accurate and comprehensive record linkage
- Design maximum weight matching algorithm specifically tailored for large-scale administrative records to merge data from different sources
- Result in a more systematic characterization and insightful understanding of multi-stage processes

Systematic Causal Framework of Racial Disparity in Policing

Cambridge, MA

Lead Researcher

Oct 2020–Sep 2022

- Develop a systematic causal framework for estimating racial bias in policing by incorporating 911 calls-for-service data
- Apply the framework to evaluate racial disparities in three major American cities: New York, Seattle, and New Orleans with over 22 million 911 call records and over 400,000 police stop records
- Emphasize the significance of analyzing 911 calls as a vital pathway to police-civilian interactions, which had often been overlooked in prior works

Open Datahub for Criminal Justice System Data

Cambridge, MA

Lead Developer

Sep 2021–Now

- Conduct Freedom of Information Act requests to over 40 major police departments across the US
- Extract and systematize diverse data sources related to law enforcement, including texts from social media, images from police body cameras, and mobility data from cell phones
- Host all relevant data on Azure Synapse Analytics, utilizing Python/R/Stata interfaces to allow users from different programming backgrounds to query the data

Causal Inference Python Package for Panel Data

Cambridge, MA

Lead Developer

Jan 2021–Now

- Develop a Python package for causal inference with panel data that includes the functionality of estimating counterfactual outcomes using synthetic interventions
- Expand the library to cover classical causal inference algorithms for panel data to allow for more efficient analysis

Explaining and Reproducing Disparities in Homophilous Networks

New York, NY

Research Assistant

Jan 2020–Feb 2021

- Analyze a few well-observed network-growing mechanisms with a simple model to identify the sufficient and necessary conditions for two phases of the chasm effect to occur
- Develop a bi-affiliation bipartite network-growth model that generalizes the simple model and successfully captures disparities at all social levels, reproducing real social networks
- Demonstrate the potential impact of the chasm effect on creating fairer systems by applying the model to advertisement and fact-checking contexts

Network influence of Diversified Seeding in Biased Networks

New York, NY

Research Assistant

Jan 2019–March 2020

- Develop a theoretical model of biased networks to investigate the intricate relationship between diversity and efficiency and identify analytical conditions for equitable choices of early adopters
- Design and test a set of algorithms leveraging network structure to optimize information diffusion while avoiding creating disparate impacts based on participant demographics, such as gender or race
- Utilize data from the DBLP network to confirm the analytical condition is often met in real networks

Anomaly Detection Models for Google Ads Risk Engine

Mountain View, CA

Software Engineer Intern

May 2019–Aug 2019

- Design and implement improved anomaly detection models for Google Ads Risk Engine
- Achieve an AUROC of 0.6 and a 13-minute reduction in suspension time compared to Google's current model on production data

Latency Analysis of Google Fuchsia OS

San Francisco, CA

Engineering Practicum Intern

May 2018–Aug 2018

- Instrument storage devices to capture relevant parameters and latency
- Analyze storage device traces to compute the latency behavior of Fuchsia, Google's next-generation of operating system
- Develop deterministic models with a success rate of 99.77% to predict latency

Honors and Awards

2021: Michael Hammer Society of Fellows

2020: Honorable Mention for 2020 Outstanding Undergraduate Researcher

Publications

- **Incorporation of Machine Learning Tools to Predict Global Outcomes for Patients with Relapsed and Refractory Peripheral T and NK/T-Cell Lymphomas in Contemporary Era**
Blood 2022
L.S. Boussi, M. Koh, **X. Han**, L. Peng, M. Koh, I. Eche, J. Ford, S. Singh, E. Miranda, C. Chiattonne, C. Van Der Weyden, M. Prince, F. Foss, S. Yoon, W. Kim, G. Panchoo, E. Verburgh, J. Cuenca Alturas, M. Al-mansour, M. Manni, M. Federico, M.E. Cabrera, B. Casadei, P.L. Zinzani, N. Yoshida, T. Okatani, M.H. Merrill, E.D. Jacobsen, O.A. O'connor, E. Marchi, and S. Jain
- **Chasm in Hegemony: Explaining and Reproducing Disparities in Homophilous Networks**
SIGMETRICS 2021
Yiguang Zhang, **Jessy Xinyi Han**, Ilica Mahajan, Priyanjana Bengani, and Augustin Chaintreau
- **Diversity and Bias in the Influence Maximization Problem**
Poster Session at WINE 2019
Ana-Andreea Stoica*, **Jessy Xinyi Han*** and Augustin Chaintreau (* The authors have contributed equally)
- **Seeding Network Influence in Biased Networks and the Benefits of Diversity**
The Web Conference 2020 (*oral presentation*)
Ana-Andreea Stoica, **Jessy Xinyi Han** and Augustin Chaintreau

Teaching and Community Service

Nov 2022: SES Graduate Application Assistance Program for the Underrepresented Groups
Sep 2022–Dec 2022: 6.7810 Algorithms for Inference, Massachusetts Institute of Technology
Nov 2021: SES Graduate Application Assistance Program for the Underrepresented Groups
Nov 2020: SES Graduate Application Assistance Program for the Underrepresented Groups
Jan 2020–May 2020: COMS E6998 Social Networks, Columbia University
Sep 2019–Dec 2019: COMS 3203 Discrete Mathematics (head TA), Columbia University
Jan 2019–May 2019: COMS 3203 Discrete Mathematics, Columbia University
Sep 2017–Dec 2018: COMS 3203 Discrete Mathematics, Columbia University

Skills

Programming languages: Proficient: Python/Java; Experience with: C/C++/Matlab/R
Languages: Native: Chinese; Advanced: English