

# LIZHOU (LEO) FAN

[Google Scholar](#) | [LinkedIn](#) | [GitHub](#) | [Harvard AIM](#) | [Homepage](#)

## EDUCATION

---

**University of Michigan** *Aug 2021 - Aug 2024*  
Doctor of Philosophy - PhD, Information  
**University of California, Los Angeles** *Sep 2016 - Jun 2020*  
Bachelor of Science - BS, Statistics

## APPOINTMENTS

---

**Harvard Medical School** *Sep 2024 - Present*  
Research Fellow  
**Mass General Brigham** *Sep 2024 - Present*  
Postdoctoral Research Fellow, AI in Medicine Program (AIM)

## RESEARCH INTERESTS

---

Artificial Intelligence, Data Technologies, Natural Language Processing, Social Computing, Health Informatics, Medical AI, Science of Science, Social Media, Digital Humanities

## RESEARCH ARTICLES

---

Note: \* marks equal contribution co-first authors; † marks articles in which I serve as a corresponding author; articles in press and under peer-review are indicated, respectively.

### Data Technologies, Data Archives, and Science of Science

- **Fan, L.**, Li, L., Ma, Z., Lee, S., Yu, H., & Hemphill, L. (2024) [A Bibliometric Review of Large Language Models Research from 2017 to 2023](#). *ACM Transactions on Intelligent Systems and Technology*. doi: 10.1145/3664930
- Yu, H.\*, **Fan, L.\***, Li, L., Zhou, J., Ma, Z., Xian, L., Hua, W., He, S., Jin, M., Zhang, Y., Gandhi, A., & Ma, X. (2024) [Large Language Models in Biomedical and Health Informatics: A Review with Bibliometric Analysis](#). *Journal of Healthcare Informatics Research*. doi: 10.1007/s41666-024-00171-8
- 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*. doi: s41597-024-03303-2
- **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
- **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/pr2.820
- Hemphill, L., Xing, J., & **Fan, L.** (2023) [Comparing Costs for Cloud-based Data Archives](#). *Preprint*
- **Fan, L.**<sup>†</sup>, 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*. doi: 10.1145/3547146

- 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/pr2.614
- Lafia, S., **Fan, L.**, Thomer, A.K., & Hemphill, L. (2022) [Subdivisions and Crossroads: Identifying Hidden Community Structures in a Data Archive's Citation Network](#). *Quantitative Science Studies*. doi: 10.1162/qss\_a.00209
- **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](#). *Working paper for the ACM CHI'22 Workshop on Data Work Across Domains*.

### Social Media, Inclusive Language, and Health Informatics

- 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](#). *ACM Transactions on the Web*. doi:10.1145/3643829
- **Fan, L.**<sup>†</sup>, Li, L., & Hemphill, L. (2024) [Characterizing Online Toxicity During the 2022 Mpox Outbreak: A Computational Analysis of Topical and Network Dynamics](#). *Journal of Medical Internet Research*. Accepted.
- 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
- **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
- 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.**, 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/pr2.475
- **Fan, L.**, Yu, H., & Yin, Z. (2020) [Stigmatization in Social Media: Documenting and Analyzing Hate Speech for COVID-19 on Twitter](#). *Proceedings of the Association for Information Science and Technology (ASIS&T)*. doi: 10.1002/pr2.313
- 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

### Social Computing and Digital Humanities

- 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.
- 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](#). Submitted to the Thirteenth International Conference on Learning Representations (ICLR 2025).

- Lin, S., Hua, W., Li, L., Chang, C., **Fan, L.**, Ji, J., Hua, H., Jin, M., Luo, J., & Zhang, Y. & Zhang, Y. (2024) [BattleAgent: Multi-modal Dynamic Emulation on Historical Battles to Complement Historical Analysis](#). *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024) Demo Track*
- **Fan, L.** & Presner, T. (2022) [Algorithmic Close Reading: Using Semantic Triplets to Index and Analyze Agency in Holocaust Testimonies](#). *Digital Humanities Quarterly*.
- Presner, T., Bonazzi, A., **Fan, L.**, Tóth, G., Deblinger, R., & Shepard, D. (2020) [Digital Humanities Methods for Analyzing Holocaust and Genocide Testimonies](#). *Panel Abstract for Digital Humanities Conference (DH2020)*.

