

# Xiaochen Wang

☎ 919-360-3666 | ✉ xcwang@psu.edu | 📍 E348 Westgate Building, State College, PA

## RESEARCH INTERESTS

---

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

## EDUCATION

---

<b>The Pennsylvania State University</b> <i>Ph.D. in Informatics</i>	Aug. 2022 – May. 2027 State College, PA
<b>University of North Carolina at Chapel Hill</b> <i>M.S. in Information Science</i>	Aug. 2020 – May 2022 Chapel Hill, NC
<b>Central South University</b> <i>B.E. in Automation (Honors Program)</i>	Sep. 2016 – Jun. 2020 Changsha, China

## EXPERIENCE

---

<b>Search ML Team, Instacart</b> <i>Machine Learning Research Intern</i>	May 2023 – Aug. 2023 San Francisco, CA
<b>College of Information Science and Technology, PSU</b> <i>Graduate Researcher   Graduate Teaching Assistant</i>	Aug. 2022 – Now State College, PA
<b>Renaissance Computing Institute, UNC</b> <i>Graduate Research Assistant   Research Specialist</i>	Aug. 2021 – Aug. 2022 Chapel Hill, NC
<b>School of Information and Library Science, UNC</b> <i>Graduate Research Assistant</i>	Aug. 2020 – Aug. 2021 Chapel Hill, NC
<b>Central South University</b> <i>Undergraduate Research Assistant</i>	May 2018 – Dec. 2018 Changsha, China

## PUBLICATIONS

---

- [1] **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*. Accepted by The 62nd Annual Meeting of the Association for Computational Linguistics (ACL'24), to appear.
- [2] **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*. Accepted by 2024 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'24), to appear.
- [3] 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*. Accepted by 2024 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'24), to appear.
- [4] 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'24, to appear.
- [5] Xiaokun Zhang, Bo Xu, Zhaochun Ren, **Xiaochen Wang**, Hongfei Lin and Fenglong Ma (2024). *Disentangling ID and Modality Effects for Session-based Recommendation*. Accepted by SIGIR 2024, to appear.
- [6] Luo, Junyu, **Xiaochen Wang**, Jiaqi Wang, Aofei Chang, Yaqing Wang, and Fenglong Ma (2024). *CoRelation: Boosting Automatic ICD Coding Through Contextualized Code Relation Learning*. Accepted by LREC-COLING 2024, to appear.

- [7] **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).
- [8] 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).
- [9] Yuan Zhong, Suhan Cui, Jiaqi Wang, **Xiaochen Wang**, Ziyi Yin, Yaqing Wang, Houping Xiao, Mengdi Huai, Ting Wang, Fenglong Ma (2023). *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] 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.
- [11] **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).
- [12] **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).
- [13] Jingwen Hou\*, **Xiaochen Wang**\*<sup>1</sup>, 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).

## SUBMISSIONS

---

- [1] **Xiaochen Wang**\*, Jiaqi Wang\*, Houping Xiao, Jinghui Chen, Fenglong Ma (2024). *FEDKIM: Adaptive Federated Knowledge Injection into Medical Foundation Models*. Submitted to EMNLP'24.
- [2] Jiaqi Wang\*, **Xiaochen Wang**\*, Lingjuan Lyu, Jinghui Chen, Fenglong Ma (2024). *FEDMEKI: A Benchmark for Scaling Medical Foundation Models via Federated Knowledge Injection*. Submitted to NeurIPS'24.

## TEACHING

---

DS 340W Applied Data Science

2022 - 2024

## SERVICE

---

Conference PC Member | Reviewer

- CIKM 2023, AAAI 2024, AAAI 2025, EMNLP 2024

Journal Article Reviewer

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

Volunteer

- KDD 2024

---

<sup>1</sup>Equal Contribution

## 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?*