

LIZHOU FAN

[Google Scholar](#) | [LinkedIn](#) | [GitHub](#)

EDUCATION

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

APPOINTMENTS

The Chinese University of Hong Kong (CUHK)	Starting Aug 2025
Vice-Chancellor Assistant Professor, Department of Psychiatry, Faculty of Medicine	
Harvard Medical School / Mass General Brigham	Sep 2024 - Mar 2025
Postdoctoral Research Fellow, AI in Medicine Program	

RESEARCH INTERESTS

Technical Interests: **Medical AI, NLP, Digital Health, Health Informatics, Data Management**
Medical Field: **Psychiatry**

PEER-REVIEWED ARTICLES

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

Medical AI, Digital Health, and Health Informatics

- 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 Parent-child Dyads Block-play Protocol and Attention-enhanced Hybrid Deep Learning Framework](#). *Engineering Applications of Artificial Intelligence*. doi: 10.1016/j.engappai.2025.110430
- Wang, X.*, **Fan, L.***, Li, H.*, Jiang, W., Bi, X., & Ma, X. (2025) [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
- 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
- Fan, L.**[†], Li, L., & Hemphill, L. (2024) [Toxicity on Social Media During the 2022 Mpox Public Health Emergency: Quantitative Study of Topical and Network Dynamics](#). *Journal of Medical Internet Research*. doi: 10.2196/52997
- 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
- 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 LGBTQ+ Elders During COVID-19](#). *BMC Public Health*. doi: 10.1186/s12889-022-14783-5

- **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/pra2.313
- 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

Natural Language Processing, Large AI Models, and Multi-agent AI Systems

- Lin, S., Hua, W., Wang, Z., Jin, M., **Fan, L.**, & Zhang, Y. (2025) [EmojiPrompt: Generative Prompt Obfuscation for Privacy-Preserving Communication with Cloud-based LLMs](#). *Proceedings of the 2025 Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL 2025)*. doi: 10.18653/v1/2024.acl-long.225
- **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)*. doi: 10.18653/v1/2024.acl-long.225
- **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
- 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*. doi: 10.18653/v1/2024.emnlp-demo.18
- **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/pra2.820
- 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

Data Management, Data Archives, Data Science, and Social Computing

- 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
- 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.
- 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
- 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.[†]**, 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

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

SELECTED PREPRINTS & ARTICLES UNDER REVIEW

- 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.**[†] (2025) [AIPatient: Simulating Patients with EHRs and LLM Powered Agentic Workflow](#). *Communications Medicine*. Under Review.
- Kim, Y., Jeong, H., Chen, S., Li, S. S., Lu, M., Alhamoud, K., Mun, J., Grau, C., Jung, M., Gameiro, R., **Fan, L.**, Park, E., Lin, T., Yoon, J., Yoon, W., Sap, M., Tsvetkov, Y., Liang, P., Xu, X., Liu, X., McDuff, D., Lee, H., Park, H. W., Tulebaev, S., & Breazeal, C. (2025) [Medical Hallucination in Foundation Models and Their Impact on Healthcare](#). Preprint.
- 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.
- Chen, S., Gao, M., Sasse, K., Hartvigsen, T., Anthony, B., **Fan, L.**, Aerts, H. J. W. L., Gallifant, J., & Bitterman, D. S. (2024). [The Risks of Medical Misinformation Generation in Large Language Models](#). *Lancet Digital Health*. Under review.
- Sun, J., Lu, T., Shao, X., Han, Y., Xia, Y., Zheng, Y., Wang, Y., Li, X., Ravindran, A., **Fan, L.**, Fang, Y., Lu, L., Zhang, X., Ravindran, N., Wang, Y., & Lu, L. (2024). Practical AI application in psychiatry: Historical review and future directions. *Molecular Psychiatry*. 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](#). Preprint.
- **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](#). Preprint.
- Hua, W.^{*}, Zhu, K.^{*}, Li, L., **Fan, L.**, Jin, M., Xue, H., Li, Z., Wang, J., & Zhang, Y., Li, L., Hemphill, L., & Zhang, Y. (2024) [Disentangling Logic: The Role of Context in Large Language Model Reasoning Capabilities](#). Preprint.
- Hua, W., Liu, O., Li, L., Amayuelas, A., Chen, J., Jiang, L., Jin, M., **Fan, L.**, Sun, F., Wang, W., Wang, X., & Zhang, Y. (2024) [Game-theoretic LLM: Agent Workflow for Negotiation Games](#). Preprint.
- Hua, W.^{*}, **Fan, L.**^{*}, Li, L., Hemphill, L., & Zhang, Y. (2023) [War and Peace \(WarAgent\): Large Language Model-based Multi-Agent Simulation of World Wars](#). Preprint.