## Natural Language Processing, Large Language Models, and Medical AI

- **Fan, L.\***, Hua, W.\*, Li, L., Ling, H., & Zhang, Y. (2024) [NPHardEval: Dynamic Benchmark on Reasoning Ability of Large Language Models via Complexity Classes](#). *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024)*.
- **Fan, L.\***, Hua, W.\*, Li, X.\*, Zhu, K., Jin, M., Li, L., Ling, H., Chi, J., Wang, J., Ma, X., & Zhang, Y. (2024) [NPHardEval4V: A Dynamic Reasoning Benchmark of Multimodal Large Language Models](#). Submitted to the *2025 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2025)*.
- Hua, W.\*, Zhu, K.\*, Li, L., **Fan, L.**, Jin, M., Xue, H., Li, Z., Wang, J., & Zhang, Y. (2024) [Disentangling Logic: The Role of Context in Large Language Model Reasoning Capabilities](#). *Preprint*.
- Yu, H., Zhou, J., Li, L., Chen, S., Gallifant, J., Shi, A., Li, X., Jin, M., Hua, W., Chen, G., Zhou, Y., Li, Z., Gupte, T., Chen, M.-L., Azizi, Z., Zhang, Y., Assimes, T. L., Ma, X., Bitterman, D. S., Lu, L., & **Fan, L.†**, [AIPatient: Simulating Patients with LLM-Powered Agentic Workflows](#). *Journal of the American Medical Association (JAMA)*. Under review.
- Chen, S., Gao, M., Sasse, K., Hartvigsen, T., Anthony, B., **Fan, L.**, Aerts, H., Gallifant, J., & Bitterman, D., [Wait, but Tylenol is Acetaminophen... Investigating and Improving Language Models' Ability to Resist Requests for Misinformation](#). *Preprint*.
- 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 an Attention-enhanced GCN-xLSTM Hybrid Deep Learning Framework](#). *Engineering Applications of Artificial Intelligence*. Under review.
- 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 Parkinson's Early Detection](#). *Biomedical Signal Processing and Control*. doi: 10.1016/j.bspc.2024.106480
- Wang, X.\*, **Fan, L.\***, Li, H.\*, Jiang, W., Bi, X., & Ma, X. (2024) [Skip-AttSeqNet: Leveraging Skip Connection and Attention-Driven Seq2seq Model to Enhance Eye Movement Event Detection in Parkinson's Disease](#). *Biomedical Signal Processing and Control*. doi: 10.1016/j.bspc.2024.106862
- Li, L., Zhou, J., Gao, Z., Hua, W., **Fan, L.**, Yu, H., Hagen, L., Zhang, Y., Assimes, T. L., Hemphill, L., & Ma, S. (2024) [A scoping review of using Large Language Models \(LLMs\) to investigate Electronic Health Records \(EHRs\)](#). *npj Digital Medicine*. Under review.
- Jin, M., Yu, Q., Dong, S., Zhang, C., **Fan, L.**, Hua, W., Zhu, S., Meng, Y., Wang, Z., & Zhang, Y. (2024) [Health-LLM: Personalized Retrieval-Augmented Disease Prediction System](#). Submitted to the 31st International Conference on Computational Linguistics (COLING 2025) System Demonstration Track.

## INVITED TALKS & PRESENTATIONS

---

- **Fan, L.** (2024) Large Language Model-powered Multi-agent Systems for Health and Social Informatics. Invited talk at UCLA Digital Humanities Program and Information Studies Department.
- **Fan, L.** (2024) War and Peace (WarAgent): Large Language Model-based Multi-Agent Simulation of World Wars. Invited talk at the Wolfram Colloquium: LLM Agents for Modeling Group Dynamics
- **Fan, L.** (2024) Advancing Medical AI: Auxiliary Diagnosis, LLM Reasoning, and AI Agents. Job talk at the Artificial Intelligence in Medicine (AIM) Program at Mass General Brigham and Harvard Medical School.
- Hua, W., & **Fan, L.** (2024) War and Peace (WarAgent): Large Language Model-based Multi-Agent Simulation of World Wars. Invited talk at the the BAAI Young Scientist Association (Qingyuan Club).
- **Fan, L.**, & Hua, W. (2023) NPHardEval: Dynamic Benchmark on Reasoning Ability of Large Language Models via Complexity Classes. Invited talk at the BAAI Young Scientist Association (Qingyuan Club).
- **Fan, L.** (2021) [Archival Data Thinking](#). A guest lecture for UCLA Information Studies (Management of Digital Records, Fall 2021).

