Mao Li

(+1) 734-882-9041 | maolee@umich.edu | limaoum.github.io

Li Mao | LiMaoUM | Google Scholar

Ann Arbor, MI, USA

Summary

- Developing and applying large language models for social text analysis, with a focus on knowledge distillation, stance detection, public opinion measurement, and the interpretation of large-scale social discourse
- Designing scalable algorithms to extract behavioral signals and opinion trends from organic social media data, advancing methods that complement traditional surveys in real-time public opinion research
- Building computational methods for statistical and public-sector data systems, including applications with the U.S. Census Bureau, to improve precision, comparability, bias assessment, and interpretability
- Designing and implementing LLM-based multi-agent systems to simulate opinion dynamics on social media, bridging computational social science, agent-based modeling, and scalable machine learning infrastructure
- Developing reusable research tools, open workflows, and training-oriented materials for text analysis, summarization, and social data science in interdisciplinary research settings

Education

- **University of Michigan** *July 2026*
  Ph.D. in Survey & Data Science and Scientific Computing (dual degree) Ann Arbor, MI
- **University of Michigan** *April 2021*
  M.Sc. in Survey Methodology Ann Arbor, MI
- **Nanjing University** *July 2019*
  Bachelor of Laws in Political Science and Administrative Science Nanjing, China
  - Exchange: Penn State University (Sep–Oct 2018), National Taiwan University (Jan 2018)

Research Experience

- **Graduate Student Research Assistant** *Sep 2021 – Present*
  New Approaches to Analyzing Social Media Content for Enhancing Census Bureau Data Ann Arbor, MI
  - Developed topic modeling pipelines and summarization frameworks to categorize and distill public discourse around the U.S. Census Bureau from large-scale Twitter data
  - Designed a probabilistic embedding-based approach to represent large social media corpora, enabling more principled identification of representative posts and filtering relevant content beyond lexical matching
  - Applied large language models with parameter-efficient fine-tuning for downstream tasks such as stance detection, text summarization, and demographic inference across heterogeneous online populations

- **Project Lead** *Nov 2024 – Present*
  Emerging Social Media Platforms Research Ann Arbor, MI
  - Collected and processed over 10 million posts from each of Bluesky and TruthSocial, including complete conversation threads and repost cascades. Developed and continue to maintain this large-scale, structured dataset to support ongoing analyses of information diffusion, motif dynamics, and collaboration with other researchers
  - Developed a network analysis framework to model information diffusion through motif analysis and structural virality quantification
  - Designed multi-agent simulation systems to model public opinion dynamics in social media conversations

- **Graduate Student Research Assistant** *Oct 2021 – Dec 2021*
  Real Estate Data for Housing Dynamics (PI: Fabian Pfeffer) Ann Arbor, MI
  - Studied feasibility of extracting real estate valuation data from web sources.

◦ Analyzed scalability of scraping algorithms.

• **Graduate Student Research Assistant** *Sep 2019 – Sep 2021*
National Neighborhood Data Archive (NaNDA) Ann Arbor, MI
◦ Built national neighborhood datasets using Python, R, ArcGIS, and scraping.
◦ Assisted analyses of neighborhood effects on cognitive function for 32,000 older adults.

• **Undergraduate Research Fellow** *Nov 2017 – Nov 2018*
Institute of Advanced Studies in Humanities & Social Science Nanjing, China
◦ Text analysis on Middle East conflicts and Jewish political traditions.

• **Undergraduate Research Assistant** *Sep 2015 – Jan 2018*
School of Government Nanjing, China
◦ Conducted surveys with 1,100 rural migrant workers and 4,000 college students.

Teaching Experience

• **Guest Lecturer** *Mar 2023*
SurvMeth 622 – Text & Sentiment Analysis, ISR Ann Arbor, MI
◦ Lecture on NLP techniques: TF-IDF, Bag of Words, Word2Vec, BERT, GPT.

