Xiyuan Yang

Ph.D. Student @ University of Illinois Urbana-Champaign

xiyuany4@illinois.edu · (+1) 217 766 5798 · Google Scholar (Citations 280+) · GitHub

Research Summary

- I am a first-year **CS Ph.D.** student at the **University of Illinois Urbana-Champaign**. My research focuses on foundation models (LLMs/VLMs/Agents), specifically in terms of scalability, predictability, and trustworthiness.

Education

- **Ph.D. in Computer Science** 2025-2030
  University of Illinois Urbana-Champaign Illinois, USA
  Advisor: Prof. Jingrui He

- **B.Eng. in Computer Science and Technology** 2021-2025
  Wuhan University Hubei, China
  Advisor: Prof. Mang Ye

Experience

- **Microsoft Research (Asia)**, Social Computing Group 2024-2025
  Research Intern, Advisor: Fangzhao Wu
  Project 1: Detecting Prompt Injections via Representation Engineering (in EMNLP 25)
  Project 2: Identifying Cross-lingual Inequalities of Foundation Models (in PNAS 25)
  Project 3: Defending Jailbreak Attacks by Vision Language Models (under review of Nature Communications)
  Project 4: Measuring Human Contributions in AI-Generated Content (under review of ACL)
  Competition: Foundation Model based News Recommendation (Ranked 1/157 groups)

- **University of Chicago**, Department of Computer Science 2024
  Research Intern, Advisor: Prof. Tian Li
  Project 1: Differentially Private Distributed Optimization (under review of ICLR)

Publication

- **Latent Collaboration in Multi-Agent Systems** [Huggingface Daily Paper #1]
  Jiaru Zou*, **Xiyuan Yang***, Ruizhong Qiu, Gaotang Li, Katherine Tieu, Pan Lu, Ke Shen, Hanghang Tong, Yejin Choi, Jingrui He, James Zou, Mengdi Wang, Ling Yang
  ArXiv Preprint 2025

- **Preconditioning Neural Tangent Kernel for Adaptive Optimization**
  Xiyuan Yang, Wenxuan Bao, Katherine Tieu, Jingrui He
  Under Review 2025

- **Defending LLMs Against Jailbreak Attacks Utilizing Cross-Modality Generalization Gap**
  Xiyuan Yang*, Chenglong Wang*, Haoyu Tang*, Yueqi Xie, Bin Zhu, Lingjuan Lyu, Mang Ye, Fangzhao Wu
  Under Review 2025

- **Uncovering Inequalities in New Knowledge Learning by Large Language Models Across Different Languages**
  Chenglong Wang, Haoyu Tang, **Xiyuan Yang**, Yueqi Xie, Jina Suh, Sunayana Sitaram, Junming Huang, Yu Xie, Pengjun Zhao, Zhaoya Gong, Xing Xie, Fangzhao Wu
  PNAS 2025

- **Defending Against Indirect Prompt Injection by Instruction Detection**
  Tongyu Wen*, Chenglong Wang*, **Xiyuan Yang***, Haoyu Tang, Yueqi Xie, Lingjuan Lyu, Zhicheng Dou, Fangzhao Wu
  EMNLP 2025

- **Differentially Private Federated Clustering with Random Rebalancing**
  Xiyuan Yang, Shengyuan Hu, Soyeon Kim, Tian Li
  Under Review 2025

- **Measuring Human Contribution in AI-Assisted Content Generation**
  Yueqi Xie, Tao Qi, Jingwei Yi, **Xiyuan Yang**, Ryan Whalen, Junming Huang, Qian Ding, Yu Xie, Xing Xie, Fangzhao Wu
  Under Review 2024

- **FedAS: Bridging Inconsistency in Personalized Federated Learning**
  Xiyuan Yang, Wenke Huang, Mang Ye
  CVPR 2024

- **Dynamic Personalized Federated Learning with Adaptive Differential Privacy**
  Xiyuan Yang*, Wenke Huang*, Mang Ye
  NeurIPS 2023

- **Robust Heterogeneous Federated Learning under Data Corruption**
  Xiuwen Fang, Mang Ye, **Xiyuan Yang**
  ICCV 2023

Community Contribution

- **PFLlib**: Personalized and Distributed Optimization Library — Github Stars **2000+**
  Core Contributor
- **LatentMAS**: Latent Communication of Multi-agent System — Github Stars **600+**
  Core Contributor and Co-first Author
- **Conference Review**: NeurIPS, ICML, ICLR, AAAI, AISTATS, CVPR, ICCV
- **Journal Review**: IEEE TMC, IEEE TNNLS, IEEE TIFS, IEEE TKDE, IEEE TDSC, Info. Fusion

Scholarship

- PhD Student Fellowship (first-year), University of Illinois Urbana-Champaign 2025
- Star of Tomorrow Award, Microsoft Research 2025
- Overseas exchange and study scholarship, Wuhan University 2024
- Leijun Computer Science Research Scholarship, Wuhan University 2024

Skills

- **Machine Learning**: PyTorch, CUDA, HuggingFace Transformers, vLLM, NumPy
- **Systems & Tools**: Linux, Distributed Systems, Docker, Git, LaTeX