LIZHOU (LEO) FAN

lizhouf@umich.edu | Website | Google Scholar | LinkedIn | GitHub | Tableau

EDUCATION

University of Michigan *2021 - Present*
Doctor of Philosophy - PhD, Information
University of California, Los Angeles *2016 - 2020*
Bachelor of Science - BS, Statistics

RESEARCH INTERESTS

Themes: **Data Curation, Archives, Social Media, Anti-Racism, Infometrics**

Fields: Data Management, Computational Social Science, Digital Humanities

PEER-REVIEWED ARTICLES

* Articles in press and under review are respectively marked.

AI-augmented Data Curation & Archiving
with focuses on research data & data discovery

- **Fan, L.**, Yin, Z., Yu, H., & Gilliland, A.J. (2022). Using Machine Learning to Enhance Archival Processing of Social Media Archives. *ACM Journal on Computing and Cultural Heritage (JOCCH)*. doi: 10.1145/3547146.

- **Fan, L.**, Lafia, S., Li, L., Yang, F., & Hemphill, L. (2023). DataChat: Prototyping a Conversational Agent for Dataset Search and Visualization. *Proceedings of the Association for Information Science and Technology (ASIS&T)*. In press.

- **Fan, L.**, Lafia, S., Wofford, M., Thomer, A.K., Yakel, E., & Hemphill, L. (2023). Mining Semantic Relations in Data References to Understand the Roles of Research Data in Academic Literature. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*. In press.

- Lafia, S., **Fan, L.**, & Hemphill, L. (2022). A Natural Language Processing Pipeline for Detecting Informal Data References in Academic Literature. *Proceedings of the Association for Information Science and Technology (ASIS&T)*. doi: 10.1002/pra2.614.

- **Fan, L.** & Presner, T. (2022). Algorithmic Close Reading: Using Semantic Triplets to Index and Analyze Agency in Holocaust Testimonies. *Digital Humanities Quarterly*.

Social Media
with focuses on anti-racism & pro-equity

- **Fan, L.**, Yu, H., & Yin, Z. (2020). Stigmatization in social media: Documenting and analyzing hate speech for COVID19 on Twitter. *Proceedings of the Association for Information Science and Technology (ASIS&T)*. doi: 10.1002/pra2.313.

- **Fan, L.**, Yu, H., Yin, Z., & Gilliland, A.J. (2021) #StopAsianHate: Archiving and Analyzing Twitter Discourse in the Wake of the 2021 Atlanta Spa Shootings. *Proceedings of the Association for Information Science and Technology (ASIS&T)*. doi: 10.1002/pra2.475.

- Yu, H., **Fan, L.**, & Gilliland, A.J. (2022). Disparities and resilience: analyzing online Health information provision, behaviors and needs of LBGTQ+ elders during COVID-19. *BMC Public Health*. doi: 10.1186/s12889-022-14783-5.

- **Fan, L.**, Li, L., & Hemphill, L. (2023). Characterizing Online Toxicity During the 2022 Mpox Outbreak: A Computational Analysis of Topical and Network Dynamics. *Journal of Medical Internet Research (JMIR)*. Under review.

- Li, L., Ma, Z., **Fan, L.**, Lee, S., Yu, H., & Hemphill, L. (2023). ChatGPT in education: A discourse analysis of worries and concerns on social media. *Education and Information Technologies*. In press.

- Li, L., **Fan, L.**, Atreja, S., & Hemphill, L. (2023). "HOT" ChatGPT: The promise of ChatGPT in detecting and discriminating hateful, offensive, and toxic comments on social media. *ACM Transactions on the Web*. Under review.

- Yin, Z.*, **Fan, L.**,* Yu H., & Gilliland A.J. (2020). Using a Three-step Social Media Similarity (TSMS) Mapping Method to Analyze Controversial Speech Relating to COVID-19 in Twitter Collections. *2020 IEEE International Conference on Big Data.*
  Recognition: Archived by the International Network for Hate Studies (INHS).

Infometrics
with focuses on bibliometrics & scientometrics

- **Fan, L.**, Li, L., Ma, Z., Lee, S., Yu, H., & Hemphill, L. (2023). A Bibliometric Review of Large Language Models Research from 2017 to 2023. *ACM Transactions on Intelligent Systems and Technology (TIST)*. Under review.

- Lafia, S., **Fan, L.**, Thomer, A.K., & Hemphill, L. (2022). Subdivisions and Crossroads: Identifying Hidden Community Structures in a Data Archives Citation Network. Quantitative Science Studies. doi: 10.1162/qss_a_00209.

BOOK CHAPTERS

- **Fan, L.**, Yu, H., & Gilliland, A.J. (2023). Aggravated Anti-Asian Hate since COVID-19 and the #StopAsianHate Movement: Connection, Disjointness, and Challenges. In book *Hate Speech on Social Media: A Global Approach.* doi: 10.25768/654-916-9.

- Presner, T. & **Fan, L.** (2024) Algorithmic Close Reading: Analyzing Vectors of Agency in Holocaust Testimonies. In book *Ethics of the Algorithm: Holocaust Memory, the Distant Witness, and the Future of Testimony.* In press, *Princeton University Press.*

WORKSHOPS, PANELS, POSTERS, PRESENTATIONS, & WORKING PAPERS

- **Fan, L.**, Lafia, S., Bleckley, D., Moss, E., Thomer, A.K. Hemphill, L. "Librarian-in-the-Loop: A Natural Language Processing Paradigm for Detecting Informal Mentions of Research Data in Academic Literature". Workshop paper presented at the Investigating Data Work Across Domains Workshop at ACM CHI, 2022.