• **Guest Lecturer** *Jun 2023*
Summer Course – Python & NLP Nanjing, China
◦ Covered Python, text classification, sentiment analysis, deep learning with Transformers.

Publications

C=Conference, J=Journal, P=Patent, W = Working Paper, T=Thesis

[C.1] **Li, M.**, Chen, X.& Chang, S. (2027 forthcoming). **Same Event, Different Truths: How Next-Generation Social Media Platforms Reframe Political Events**. Manuscript accepted for publication in *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM 2027)*.

[C.2] **Li, M.** & Chen, X. (2026). **Who Talks to Whom: Quantifying Echo Chamber Effects in Emerging Social Media Platforms**. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM 2026)*.

[C.3] **Li, M.**, Conrad, F., & Gagnon-Bartsch, J. (2025). **FastLexRank: Bring Order into Social Media Posts Using Lexical Ranking Algorithm**. In *Proceedings of the ICWSM 2025 Workshop on R2CASS*. https://doi.org/10.36190/2025.48

[C.4] **Li, M.**, & Chen, X. (2025). **Partitioned realities: How next generation social media platforms shape competing narratives of "truth"**. In *Proceedings of the ICWSM 2025 Workshop on Next-Gen and Alternative Social Media*. https://doi.org/10.36190/2025.42

[J.1] **Li, M.**, Morales, S., Nurumov, K., & Lee, S. (Accepted). **Using Large Language Models to Autocode Probing Responses to the Self-Rated Health Question: Strengths and Weaknesses in a Multilingual Survey Context**. Manuscript under minor revision for publication in *Journal of Survey Statistics and Methodology (JSSAM)*.

[J.2] **Li, M.**, Conrad, F., & Owens, V. L. (2026). **Effects of interviewer language and dialect choice on questions about political trust: Examining the Asian Barometer Survey in China and the Philippines**. *Journal of Survey Statistics and Methodology, 14*(1), 116–136.

[J.3] **Li, M.**, & Conrad, F. (2024). **Tracking census online self-completion using Twitter posts**. *Social Science Computer Review, 42*(1), 45–60. https://doi.org/10.1177/08944393241268461

[J.4] Finlay, J., Esposito, M., **Li, M.**, et al. (2021). **Neighborhood active aging infrastructure and cognitive function: A mixed-methods study of older Americans**. *Preventive Medicine, 150*, 106669.

[J.5] Finlay, J., Esposito, M., **Li, M.**, et al. (2021). **Can Neighborhood Social Infrastructure Modify Cognitive Function? A Mixed-Methods Study of Urban-Dwelling Aging Americans** *Journal of Aging and Health, 33*(9), 772–785.

[J.6] Finlay, J. M., Yu, W., **Li, M.**, et al. (2021). **Neighborhood cognitive amenities? A mixed-methods study of intellectually-stimulating places and cognitive function among older Americans**. *Well-being, Space and Society, 2*, 100040.

[J.7] Yu, R., Esposito, M., **Li, M.**, et al. (2023). **Neighborhood "disamenities": Local barriers and cognitive function among Black and White aging adults**. *BMC Public Health* 23, 197 (2023).

[W.1] **Li, M.**, & Conrad, F. (Under review). **Advancing Annotation of Stance in Social Media Posts: A Comparative Analysis of Large Language Models and Crowd Sourcing**. Manuscript submitted for publication in *Journal of Computational Social Science*.

[W.2] **Li, M.**, Chen, X.& Budak, C. (Revise and resubmit). **Depth, Breadth, and Bias: Structural Diffusion of Political Content on Divergent Platforms**. Manuscript invited for revision and resubmission (revision due September 2026) at *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM 2027)*.