INVITED TALK & PRESENTATIONS

- **Fan, L.** (2025) AIPatient: Simulating Patients with EHRs and LLM Powered Agentic Workflow. *An invited talk at University of Massachusetts Lowell, Miner School of Computer & Information Sciences*
- 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 EHRs and LLM Powered Agentic Workflow](#). Presented at the AAAI 2025 Workshop on Advancing LLM-Based Multi-Agent Collaboration.
- **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. Presented at the Wolfram Colloquium: LLM Agents for Modeling Group Dynamics
- **Fan, L.** (2024) Advancing Medical AI: Auxiliary Diagnosis, LLM Reasoning, and AI Agents. Presented to 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. Presented to 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. Presented to the the BAAI Young Scientist Association (Qingyuan Club).
- **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.**, Tóth, G., Deblinger, R., & Shepard, D. (2020) [Digital Humanities Methods for Analyzing Holocaust and Genocide Testimonies](#). Presented at Digital Humanities Conference (DH2020).

RESEARCH EXPERIENCES

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

Postdoctoral Research Fellow

2024 - 2025

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

Developing New Technologies for Improving Inclusive Language in Clinical Trial Materials

Postdoctoral Research Fellow

2024 - 2025

- 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

ADVISEES

Wanxin Xie

Undergraduate Student Advisee at University of California, Los Angeles

Now: MS in Data Science, Harvard University

Jun 2020 - July 2021

Michelle Lee

Undergraduate Student Advisee at University of California, Los Angeles

Now: Senior Analyst at Analysis Group

Jun 2020 - July 2021

Jinkui Chi

Master Student Advisee at University of Michigan

Now: MS in Information Science at University of Michigan

Sep 2023 - May 2024

REVIEWER EXPERIENCES

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

ACM Computing Surveys

Reviewer 2025 (1)

IEEE Signal Processing Magazine

Reviewer 2025 (1)

BMC Public Health

Reviewer 2025 (1)

BMJ Oncology

Reviewer 2025 (1)

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)

Language Resources and Evaluation

Reviewer 2024 (1)

Humanities and Social Sciences Communications

Reviewer 2024 (1)

European Heart Journal Digital Health

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), 2025 (1)

Digital Humanities Quarterly*Reviewer**2021 (2), 2022 (1)***The 87th ASIS&T Annual Meeting***Reviewer**October 25-29, 2024 (6)***The 86th ASIS&T Annual Meeting***Reviewer**October 27-31, 2023 (5)***Digital Humanities 2022 Conference***Reviewer**July 25-29, 2022 (3)*

INDUSTRY EXPERIENCES**Procter & Gamble***PhD Data Scientist Intern**Cincinnati, OH**May - Aug, 2023***Data Application Lab***Data Analyst Intern**Monterey Park, CA**Sep 2018 - Feb 2019*

FELLOWSHIPS, GRANTS, SCHOLARSHIPS & AWARDS

- AWS Computing Fund for BigANN Competition (\$1000), NeurIPS, 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.

Last Updated: Apr 8, 2024