LIZHOU (LEO) FAN

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

EDUCATION

University of Michigan *2021 - May 2025 (Expected)*
Doctor of Philosophy - PhD, Information
University of California, Los Angeles *2016 - 2020*
Bachelor of Science - BS, Statistics

RESEARCH INTERESTS

Keywords: **GenAI, LLMs, AI Agents, Auxiliary Diagnosis, Anti-Racisim, Pro-Equity**

Fields of Research: **NLP, Multimodal ML, Multi-agent Systems, Data Archives, Health Informatics, Biomedical Informatics, Social Media, Computational Social Science**

PEER-REVIEWED ARTICLES

* Articles in press and under review are respectively marked.

Large Language Models, Multimodality, and AI Agents

- **Fan, L.**, Hua, W., Li, L., Ling, H., Zhang, Y., & Hemphill, L. (2024) NPHardEval: Dynamic Benchmark on Reasoning Ability of Large Language Models via Complexity Classes. In preparation for submission to the 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024).
- Hua, W., **Fan, L.**, Li, L., Hemphill, L., & Zhang, Y. (2024) War and Peace (WarAgent): Large Language Model-based Multi-Agent Simulation of World Wars. In preparation for submission to the 41st International Conference on Machine Learning (ICML 2024).
- Li, L., **Fan, L.**, Atreja, S., & Hemphill, L. (2024) "HOT" ChatGPT: The promise of ChatGPT in detecting and discriminating hateful, offensive, and toxic comments on social media. *Manuscript accepted by ACM Transactions on the Web.*
- 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.* doi: 10.1007/s10639-023-12256-9

AI-augmented Archiving and Data Curation

- **Fan, L.**, Yin, Z., Yu, H., & Gilliland, A.J. (2022) Using Machine Learning to Enhance Archival Processing of Social Media Archives. *Journal on Computing and Cultural Heritage.* 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).* doi: 10.1002/pra2.820
- **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.* doi: 10.1109/JCDL57899.2023.00039
- 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

Health and Biomedical Informatics

- **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
- 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*. Under review. Manuscript available upon request.
- Li, X., **Fan, L.**, Wu, H., Chen, K., Yu, X., Chao, C., Cai, Z., Niu, X., Cao, A., & Ma, X. (2024) Enhancing Autism Spectrum Disorder Early Detection with the Parent-Child Dyads Block-Play Protocol and a Hybrid Deep Learning Framework. *IEEE Journal of Biomedical and Health Informatics*. Under review. Manuscript available upon request.
- Lv, C., **Fan, L.**, Li, H., Ma, J., Jiang, W., & Ma, X. (2024) Leveraging Multimodal Deep Learning Framework and a Comprehensive Audio-Visual Dataset to Advance Parkinsons Early Detection. *Biomedical Signal Processing and Control*. Under review. Manuscript available upon request.
- Wang, X., **Fan, L.**, Li, H., Jiang, W., Bi, X., & Ma, X. (2024) AttSeqNet: Leveraging Attention-Driven and Time-wise Splitting Seq2seq Model to Enhance Eye Movement Event Detection in Parkinson's Disease. *Biomedical Signal Processing and Control*. Under review. Manuscript available upon request.

Infometrics (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*. 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
- Hemphill, L., Thomer, A., Lafia, S., **Fan, L.**, Bleckley, D., & Moss, E. (2024) A Dataset for Measuring the Impact of Research Data and their Curation. *Scientific Data*. Under review. Manuscript available upon request.

Anti-Racism via Computational Social Science

- **Fan, L.**, Yu, H., & Gilliland, A.J. (2022) 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
- **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
- 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. *Proceedings of the IEEE International Conference on Big Data (Big Data)*. doi: 10.1109/BigData50022.2020.9377930
- **Fan, L.** & Presner, T. (2022) Algorithmic Close Reading: Using Semantic Triplets to Index and Analyze Agency in Holocaust Testimonies. *Digital Humanities Quarterly*.

- 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*. Princeton University Press.

WORKSHOPS, PANELS, POSTERS, & PRESENTATIONS

- Hemphill, L., Xing, J., & **Fan, L.** (2023) Comparing Costs for Cloud-based Data Archives.
- **Fan, L.**, Lafia, S., Bleckly, D., Moss, E., Thomer, A.K., & Hemphill, L. (2022) Librarian-in-the-Loop: A Natural Language Processing Paradigm for Detecting Informal Mentions of Research Data in Academic Literature. Presented at the ACM CHI'22 Workshop on Data Work Across Domains.
- **Fan, L.** (2021) Archival Data Thinking. *An invited talk at an UCLA Ed&IS lecture (Management of Digital Records, Fall 2021).*
- Presner, T., Bonazzi, A., **Fan, L.**, Tth, G., Deblinger, R., & Shepard, D. (2020) Digital Humanities Methods for Analyzing Holocaust and Genocide Testimonies. Presented at Digital Humanities Conference (DH2020).

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.

Benchamarking and Enhancing LLM Reasoning Abilities
Graduate Researcher *2023 -*

- Evaluates foundation models' reasoning ability through dynamic benchmarks with complexity classes, including P (polynomial), NP (non-polynomial)-complete, and NP-hard.
- Improves LLM reasoning ability through multi-agent interactions.
- Extends the scope of evaluation and enhancement to multimodal LLMs, or Large Multi-modal Models (LMMs).

Multi-agent LLM Systems
Graduate Researcher *2023 -*

- Creates LLM and LMM-based multi-agent systems (MAS) for history simulation and historical event emulation.

AI for Disease Diagnostics
Graduate Researcher *2023 -*

- Applies deep learning and large multimodal models to improve disease early detection, such as Autism Spectrum Disorder (ASD) and Parkinson's Disease (PD).

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 507. Data Mining: Methods and Applications
University of Michigan *Winter 2024 Syllabus*

SI 671. Data Mining: Methods and Applications
University of Michigan *Fall 2023 Syllabus—Teaching Evaluation*

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

Data Science Journal
Reviewer *2024 (1)*

Cogent Social Sciences
Reviewer *2024 (1)*

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: January 19, 2024