

Xiaochen Wang

📞 919-360-3666 | ✉️ xcwang@psu.edu | 📍 E348 Westgate Building, University Park, PA

RESEARCH INTERESTS

Multimodal Learning, Natural Language Processing, Information Retrieval, Health Informatics

EDUCATION

The Pennsylvania State University

Aug. 2022 – May. 2027

Ph.D. in Informatics

University Park, PA

Supervisor: Dr. Fenglong Ma

University of North Carolina at Chapel Hill

Aug. 2020 – May 2022

M.S. in Information Science

Chapel Hill, NC

Supervisor: Dr. Yue Wang

Central South University

Sep. 2016 – Jun. 2020

B.E. in Automation (Honors Program)

Changsha, China

Supervisor: Dr. Fan Guo

INTERNSHIP

Search ML Team, Instacart

May 2023 – Aug. 2023

Machine Learning Research Intern

Mentor: Dr. Taesik Na

Renaissance Computing Institute, UNC

May 2022 – Aug. 2022

Research Specialist

Mentor: Dr. Yue Wang

PUBLICATIONS

- [1] **Xiaochen Wang**(2024). *Developing Multimodal Healthcare Foundation Model: From Data-driven to Knowledge-enhanced*. Accepted by AAAI 2025 Doctoral Consortium, to appear.
- [2] **Xiaochen Wang**^{*1}, Jiaqi Wang*, Houping Xiao, Jinghui Chen, Fenglong Ma (2024). *FEDKIM: Adaptive Federated Knowledge Injection into Medical Foundation Models*. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 8141-8154).
- [3] Jiaqi Wang*, **Xiaochen Wang***, Lingjuan Lyu, Jinghui Chen, Fenglong Ma (2024). *FEDMEKI: A Benchmark for Scaling Medical Foundation Models via Federated Knowledge Injection*. Accepted by NeurIPS 2024 as Spotlight, to appear.
- [4] **Xiaochen Wang**, Junyu Luo, Jiaqi Wang, Yuan Zhong, Xiaokun Zhang, Yaqing Wang, Parminder Bhatia, Cao Xiao, Fenglong Ma (2024). *Unity in Diversity: Collaborative Pre-training Across Multimodal Medical Sources*. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL) (Volume 1: Long Papers) (pp. 3644-3656).
- [5] **Xiaochen Wang**, Xiao Xiao, Ruhan Zhang, Xuan Zhang, Taesik Na, Tejaswi Tenneti, Haixun Wang, Fenglong Ma (2024). *Mitigating Pooling Bias in E-commerce Search via False Negative Estimation*. In Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) (pp. 5917-5925).

¹Equal Contribution

- [6] Yuan Zhong, **Xiaochen Wang**, Jiaqi Wang, Xiaokun Zhang, Yaqing Wang, Mengdi Huai, Cao Xiao, Fenglong Ma (2024). *Synthesizing Multimodal Electronic Health Records via Predictive Diffusion Models*. In Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) (pp. 4607-4618).
- [7] Jiaqi Wang, Junyu Luo, Muchao Ye, **Xiaochen Wang**, Yuan Zhong, Aofei Chang, Guanjie Huang, Ziyi Yin, Cao Xiao, Jimeng Sun, Fenglong Ma (2024). *Recent Advances in Predictive Modeling with Electronic Health Records*. Accepted by IJCAI 2024, to appear.
- [8] Xiaokun Zhang, Bo Xu, Zhaochun Ren, **Xiaochen Wang**, Hongfei Lin and Fenglong Ma (2024). *Disentangling ID and Modality Effects for Session-based Recommendation*. In Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR) (pp. 1883-1892).
- [9] Yuan Zhong, Suhan Cui, Jiaqi Wang, **Xiaochen Wang**, Ziyi Yin, Yaqing Wang, Houping Xiao, Mengdi Huai, Ting Wang, Fenglong Ma (2024). *MedDiffusion: Boosting Health Risk Prediction via Diffusion-based Data Augmentation*. In Proceedings of the 2024 SIAM International Conference on Data Mining (SDM) (pp. 499-507). Society for Industrial and Applied Mathematics.
- [10] Luo, Junyu, **Xiaochen Wang**, Jiaqi Wang, Aofei Chang, Yaqing Wang, and Fenglong Ma (2024). *CoRelation: Boosting Automatic ICD Coding Through Contextualized Code Relation Learning*. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024) (pp. 3997-4007).
- [11] **Xiaochen Wang**, Junyu Luo, Jiaqi Wang, Ziyi Yin, Suhan Cui, Yuan Zhong, Yaqing Wang, Fenglong Ma (2023). *Hierarchical Pretraining on Multimodal Electronic Health Records*. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 2839-2852).
- [12] Junyu Luo, Cheng Qian, **Xiaochen Wang**, Lucas Glass and Fenglong Ma (2023). *pADR: Towards Personalized Adverse Drug Reaction Prediction*. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM) (pp. 4724-4730).
- [13] Lei Wang, Jose Luis Ambite, Abhishek M. Appaji, Janine Bijsterbosch, Jérôme Dockès, Rick Herrick, Alex Kogan, Howard M. Lander, Petra Lenzini, Daniel Marcus, Stephen M. Moore, J- B Poline, Arcot Rajasekar, Satya S. Sahoo, Matthew D. Turner, **Xiaochen Wang**, Yue Wang, Jessica A. Turner(2023). *NeuroBridge: A Prototype Platform for Discovery of The Long-Tail Neuroimaging Data Authors*. *Frontiers in Neuroinformatics*, 17, 1215261.
- [14] **Xiaochen Wang**, Yue Wang (2022). *Sentence-Level Resampling for Named Entity Recognition*. The 2022 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL) (pp. 2151-2165).
- [15] **Xiaochen Wang**, Yue Wang, José-Luis Ambite, Abhishek Appaji, Howard Lander, Stephen M. Moore, Arcot K. Rajasekar, Jessica A. Turner, Matthew D. Turner, Lei Wang, Satya S. Sahoo (2022). *Enabling Scientific Reproducibility through FAIR Data Management: An ontology-driven deep learning approach in the NeuroBridge Project*. *American Medical Informatics Association 2022 Annual Symposium (AMIA)* (Vol. 2022, p. 1135).
- [16] Jingwen Hou*, **Xiaochen Wang***, Jean-Jacques Dubois, R. Byron Rice, Amanda Haddock, Yue Wang (2022). *Resource: Extreme Systematic Reviews: A Large Literature Screening Dataset to Support Environmental Policymaking*. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management (CIKM) (pp. 4029-4033).

