# PAIHENG XU

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

## RESEARCH INTEREST

Natural Language Processing, Computational Social Science, Causal Inference

## 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 & Jing Liu — *Sep. 2021 - present*

- Discover latent and **distributional patterns** in text by combining LLMs with **causality-inspired** frameworks.
- Investigate **bias** and **spurious correlation** mitigation for text **classification** and LLM **alignment**.
- AI for Edu - LLM-based transcript analysis to identify effective teaching practices and measure teaching quality.

**Research Assistant** – Advisor: Louiqa Raschid & Vanessa Frias-Martinez — *Jan. 2023 - present*

- Model correlations between geo-co-location and social media engagement using **statistical** & **LLM**-based methods.
- Link online interaction patterns with offline community outcomes to promote pro-social behaviors.

JOHNS HOPKINS UNIVERSITY — Baltimore, MD
**Full-time Research Developer** – Advisor: Mark Dredze — *Jun. 2020 - Aug. 2021*

- Develop Twitter-based social mobility indices to study demographic differences in social distancing behavior. [link]
- Apply embedding- and transformer-based models to discover e-cig brands and track sentiment shifts on social media.

**Research Assistant** – Mentor: Zach Wood-Doughty — *Nov. 2018 - May 2020*

- Demographic inference on Twitter with distant supervision (Package); study linguistic variation across racial groups.

## INDUSTRY EXPERIENCE

ADOBE RESEARCH — San Jose, CA
**Research Scientist Intern** – Mentor: Haoliang Wang — *May 2025 - Aug. 2025*

- Design style understanding and **personalized** suggestion generation through LVLM-based design pattern uncovering.
- **User-centric evaluation** for personalization using paired image comparisons to capture design preferences.

ADOBE RESEARCH — San Jose, CA
**Research Scientist Intern** – Mentor: Vishy Swaminathan, Gang Wu, & Xiang Chen — *May 2024 - Aug. 2024*

- **Skill development** for software scripting through code generation & offline simulations using **LLM-based agents**.

ADOBE RESEARCH — Remote
**Research Scientist Intern** – Mentor: Tong Yu & 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**, Gang Wu, Xiang Chen, Tong Yu, Chang Xiao, Franck Dernoncourt, Tianyi Zhou, Wei Ai, Vishy Swaminathan, Skill Discovery for Software Scripting Automation via Offline Simulations with LLMs. *arXiv:2504.20406*, 2025.

**Paiheng Xu***, Xiaoyu Liu*, 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***, 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*

**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.

**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.

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.

**Paiheng Xu***, Zach Wood-Doughty*, 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#, LaTeX, Markdown, Bash |
| **Tools** | Transformers, PyTorch, TRL, HF Agents, Scikit-learn, NLTK, Statsmodels, NetworkX |

## 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; TPAMI
**Sub-Reviewer:** ACL 2021; KDD 2022; SIGIR 2022-2025