## RESEARCH EXPERIENCES

---

### **Informatics Strategies to Improve Immune-related Adverse Event Detection in Cancer Patients**

*Postdoctoral Research Fellow*

2024 -

- Funded by NIH R01 Grant [1R01CA294033-01](#)

### **Developing New Technologies for Improving Inclusive Language in Clinical Trial Materials**

*Postdoctoral Research Fellow*

2024 -

- Funded by the Harvard Catalyst Grant, a NIH/NCATS (NIH National Center for Advancing Translational Sciences) funding on the topic [Expanding Inclusivity in Research Studies RFA](#)

### **Generative AI-augmented and User-centric Research Data Discovery and Reuse**

*PhD Dissertation Research*

2023 - 2024

- Structured Scholarly Knowledge Graphs (SKGs) for data discovery
- Leveraged and customized Large Language Models (LLMs) for enhanced data recommendation
- Communicated data recommendations and reuse suggestions to users in both written and visual formats
- [Dissertation Results](#) are partially used to support the application to the NSF/NIH SOS:BIO Grant [Identifying and Encouraging Connections among Data Reuse, Scientific Innovation, and Scientific Careers](#)

### **Large AI Models for Enhancing Clinical Diagnostics and Training**

*Lead Researcher*

2023 - 2024

- Funded by [UMSI PhD Student Research Fund](#)
- Applied deep learning and large multimodal models to improve disease early detection, such as Autism Spectrum Disorder (ASD) and Parkinson's Disease (PD)

- Developed LLM-based multi-agent systems (MAS) for emergency management services (EMS)
- Developed the AIPatient system to facilitate patient simulation for medical education and clinical research

### **Benchmarking LLMs and Developing Intelligent Multi-agent Systems (MAS)**

*Lead Researcher*

*2023 - 2024*

- Funded by [UMSI PhD Student Research Fund](#)
- Evaluated foundation models' reasoning ability through dynamic benchmarks with complexity classes, including P (polynomial), NP (non-polynomial)-complete, and NP-hard
- Extended the scope of evaluation and enhancement to multimodal LLMs, or Large Multi-modal Models (LMMs)
- Created LLM and LMM-based multi-agent systems (MAS) for history simulation and historical event emulation

### **Transforming Data Discovery Through Behavior Modeling and Recommendation**

*Graduate Researcher*

*2021 - 2023*

- Funded by NSF Grant [2121789](#)

### **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*

*2021 - 2023*

- Funded by Meta, through [SOMAR, MIDAS, University of Michigan](#)

### **Documenting and Analyzing Social Media: Anti-Racism and Pro-Equity during COVID**

*Lead Researcher*

*2020 - 2023*

- Formed and led a team of researchers from UCLA in early 2023 to document and research the unfolding of COVID-19 and the escalation of anti-Asian hate

### **Ethics of the Algorithm: Holocaust Memory, the Distant Witness, and the Future of Testimony**

*Data Scientist & Pre-doctoral Researcher*

*2019 - 2021*

- Conducted pre-doctoral research 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 EXPERIENCES**

---

### **SI 507. Intermediate Programming**

*University of Michigan*

*Winter 2024 Syllabus*

**SI 671. Data Mining: Methods and Applications**

*University of Michigan*

*Fall 2023 Syllabus*

**DH 199. Capstone: The Online Statistical Computing Reference (OSCR) for digital humanities and social sciences**

*University of California, Los Angeles*

*Winter 2020 Syllabus*

**PROFESSIONAL SERVICES**

---

**Frontiers in Research Metrics and Analytics**

*Review Editor, Editorial Board*

*2024 - Present*

Section: Emerging Technologies and Transformative Paradigms in Research

**The 25th International conference on Web Information Systems Engineering**

*Member, Program Committee*

*December 2-5, 2024*

WISE-2024

**The 1st International Workshop on Large Language Models Applications in Medical Informatics**

*Member, Program Committee*

*November 26-29, 2024*

LLMs4MI2024, co-located with the 2nd International Conference on Foundation and Large Language Models (FLLM2024)