- **Fan, L.**. "Archival Data Thinking and Practices for Social Media Collections". Talk given at Information Studies Department, UCLA, 2021.

- Presner, T., Bonazzi, A., Deblinger, R., **Fan, L.**, Shepard, D., & Toth, G. "Digital Humanities Methods for Analyzing Holocaust and Genocide Testimonies". Panel presented at the DH 2020 Conference.

- Bonazzi, A.*, **Fan, L.**,* Presner, T. "Language Use and Narrative Structures in Genocide Interviews via Digital Humanities". Poster presented at Early Career Research Conference, Institute for Digital Research and Education, UCLA, 2019.

- **Fan, L.**, Presner, T. "Using Text Mining and Information Extraction Methods to Analyze Holocaust and Genocide Testimonies". Oral presentation nominated by the Digital Humanities Program and presented at Undergraduate Research Showcase, UCLA, 2020.

- **Fan, L.**, "Classification and Clustering for Text Mining". Workshop presented at the Text Analysis Research Cluster (TARC) Session, Digital Research Consortium, UCLA, 2020.

- Hemphill, L., Xing, J., & **Fan, L**. "Comparing Costs for Cloud-based Data Archives". Working paper, 2023.

RESEARCH EXPERIENCE

AI-augmented Data Management in Institutional Repositories: Transforming Research Data Use Detection, Categorization, and Recommendation
Dissertation Research   *2023 -*

- Aims to enhance the efficiency of data management by refining processes related to archiving, curation, and reuse, thereby augmenting the discoverability of research data.

- Aspires to allocate human labor to tasks centered on ensuring precision and context, while employing AI to bolster speed and scalability.

- Additionally, endeavors to develop suitable AI models and pipelines, ranging from rule-based systems to machine learning, deep learning, and large language models, aligning with the aforementioned objectives.

Measuring and Improving the Efficacy of Curation Activities in Data Archives (MICA)
Graduate Researcher   *2021 - 2023*

- Funded by IMLS Grant LG-37-19-0134-19

Developing Evidence-based Data Sharing and Archiving Policies
Graduate Researcher   *2021 - 2022*

- Funded by NSF Grant 1930645

Democratizing Social Media Data for Research Impact
Graduate Researcher   *2022 - 2023*

- Funded by META, through SOMAR, MIDAS, University of Michigan

Documenting and Analyzing Social Media: Anti-Racism and Pro-Equity
Lead Researcher   *2020 - 2023*

- Started with a group of voluntary researchers from UCLA in early 2023, when COVID-19 unfolded and anti-Asian hate escalated.

Ethics of the Algorithm: Holocaust Memory, the Distant Witness, and the Future of Testimony
Data Scientist & Researcher   *2019 - 2023*

- As a part of my pre-doctoral work supported by the UCLA Holocaust Research Lab

Online Statistical Computing Reference (OSCR) for Digital Humanities and Social Sciences
Lead Developer & Researcher   *2019 - 2021*

- Funded by UCLA Ed&IS Digital Resource Development Initiative Project

TEACHING EXPERIENCE

SI 671. Data Mining: Methods and Applications
University of Michigan   *Fall 2023*

DH 199. Capstone: The Online Statistical Computing Reference (OSCR) for digital humanities and social sciences
University of California, Los Angeles *Winter 2020*

REVIEWER AND VOLUNTEER EXPERIENCE

Knowledge Organization
Reviewer *2023 (1)*

IEEE Transactions on Computational Social Systems
Reviewer *2021 (1), 2022 (1)*

Digital Humanities Quarterly
Reviewer *2021 (2), 2022 (1)*

The 86th ASIS&T Annual Meeting
Reviewer *October 27-31, 2023*

Digital Humanities 2022 Conference
Reviewer *July 25-29, 2022*

The 44th ACM SIGIR Conference
Student Volunteer *July 11-15, 2021*

International Data Engineering and Science Association (IDEAS)
Volunteer *Sep 2018 - Feb 2019*

INDUSTRY EXPERIENCE

Procter & Gamble Cincinnati, OH
PhD Data Scientist Intern *May - Aug, 2023*

Data Application Lab Monterey Park, CA
Data Analyst Intern *Sep 2018 - Feb 2019*

FELLOWSHIPS, GRANTS, SCHOLARSHIPS & AWARDS

- AWS Computing Fund for BigANN Competition ($1000), NeuraIPS, 2023

- Travel Fund for JCDL 2023 ($1770), ACM SIGIR, 2023

- Azure Credit Award ($5000), Microsoft Azure & University of Michigan, 2023

- School of Information Research Fund ($2000), University of Michigan, 2021-2025.

- School of Information Fellowship ($11600), University of Michigan, 2021.

- Rackham Conference Travel Grant ($900), University of Michigan, 2021, 2022 & 2023.

- Dean's Prize for Excellence in Research Award ($2000), UCLA, 2020.

- Digital Resource Module Award ($1000), UCLA, 2020.

- Best Poster Award, Early Career Researcher Conference, UCLA, 2019.

SKILLS

- **Research Skills:** Grant Writing, Experimental Design, Qualitative Methods

- **Programming Skills:** Python, R, MySQL, C++, Git, Linux, LaTeX

- **Computing Tools:** TensorFlow, PyTorch, Scikit-learn, NumPy, Pandas, AWS, Azure
- **Data Science and Engineering Skills/Tools:** Deep Learning, Machine Learning, NLP, LLM, OpenAI, WandB, Tableau, Streamlit, Neo4j, ElasticSearch

AFFILIATIONS

- Member of the Association for Computing Machinery (ACM)
- Member of the Association for Information Science and Technology (ASIS&T)

Last Updated: October 3, 2023