[W.3] Conrad, F., **Li, M.**, Schober, M., Gagnon-Bartsch, J., Dolgin, R., Zhou, E., Chen, P. & Beatty, Paul. (Under review). **Capturing public opinion by automatically summarizing social media posts**. Manuscript submitted for publication in *Computers in Human Behavior*. (equal contribution first author)

[W.4] **Li, M.**, & Conrad, F. G. (2026). **Persona-based simulation of human opinion at population scale**. *arXiv*.

[W.5] **Li, M.**, Nurumov, K., Morales S. & Lee, S. **Can Large Language Models Act as Human Respondents? A Cross-Cultural Comparison of Synthetic and Real Survey Responses on Life Expectancy**

[T.1] **Li, M.**, **Assessing the Representativeness of Social Media Data for Measuring Public Opinion about Public Events**.

Conference Talks

[4] **Li, M.** (2026). **From Information Exposure to Debate: Modeling Opinion Change with Multi-Agent Large Language Model (LLM) Simulations**. Presented at the *AAPOR 81st Annual Conference*, Atlanta, GA, USA.

[1] **Li, M.** (2025). **Who posted that? Automatically inferring characteristics of social media users**. Presented at the *AAPOR 80th Annual Conference*, St. Louis, MO, USA & *The 11th Conference of the European Survey Research Association (ESRA)*, Utrecht, The Netherlands.

[2] **Li, M.** (2025). **Exploring cross-cultural social science research with LLM-generated synthetic respondents: Insights from subjective life expectancy probing**. Presented at the *WAPOR Annual Conference*, St. Louis, MO, USA.

[3] **Li, M.** (2024). **Automating self-rated health coding with large language models: Strengths and weaknesses under multilingual survey scenario**. Presented at the *WAPOR Annual Conference*, Seoul, South Korea.

[4] **Li, M.** (2024). **Capturing public opinion by automatically summarizing social media posts**. Presented at the *AAPOR 79th Annual Conference*, Atlanta, GA, USA.

[5] **Li, M.** (2023). **Tracking census self-completion rate using Twitter posts**. Presented at the *AAPOR 78th Annual Conference*, Philadelphia, PA, USA & *47th Annual Conference of the Midwest Association for Public Opinion Research (MAPOR)*, Chicago, IL, USA. & *Michigan Student Symposium for Interdisciplinary Statistical Sciences (MSSISS)*, Ann Arbor, MI, USA.

[6] **Li, M.** (2021). **Effects of interviewer language and dialect on questions about political trust**. Presented at the *46th Annual Conference of the Midwest Association for Public Opinion Research (MAPOR)*, Chicago, IL, USA.

Published Data Contributions

Multiple datasets in ICPSR National Neighborhood Data Archive (NaNDA): broadband access, health care, social services, justice, retail, and more.

Software Packages

- **FastLexRank (Python Package)** *2023*
  Algorithm development [GitHub]
  ◦ Fast, streamlined implementation of the LexRank summarization algorithm.
- **Twitter Browser** *Ongoing*
  Web App [GitHub]
  ◦ Web application to summarize and visualize Twitter data.
- **KRE (R package)** *2023*
  Adaptive kernel ratio estimation [GitHub]
  ◦ Adaptive bandwidth method for inhomogeneous populations.

Skills

- **Programming Languages:** Python, C++, R
- **GIS/Big Data:** ArcGIS, PySpark, Hadoop
- **Data Collection:** Survey, Interviews, Web Scraping
- **Languages:** Mandarin (Fluent), English (Fluent)

Honors and Awards

- **Janet A. Harkness Student Paper Award** *2025*
  World Association for Public Opinion Research (WAPOR)
  ◦ Awarded for outstanding student paper contribution at the WAPOR Annual Conference.
- **Dow Distinguished Award** *2021*
  University of Michigan
  ◦ Awarded for interdisciplinary research on sustainable green infrastructure.

Professional Memberships

- **AAPOR** (American Association for Public Opinion Research) *Member*
- **MAPOR** (Midwest AAPOR) *Member*
- **AAAI** (Association for the Advancement of Artificial Intelligence)) *Member*