

XIAOCHENG YANG

Email: xy61@illinois.edu | Main Page: yxc-cyber.github.io

EDUCATION

University of Illinois Urbana-Champaign

Master of Science in Computer Science

Related Coursework:

Advanced Topics in NLP, Numerical Analysis, Text Information Systems

Aug 2024 - Expected May 2026

Current GPA: 4.0/4.0

New York University Shanghai

Bachelor of Science in Computer Science, Minor in Mathematics

Related Coursework:

Natural Language Processing, Reinforcement Learning, Computer Vision

Aug 2020 - May 2024

GPA: 3.92/4.0, In-Major GPA: 4.0/4.0

Study Away:

New York University, New York, USA

Sep 2022 - Dec 2022

PUBLICATIONS

- Kunlun Zhu, Hongyi Du, Zhaochen Hong, **Xiaocheng Yang**, ..., Xiangru Tang, Heng Ji, Jiaxuan You. 2025. [MultiAgentBench: Evaluating the Collaboration and Competition of LLM agents](#).
- Cheng Qian, Peixuan Han, Qinyu Luo, ..., **Xiaocheng Yang**, Denghui Zhang, Yunzhu Li, Heng Ji. 2024. [EscapeBench: Pushing Language Models to Think Outside the Box](#).
- Vardhan Dongre, **Xiaocheng Yang**, Emre Can Acikgoz, Suvodip Dey, Gokhan Tur, Dilek Hakkani-Tur. 2024. [ReSpAct: Harmonizing Reasoning, Speaking, and Acting](#).
- Xiaocheng Yang**, Bingsen Chen, and Yik-Cheung Tam. 2024. [Arithmetic Reasoning with LLM: Prolog Generation & Permutation](#). In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers), pages 699–710, Mexico City, Mexico. Association for Computational Linguistics.
- Bingsen Chen, Peiyang Wu, **Xiaocheng Yang**, Yik-Cheung Tam, Hongyi Wen. 2023. [Dialogue State Tracking using Large Language Models](#).

RESEARCH EXPERIENCE

Siebel School of Computing and Data Science, University of Illinois Urbana-Champaign

Graduate Research Assistant | Advisor: [Prof. Gokhan Tur](#)

Aug 2024 - Present

- Reproduced LLM ReAct framework for the task-oriented dialogue domain
- Conducted prompt engineering, by introducing friction rules in the prompt, to promote machine-user interaction and mitigate ambiguity in task-oriented dialogues; Observed an increment in inform and success scores on the task-oriented dialogue benchmark MultiWOZ by 5.5% and 3% respectively when proper rules are given in the prompt
- Built and experimented with multi-agent systems for task-oriented dialogues, involving a verifier, friction generator, and ReAct agent; Observed improved inform scores when multi-agents are properly involved

Department of Computer Science, New York University Shanghai

Undergraduate Research Assistant | Advisor: [Prof. Yik-Cheung \(Wilson\) Tam](#)

Jun 2023 – May 2024

- Finetuned LoRAs for large language models to investigate model mathematic reasoning ability on the arithmetic reasoning benchmark GSM8K; Experimented with different output settings, including Chain-of-Thought, Prolog generation, and combinations of both output strategies; Found that prolog generation outstripped other strategies; Open-sourced the GSM8K-Prolog dataset
- Experimented with data augmentations on Prolog data; Observed a 10.9% margin of accuracy on the GSM8K test set and a 22.6% margin on the GSM-HARD test set over the Chain-of-Thought baseline with Prolog permutation strategy; Investigated the divergence between cross entropy loss and the actual accuracy of Prolog codes

- Finetuned LoRAs for large language models to solve dialogue state tracking (DST) using the task-oriented dialogue benchmark MultiWOZ; Experimented with different LoRA settings, model scales, data scales, and different output settings, including slot-level QA and JSON format; Achieved a new SOTA in end-to-end DST methods, obtaining an 82.4% joint goal accuracy

Department of Economics, New York University Shanghai

Undergraduate Research Assistant | Advisor: [Prof. Guodong Chen](#), [Prof. Yu Zhou](#) Jun 2022 – Dec 2023

- Conducted quasi-experimental study on the impact of the school enrollment lottery policies on household expenditures
- Experimented with data cleaning strategies, including outlier filtering, winsorization, and logarithms
- Utilized propensity score matching and difference-in-difference method and observed declining household education expenditures after the policy was released

HONORS AND AWARDS

- **Summa Cum Laude**, New York University Shanghai May 2024
- **Honors in Computer Science**, New York University Shanghai May 2024
- **NYU Shanghai Excellence Award**, New York University Shanghai May 2024
- **Dean's List for 2022-2023 Academic Year**, New York University Shanghai May 2023
- **Dean's List for 2021-2022 Academic Year**, New York University Shanghai May 2022
- **Dean's List for 2020-2021 Academic Year**, New York University Shanghai May 2021

TECHNICAL SKILLS

Programming Languages: Python, LaTeX, SQL, R, Java

Machine Learning Libraries: PyTorch, Huggingface Python library, Peft, Torchrun, Deepspeed, Pandas, NumPy, Matplotlib, Scikit-Learn

Spoken Languages: Mandarin Chinese (Native), English (TOEFL 104)