**The 44th ACM SIGIR Conference**

*Assistant Session Facilitator*

*July 11-15, 2021*

**International Data Engineering and Science Association (IDEAS)**

*Organizer*

*Sep 2018 - Feb 2019*

**REVIEWER EXPERIENCES**

---

\* The number in parenthesis after the year indicate the number of manuscripts reviewed.

**Social Network Analysis and Mining**

*Reviewer*

*2024 (4)*

**Scientific Reports**

*Reviewer*

*2024 (3)*

**Data Science Journal**

*Reviewer*

*2024 (2)*

**JMIR Research Protocols**

*Reviewer*

*2024 (1)*

**Journal of Computational Social Science**

*Reviewer*

*2024 (1)*

**PeerJ Computer Science**

*Reviewer*

*2024 (2)*

**Discover Artificial Intelligence**

*Reviewer*

*2024 (1)*

<b>Language Resources and Evaluation</b> <i>Reviewer</i>	2024 (1)
<b>Humanities and Social Sciences Communications</b> <i>Reviewer</i>	2024 (1)
<b>European Heart Journal Digital Health</b> <i>Reviewer</i>	2024 (1)
<b>Cogent Social Sciences</b> <i>Reviewer</i>	2024 (1)
<b>Knowledge Organization</b> <i>Reviewer</i>	2023 (1)
<b>IEEE Transactions on Computational Social Systems</b> <i>Reviewer</i>	2021 (1), 2022 (1)
<b>Digital Humanities Quarterly</b> <i>Reviewer</i>	2021 (2), 2022 (1)
<b>The 87th ASIS&amp;T Annual Meeting</b> <i>Reviewer</i>	October 25-29, 2024 (6)
<b>The 86th ASIS&amp;T Annual Meeting</b> <i>Reviewer</i>	October 27-31, 2023 (5)
<b>Digital Humanities 2022 Conference</b> <i>Reviewer</i>	July 25-29, 2022 (3)

## INDUSTRY EXPERIENCES

---

<b>Procter &amp; Gamble</b> <i>PhD Data Scientist Intern</i>	Cincinnati, OH May - Aug, 2023
<b>Data Application Lab</b> <i>Data Analyst Intern</i>	Monterey Park, CA Sep 2018 - Feb 2019

## FELLOWSHIPS, GRANTS, SCHOLARSHIPS & AWARDS

---

- AWS Computing Fund for BigANN Competition (\$1000), NeuralIPS, 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.
- Wiley Top Downloaded Article Award, John Wiley & Sons, Inc., 2021
- 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.

## AFFILIATIONS

---

- Member of the Association for Computing Machinery (ACM)
- Member of the Association for Computational Linguistics (ACL)
- Member of the American Medical Informatics Association (AMIA)
- Member of the Association for Information Science and Technology (ASIS&T)

## REFERENCES

---

- **Libby Hemphill, PhD**  
*Associate Professor*  
Director, Resource Center for Minority Data, ICPSR  
Director, Social Media Archive, ICPSR  
School of Information, University of Michigan  
Email: libbyh@umich.edu  
Phone: (734) 763-2285
- **Anne Gilliland, PhD**  
*Professor*  
Director, Center for Information as Evidence  
Graduate School of Education & Information Studies  
University of California, Los Angeles  
Email: gilliland@gseis.ucla.edu  
Phone: (310) 825-8799
- **Yongfeng Zhang, PhD**  
*Associate Professor*  
Department of Computer Science, Rutgers University  
Email: yongfeng.zhang@rutgers.edu  
Phone: (848) 445-8309
- **Todd Presner, PhD**  
*Chair and Professor*  
Advisor, Office of the Vice Chancellor for Research and Creative Activities  
Department of European Languages and Transcultural Studies  
University of California, Los Angeles  
Email: presner@ucla.edu  
Phone: (310) 794-6051

Last Updated: Oct 30, 2024