Fangyuan Xu

carrie.xfy@gmail.com 609-582-8635 Website GitHub Google Scholar

Education

New York University *Jan 2025 - 2026 (Expected)*
Ph.D. in Computer Science
- Advisor: Eunsol Choi; Research interest: Natural Language Processing
- GPA: 4.0/4.0.

The University of Texas at Austin *Sep 2022 - Dec 2024*
Ph.D. in Computer Science
- Advisor: Eunsol Choi; Research interest: Natural Language Processing
- Transferred to NYU.

Cornell University *Aug 2019 - May 2020*
Master of Engineering (Computer Science)
- GPA: 4.06/4.30 (Overall), 4.1/4.3 (CS)

The University of Hong Kong *Sep 2013 - June 2017*
Bachelor of Engineering (Computer Science with Minor in French)
- GPA: 3.73/4.3 (Overall), 3.9/4.3 (CS)
- Graudated with First Class Honor.

Publications

[1] DP-RFT: Learning to Generate Synthetic Text via Differentially Private Reinforcement Fine-tuning *In submission*
Fangyuan Xu, Sihao Chen, Zinan Lin, Taiwei Shi, Sydney Graham, Pei Zhou, Mengting Wan, Alexander Stein, Virginia Estellers, Charles Chen, Tadas Baltrusaitis, Richard Speyer, Jennifer Neville, Eunsol Choi, Longqi Yang

[2] SAGE: Steerable Agentic Data Generation for Deep Search with Execution Feedback EACL 2026 (Findings)
Fangyuan Xu, Rujun Han, Yanfei Chen, Zifeng Wang, I-Hung Hsu, Jun Yan, Vishy Tirumalashetty, Eunsol Choi, Tomas Pfister, Chen-Yu Lee
paper link

[3] RefreshKV: Updating Small KV Cache During Long-form Generation ACL 2025
Fangyuan Xu, Tanya Goyal*, Eunsol Choi*
paper link

[4] RECOMP: Improving Retrieval-Augmented LMs with Compression and Selective Augmentation ICLR 2024
Fangyuan Xu, Weijia Shi, Eunsol Choi
paper link

[5] KIWI: A Dataset of Knowledge-Intensive Writing Instructions for Answering Research Questions ACL 2024 (Findings)
Fangyuan Xu, Kyle Lo, Luca Soldaini, Bailey Kuehl, Eunsol Choi, David Wadden
paperlink

[6] Understanding Retrieval Augmentation for Long-Form Question Answering COLM 2024
Hung-Ting Chen, **Fangyuan Xu***, Shane Arora*, Eunsol Choi
paperlink

[7] Long-form Answers to Visual Question from Blind and Low Vision People — COLM 2024 (Spotlight)

Mina Huh, **Fangyuan Xu**, Yi-Hao Peng, Chongyan Chen, Hansika Murugu, Danna Gurari, Eunsol Choi, Amy Pavel

paperlink

[8] Contrastive Learning to Improve Retrieval for Real-world Fact Checking — EMNLP 2024 FEVER Workshop

Aniruddh Sriram, **Fangyuan Xu**, Eunsol Choi, Greg Durrett

paperlink

[9] A Critical Evaluation of Evaluations for Long-form Question Answering — ACL 2023 (Oral)

Fangyuan Xu*, Yixiao Song*, Mohit Iyyer, Eunsol Choi

paperlink

[10] Concise Answers to Complex Questions: Summarization of Long-form Answers — ACL 2023

Abhilash Potluri*, **Fangyuan Xu***, Eunsol Choi

paperlink

[11] Modeling Exemplification in Long-form Question Answering via Retrieval — NAACL 2022 (Oral)

Shufan Wang, **Fangyuan Xu**, Eunsol Choi, Mohit Iyyer

paperlink

[12] How Do We Answer Complex Questions: Discourse Structure of Long-form Answers — ACL 2022 (Oral)

Fangyuan Xu, Jessy Li, Eunsol Choi

paperlink

** denotes equal contribution.*

Research Experience

Google Cloud AI Research — *Sunnyvale, CA*
Student Researcher — *May 2025 - Oct 2025*

- Mentor: Rujun Han
- Worked on synthetic data generation for training search agents with reinforcement learning, published paper [2] at EACL 2026.

Microsoft Office of Applied Research — *Redmond, Washington*
Research Intern — *Feb 2025 - May 2025*

- Mentor: Sihao Chen
- Worked on differentially private synthetic data generation via reinforcement learning, paper [1] in submission.

Allen Institute for AI — *Redmond, Washington*
Research Intern — *May 2023 - Aug 2023*

- Mentor: David Wadden
- Worked on evaluation for LLMs on multi-turn instruction following for document grounded writing task, published paper [5] at ACL 2024.

Work Experience

Twitter — *San Francisco, CA*
Machine Learning Engineer II — *July 2020 - Aug 2021*

Morgan Stanley *Hong Kong*
Technology Associate *Aug 2017 - June 2019*

Morgan Stanley *Hong Kong*
Technology Analyst Intern *June 2016 - Aug 2016*

IOIO Creative *Hong Kong*
Software engineering Intern *July 2015 - Aug 2015*

Teaching Experience

CS395T Topics in NLP *Fall 2023*
Teaching Assistant

- Instructor: Eunsol Choi; Course Website

Services

Reviewer: ARR, ICML, ICLR, COLING, COLM, COLING
Program Committee: PC Assitant COLM 2025, In2Writing workshop 2025, LCFM workshop 2024, KnowledgeLM workshop 2024, ME-FoMO Workshop 2024
Organizer: ML^2 meeting @ NYU (Spring 2025), NLL meeting @ UT Austin (Fall 2021)

Mentoring

Yuhan Liu, NYU Ph.D. student *May 2025 - Present*
Aniruddh Sriram, UT Austin undergraduate student; Published [8] at FEVER. *Jan - June 2024*
Abhilash Potluri, UT Austin undergraduate student; Published [10] at ACL. *Jan - May 2022*

Invited Talks

"KIWI: A Dataset of Knowledge-Intensive Writing Instructions for Answering Research Questions" *Nov 2024*
Samaya AI (hosted by Ashwin Paranjape)