Xiaoxuan Wang

(312)937-5271 | xw27@cs.ucla.edu

Education:

University of Illinois Urbana-Champaign *Auguest,2018*--May, 2022
B.S. of Computer Science in Grainger Engineering College
Minor in Mathematics James Scholar Honor Dean List

University of California Los Angeles *Sept, 2022--Present*
Ph.D. of Computer Science
Advisor: Wei Wang

Manuscripts:

- ***SciBench: Evaluating College-Level Scientific Problem-Solving Abilities of Large Language Model***
  Xiaoxuan Wang*, Ziniu Hu*, Pan Lu*, Yanqiao Zhu*, Jieyu Zhang,
  Satyen Subramaniam, Arjun R. Loomba, Shichang Zhang, Yizhou Sun, Wei Wang
  ICML 2024
  Media Covered by Nature News Feature

- ***CliBench: Multifaceted Evaluation of Large Language Models in Clinical Decisions on Diagnoses, Procedures, Lab Tests Orders and Prescriptions***
  Mingyu Derek Ma, Chenchen Ye, Yu Yan, **Xiaoxuan Wang**, Peipei Ping, Timothy S Chang, Wei Wang

- ***Memorize and Rank: Evaluating Large Language Models For Clinical Diagnosis Prediction***
  Mingyu Derek Ma, Xiaoxuan Wang, Yijia Xiao, Anthony Cuturrufo, Wei Wang
  NeurIPS GenAI4Health, 2024 and *AAAI 2024 Spring Symposium*

- ***STAR: Boosting Low-Resource Event Extraction by Structure-to-Text Data Generation with Large Language Models***
  Mingyu Derek Ma, **Xiaoxuan Wang**, Po-Nien Kung, P.Jefferey Brantingham, Nanyun Peng, Wei Wang
  AAAI 2024

- ***Learning under Label Proportions for Text Classification***
  Jatin Chauhan, **Xiaoxuan Wang**, Wei Wang
  EMNLP -Findings 2023

- ***Global Responses to the COVID-19 Pandemic: A Case Study of Spatiotemporal Evidence Finding and Verification***
  Rotem Dror, **Xiaoxuan Wang**, Dan Roth

- **Seamless Equal Accuracy Ratio for Inclusive CTC Speech Recognition**
  Heting Gao, **Xiaoxuan Wang**, Sunghun Kang, Rusty Mina, Dias Issa, John Harvill, Leda Sarı, Mark Hasegawa-Johnson, Chang D. Yoo
  Speech Communication 2022

Internship Experience:

Amazon AWS June 2024— Sep 2024
Mentored by Boran Han, Applied Scientist Intern

- Implement and evaluate various Reinforcement Learning Human Feedback (RLHF) methods, such as REST-EM, Iterative DPO and its variants (KTO, cDPO, IPO, DPO+NLL, etc.), PPO, to determine the most effective approach for integrating tools in problem-solving capabilities of Large Language Models (LLMs).
- Conducted experiments demonstrating that REST-EM exhibits greater sensitivity to data quality and quantity compared to DPO and its variants. Observed that self-generated answers may introduce noise, and applying regularization loss(IPO, cDPO) enhances tolerance in the iterative pipeline. Incorporated NLL loss to mitigate the effects of chosen log-likelihood.

Research Experience:

SciBench: Scientific Computing Problems Benchmark March 2023— July 2023

Advised by Professor **Wei Wang** and Professor **Yizhou Sun**, University of California Los Angeles

- Create a college-level dataset comprised of challenging scientific problems in chemistry, physics, and mathematics, offering a higher degree of difficulty compared to currently available datasets.
- Evaluate the performance and constraints of Large Language Models using the created datasets and design a self-assessment and error detection mechanism.
- Design an evaluation protocol that systemically and automatically examines the problem-solving abilities of Large Language Model.

Low-Resource Event Extraction Using LLM February 2023— June 2023

Advised by Professor **Wei Wang**, University of California Los Angeles

- Devise the STAR method, a structure-to-text data generation technique, utilizing Large Language Models for low-resource event extraction tasks.
- Implement fine-grained step-by-step instructions, with a self-reflection and self-refinement mechanism for error identification and quality enhancement.
- Validate through experiments that STAR outperforms human-curated data point in specific scenarios, significantly boosting low-resource event extraction performance.

Global Government Response to Covid-19 Entailment May 2021—May 2022

Advised by Professor **Dan Roth**, University of Pennsylvania

- Develop temporal tagger to extract temporal expressions from document and normalize them to the consistent annotation format
- Use state-of-art ranking systems to match government policy and news article with spatial and temporal inference
- Apply different state-of-art entailment methods to the top-ranking articles.

Fairness Speech Recognition August 2020—December 2021

Advised by Professor **Mark Hasegawa-Johnson**, University of Illinois Urbana Champaign

- Cooperate with Korea Advanced Institute of Science and Technology
- Help develop a novel inclusiveness measure for ASR System integrated with the standard CTC training pipeline to lower accuracy gap between protected attributes
- Help create novel multi-dialect dataset of ASR by combining existing corpora in seven dialects