SUBMISSION

- [1] Yuan Zhong, **Xiaochen Wang**, Jiaqi Wang, Xiaokun Zhang, Fenglong Ma (2025). *MedDiTPro: A Prompt-Guided Diffusion Transformer for Multimodal Longitudinal Medical Data Synthesis*. Submitted to KDD'25.

TUTORIAL

- [1] Jiaqi Wang, **Xiaochen Wang**, Yuan Zhong, Ziyi Yin, Aofei Chang, Cao Xiao, Fenglong Ma (2024). *Multimodal Artificial Intelligence in Healthcare*. AAAI 2025 Tutorial and Lab Forum.

TEACHING

Teaching Assistant

DS 340W: Applied Data Science

2022 - 2025

Instructor: Dr. Kaamran Raahemifar

SERVICE

Conference PC Member | Reviewer

- CIKM'23, AAAI'24, AAAI'25, EMNLP'24, AMIA'24, NAACL'24, AISTATS'25, PAKDD'25, KDD'25, IJCAI'25

Journal Article Reviewer

- Progress in Artificial Intelligence, Heliyon, Human-centric Computing and Information Sciences

Volunteer

- KDD'24, EMNLP'24

PRESENTATIONS & TALKS

- [1] *Developing Multimodal Healthcare Foundation Model: From Data-driven to Knowledge-enhanced Poster Presentation at AAAI 2025.*
- [2] *Multimodal Artificial Intelligence in Healthcare Tutorial at AAAI 2025.*
- [3] *FEDKIM: Adaptive Federated Knowledge Injection into Medical Foundation Models Poster Presentation at EMNLP 2024.*
- [4] *Unity in Diversity: Collaborative Pre-training Across Multimodal Medical Sources Poster Presentation at ACL 2024.*
- [5] *Mitigating Pooling Bias in E-commerce Search via False Negative Estimation Oral & Poster Presentation at KDD 2024*
- [6] *Synthesizing Multimodal Electronic Health Records via Predictive Diffusion Models Oral & Poster Presentation at KDD 2024*
- [7] *Hierarchical Pretraining on Multimodal Electronic Health Records Poster Presentation at EMNLP 2023*
- [8] *Sentence-Level Resampling for Named Entity Recognition Poster presentation at NAACL 2022*
- [9] *Enabling Scientific Reproducibility through FAIR Data Management: An ontology-driven deep learning approach in the NeuroBridge Project Oral Presentation at AMIA 2022*
- [10] *Boosting Search Embedding via Cross Encoder Oral Presentation at Instacart*

SELECTED HONORS

Awards

- Student Travel Award for KDD 2024
- SIGIR Student Travel Grant for CIKM 2022
- Graduate Programs Travel Award for AMIA 2022

- Third Prize in The Chinese College Student Intelligent Design Competition
- Central South University Academic Scholarship

Media Coverage

- *Can Machine Learning Improve Air Quality?*