Curriculum Vitae

Yuancheng Xu
Email: ycxu@umd.edu
Website: https://yuancheng-xu.github.io/

Education

University of Maryland, College Park
Ph.D. student in Applied Mathematics & Statistics, and Scientific Computation *2020-2025 (expected)*
Advisor: Prof. Furong Huang (Computer Science)

Southern University of Science and Technology, China *2016-2020*
B.S. in Mathematics and Applied Mathematics
GPA 3.94/4.00 (1/909)
Summa Cum Laude (1%)

New York University *Spring, 2019*
Visiting Student at the Courant Institute of Mathematical Sciences

Research Interests

My research focuses on large language model alignment, AI safety and fairness, as well as enhancing the reasoning and planning capabilities of AI systems. Additionally, I work on controllable video generation, aiming to recreate the world through representations of multimodal content and user-model interactions.

Publications and Preprints

1. **Yuancheng Xu**, Jiarui Yao, Manli Shu, Yanchao Sun, Zichu Wu, Ning Yu, Tom Goldstein, Furong Huang. " Shadowcast: Stealthy Data Poisoning Attacks Against Vision-Language Models". In *Neural Information Processing Systems (NeurIPS)*, 2024.

2. Bang An, Sicheng Zhu, Ruiyi Zhang, Michael-Andrei Panaitescu-Liess, **Yuancheng Xu**, Furong Huang. "Automatic Pseudo-Harmful Prompt Generation for Evaluating False Refusals in Large Language Models". In *Conference on Language Modeling (COLM)*, 2024

3. **Yuancheng Xu**, Chenghao Deng, Yanchao Sun, Ruijie Zheng, Xiyao Wang, Jieyu Zhao, Furong Huang. "Adapting Static Fairness to Sequential Decision-Making: Bias Mitigation Strategies towards Equal Long-term Benefit Rate". In *International Conference on Machine Learning (ICML)*, 2024.

4. Bang An, Mucong Ding, Tahseen Rabbani, Aakriti Agrawal, **Yuancheng Xu**, Chenghao Deng, Sicheng Zhu, Abdirisak Mohamed, Yuxin Wen, Tom Goldstein, Furong Huang. "Benchmarking the Robustness of Image Watermarks". In *International Conference on Machine Learning (ICML)*, 2024.

5. Xiyao Wang, Yuhang Zhou, Xiaoyu Liu, Hongjin Lu, **Yuancheng Xu**, Feihong He, Jaehong Yoon, Taixi Lu, Gedas Bertasius, Mohit Bansal, Huaxiu Yao, Furong Huang. "Mementos: A Comprehensive Benchmark for Multimodal Large Language Model Reasoning over Image Sequences". In *Association for Computational Linguistics (ACL)*, 2024.

6. Mucong Ding, Bang An, **Yuancheng Xu**, Anirudh Satheesh, Furong Huang. "SAFLEX: Self-Adaptive Augmentation via Feature Label Extrapolation". In *International Conference on Learning Representations (ICLR)*, 2024.

7. Xiaoyu Liu, Jiaxin Yuan, Bang An, **Yuancheng Xu**, Yifan Yang, Furong Huang. "C-Disentanglement: Discovering Causally-Independent Generative Factors under an Inductive Bias of Confounder". In *Neural Information Processing Systems (NeurIPS)*, 2023.

8. **Yuancheng Xu**, Yanchao Sun, Micah Goldblum, Tom Goldstein, Furong Huang. "Exploring and Exploiting Decision Boundary Dynamics for Adversarial Robustness". In *International Conference on Learning Representations (ICLR)*, 2023.

9. **Yuancheng Xu**, Yanchao Sun, and Furong Huang. "Everyone Matters: Customizing the Dynamics of Decision Boundary for Adversarial Robustness". In *International Conference on Machine Learning (ICML) Workshop on Continuous Time Perspectives in Machine Learning*, 2022.

10. **Yuancheng Xu**, Athanasse Zafirov, R. Michael Alvarez, Dan Kojis, Min Tan, and Christina M. Ramirez. "FREEtree: a Tree-Based Approach for High Dimensional Longitudinal Data with Correlated Features". *Preprint*, 2020.

Research Experience

- **AI Research Intern** — Netflix Eyeline Studios

Advisor: Ning Yu — *Sept 2024 - 2025 (on-going)*

  - Controllable video generation

- **AI Research Intern** — JPMorgan Chase & Co.

Advisor: Dr. Udari Madhushani Sehwag and *Alec Koppel* — *June – Sept 2024*

  - Test-time alignment of large language model

- **Machine Learning Research Intern** — Comcast Applied AI

Advisor: Dr. Mahmudul Hasan — *June – Sept 2023*

  - Scene-text understanding via vision-language models

- **Research Intern** — University of California, Los Angeles

Cross-disciplinary Scholars in Science and Technology (CSST) Program — *May – Sept 2019*

Advisor: Prof. Christina Ramirez (Biostatistics)

  - Tree-based Methods for Longitudinal Analysis

- **Research Intern** — New York University

Undergraduate Research program *May – Sept 2018*

Advisor: Prof. Sukbin Lim (Neuroscience)

- Computational Mechanisms for Working Memory

Awards

Dean's fellowship, University of Maryland, College Park	2020,2021
Summa Cum Laude at Southern University of Science and Technology (10/1000)	2020
China National Scholarship (0.2%, Highest honor of Chinese undergraduate students)	2019
National Mathematical Olympiad (National Second Prize)	2015