

PAIHENG XU

Email: paiheng@cs.umd.edu ◇ Homepage: paihengxu.github.io ◇ Telephone: (410) · 900 · 3657 ◇ Github: [paihengxu](https://github.com/paihengxu)

RESEARCH INTEREST

Natural Language Processing, Computational Social Science

EDUCATION

University of Maryland, College Park, USA

2021 - present

Ph.D. Student in Computer Science

Johns Hopkins University, Baltimore, USA

2018 - 2020

M.S.E in Computer Science

Southwest University, Chongqing, China

2014 - 2018

B.E in Computer Science

RESEARCH EXPERIENCE

UNIVERSITY OF MARYLAND, COLLEGE PARK

College Park, MD

Research Assistant – Advisor: Wei Ai

Sep. 2021 - present

- Discover abstract and implicit patterns from (textual) data with LLMs.
- AI for Education - Identify effective teaching practices & measure teaching qualities with multimodal data.
- Study how social media conversations influence and reflect pro-social behaviors, such as public health engagement.

JOHNS HOPKINS UNIVERSITY

Baltimore, MD

Full-time Research Developer – Advisor: Mark Dredze

Jun. 2020 - Aug. 2021

- Track and analyze mobility change in response to social distancing using Twitter data. [\[link\]](#)
- Discovery, semantic shift of e-cigarette brands from social media.

Research Assistant – Mentor: Zach Wood-Doughty

Nov. 2018 - May 2020

- Demographics inference (Python package [Demographer](#)), linguistics difference for racial groups on Twitter.
- Developed a [reproducible method](#) to collect large-scale dataset using noisy self-report lexical co-occurrence.

UNIVERSITY OF ELECTRONIC SCIENCE & TECHNOLOGY

Chengdu, China

Research Intern on Human Mobility – Advisor: Tao Zhou

Jul. 2017 - Jun. 2018

- Verified and fixed the overestimation of the predictability of human mobility trajectory.
- Showed the intrinsic deviation of using Lempel-Ziv entropy to estimate predictability of highly random time series.

INTERNSHIP

ADOBE RESEARCH

San Jose, CA

Research Scientist Intern – Mentor: Vishy Swaminathan, Gang Wu, and Xiang Chen

May 2024 - Aug. 2024

- Skill development through code generation and offline simulations using LLM-based agents.

ADOBE RESEARCH

Remote

Research Scientist Intern – Mentor: Tong Yu and Haoliang Wang

May 2022 - Aug. 2022

- Learning knowledge from Critical Service Outage reports for auto-remediation.
- Adaptive Question Generation for Extractive Question Answering using In-context Learning.

SELECTED PUBLICATIONS

* denotes equal contribution

Paiheng Xu, Jing Liu, Nathan Jones, Julie Cohen, Wei Ai, The Promises and Pitfalls of Using Language Models to Measure Instruction Quality in Education. *NAACL 2024*.

Yuhang Zhou, **Paiheng Xu**, Xiaoyu Liu, Bang An, Wei Ai, Furong Huang, Explore Spurious Correlations at the Concept Level in Language Models for Text Classification. *ACL 2024*.

Paiheng Xu, Louiqa Raschid, Vanessa Frias-Martinez, Does Geo-co-location Matter? A Case Study of Public Health Conversations during COVID-19. *arXiv:2405.17710*, 2024.

Xiaoyu Liu*, **Paiheng Xu***, Junda Wu, Jiaxin Yuan, Yifan Yang, Yuhang Zhou, Fuxiao Liu, Tianrui Guan, Haoliang Wang, Tong Yu, Julian McAuley, Wei Ai, Furong Huang, Large Language Models and Causal Inference in Collaboration: A Comprehensive Survey. *Findings of NAACL 2025*.

Paiheng Xu, David A. Broniatowski, Mark Dredze, Twitter Social Mobility Data Reveal Demographic Variations in Social Distancing Practices During the COVID-19 Pandemic. *Scientific Reports*, 2024.

Paiheng Xu*, Yuhang Zhou*, Bang An, Wei Ai, Furong Huang, GFairHint: Improving Individual Fairness for Graph Neural Networks via Fairness Hint. *Transactions on Knowledge Discovery from Data*, 2025

Yuhang Zhou, **Paiheng Xu**, Xiyao Wang, Xuan Lu, Ge Gao, Wei Ai, Emojis Decoded: Leveraging ChatGPT for Enhanced Understanding in Social Media Communications. *ICWSM*, 2025.

Yuhang Zhou*, Jing Zhu*, **Paiheng Xu**, Xiaoyu Liu, Xiyao Wang, Danai Koutra, Wei Ai, Furong Huang, Multi-Stage Balanced Distillation: Addressing Long-Tail Challenges in Sequence-Level Knowledge Distillation. *Findings of EMNLP 2024*.

Zach Wood-Doughty*, **Paiheng Xu***, Xiao Liu, Mark Dredze, Using Noisy Self-Reports to Predict Twitter User Demographics. *SocialNLP@NAACL*, 2021.

Paiheng Xu, Mark Dredze, David A. Broniatowski, The Twitter Social Mobility Index: Measuring Social Distancing Practices from Geolocated Tweets. *Journal of Medical Internet Research*, 2020.

Paiheng Xu, Likang Yin, Zhongtao Yue, Tao Zhou, On Predictability of Time Series. *Physica A: Statistical Mechanics, its Applications* 523, 345-351 (2019).

Paiheng Xu, Rong Zhang, Yong Deng, A Novel Visibility Graph Transformation of Time Series into Weighted Networks. *Chaos, Solitons & Fractals* 117, 201-208 (2018).

TECHNICAL SKILLS

Languages	Python, MATLAB, R, C/C++, C#, L ^A T _E X, Markdown, Bash
Tools	Git, PyTorch, Tensorflow, Scikit-learn, NLTK, Spacy, Jupyter Notebook, MySQL

HONOR & AWARDS

Dean's Fellowship, University of Maryland
National Scholarship, Ministry of Education, China
Outstanding Graduates, Southwest University
Meritorious Winner, Interdisciplinary Contest in Modeling (ICM)

TEACHING & SERVICE

Teaching Assistant: Introduction to Artificial Intelligence, Fall 2022
Program Committee: ICWSM 2022, 2025
Reviewer: ARR 2025; KDD 2025; TheWebConf 2025
Sub-Reviewer: ACL 2021; KDD 2022; SIGIR 2022-2025