

Fangyuan Xu

[✉ fx2079@nyu.edu](mailto:fx2079@nyu.edu) [📞 609-582-8635](tel:609-582-8635) [🔗 Website](#) [🔗 GitHub](#) [🔗 Google Scholar](#)

Education

New York University

Ph.D. in Computer Science

*Jan 2025 - 2026
(Expected)*

- Advisor: [Eunsol Choi](#); Research interest: Natural Language Processing
- GPA: 4.0/4.0.

The University of Texas at Austin

Ph.D. in Computer Science

Sep 2022 - Dec 2024

- Advisor: [Eunsol Choi](#); Research interest: Natural Language Processing
- Transferred to NYU.

Cornell University

Master of Engineering (Computer Science)

Aug 2019 - May 2020

- GPA: 4.06/4.30 (Overall), 4.1/4.3 (CS)

The University of Hong Kong

Bachelor of Engineering (Computer Science with Minor in French)

Sep 2013 - June 2017

- GPA: 3.73/4.3 (Overall), 3.9/4.3 (CS)
- Graduated with First Class Honor.

Publications

[1] Steerable Agentic Data Generation for Deep Search with Execution Feedback

In submission

Fangyuan Xu, Rujun Han, Yanfei Chen, Zifeng Wang, I-Hung Hsu, Jun Yan, Vishy Tirumalashetty, Eunsol Choi, Tomas Pfister, Chen-Yu Lee

[2] 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

[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 | Mina Huh, Fangyuan Xu , Yi-Hao Peng, Chongyan Chen, Hansika Murugu, Danna Gurari, Eunsol Choi, Amy Pavel | COLM 2024 (Spotlight) |
| paperlink ↗ | | |
| [8] Contrastive Learning to Improve Retrieval for Real-world Fact Checking | Aniruddh Sriram, Fangyuan Xu , Eunsol Choi, Greg Durrett | EMNLP 2024 FEVER Workshop |
| paperlink ↗ | | |
| [9] A Critical Evaluation of Evaluations for Long-form Question Answering | Fangyuan Xu *, Yixiao Song*, Mohit Iyyer, Eunsol Choi | ACL 2023 (Oral) |
| paperlink ↗ | | |
| [10] Concise Answers to Complex Questions: Summarization of Long-form Answers | Abhilash Potluri*, Fangyuan Xu *, Eunsol Choi | ACL 2023 |
| paperlink ↗ | | |
| [11] Modeling Exemplification in Long-form Question Answering via Retrieval | Shufan Wang, Fangyuan Xu , Eunsol Choi, Mohit Iyyer | NAACL 2022 (Oral) |
| paperlink ↗ | | |
| [12] How Do We Answer Complex Questions: Discourse Structure of Long-form Answers | Fangyuan Xu , Jessy Li, Eunsol Choi | ACL 2022 (Oral) |
| paperlink ↗ | | |

* denotes equal contribution.

Research Experience

| | |
|--|----------------------------|
| Google Cloud AI Research | <i>Sunnyvale, CA</i> |
| <i>Student Researcher</i> | <i>May 2025 - Oct 2025</i> |
| ○ Mentor: Rujun Han ↗ | |
| ○ Worked on synthetic data generation for training search agents with reinforcement learning, paper [1] in submission. | |
| Microsoft Office of Applied Research | <i>Redmond, Washington</i> |
| <i>Research Intern</i> | <i>Feb 2025 - May 2025</i> |
| ○ Mentor: Sihao Chen ↗ | |
| ○ Worked on differentially private synthetic data generation via reinforcement learning, paper [2] in submission. | |
| Allen Institute for AI | <i>Redmond, Washington</i> |
| <i>Research Intern</i> | <i>May 2023 - Aug 2023</i> |
| ○ 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 | <i>San Francisco, CA</i> |
| <i>Machine Learning Engineer II</i> | <i>July 2020 - Aug 2021</i> |

Morgan Stanley
Technology Associate

Hong Kong
Aug 2017 - June 2019

Morgan Stanley
Technology Analyst Intern

Hong Kong
June 2016 - Aug 2016

IOIO Creative
Software engineering Intern

Hong Kong
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 Assistant COLM 2025, In2Writing workshop 2025, LCFM workshop 2024, KnowledgeLM workshop 2024, ME-FoMO Workshop 2024

Organizer: ML² 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)