Cathy Jiao

cljiao@cs.cmu.edu | cathyjiao@github.io | (412) 720 8665

EDUCATION

Ph.D. Carnegie Mellon University   September 2023 – Present
Language Technologies Institute, School of Computer Science
Advisor: Chenyan Xiong

M.S. Carnegie Mellon University   August 2021- August 2023
Language Technologies Institute, School of Computer Science
Advisors: Maxine Eskenazi, Aaron Steinfeld

B.S. **University of British Columbia**   December 2019
Major: Computer Science and Mathematics (graduated with distinction)

RESEARCH INTERESTS

I work on data attribution methods for large language models (i.e., mapping training data to model outputs), with applications for dataset curation and data valuation.

EXPERIENCE

Graduate Research Assistant
CX Research Group (advised by Chenyan Xiong)   2023 – Present
- *Research areas*: data attribution for LLMs, data-centric evaluation, training data valuation/pricing.
DialRC Group, Carnegie Mellon University (advised by Maxine Eskenazi, Aaron Steinfeld)   2021 –2023
- *Research areas*: dialogue systems, conversational NLP, interaction, accessibility.

Data Scientist   2020
Motorola Solutions, Language-Driven Learning Team   Vancouver, Canada
- Adapted semantic data curation methods to create a large-scale text-to-SQL dataset for video search.
- Developed NLP solutions for public safety, including 911 dispatch data extraction and summarization

NLP Research Intern   Summer 2018, Summer 2019
Motorola Solutions, Language-Driven Learning Team   Vancouver, Canada
- Developed an unsupervised method for converting text to security database queries (patent granted).
- Created a voice assistant for video security navigation (received the President's Project Award '19).
- Detected toxic speech from video intercoms in real time (received the AI/ML Project Award '18).

Software Engineer Intern   Spring, Summer 2017
Kongsberg Maritime   *Vancouver, Canada*
- Developed a C# sensor emulator that generated and transmitted pseudo-GPS metadata patterns.
- Designed MATLAB tests for beamform processing on sonar images.

PUBLICATIONS AND PREPRINTS

*= equal contribution

Fairshare Data Pricing for Large Language Models
Luyang Zhang, **Cathy Jiao***, Beibei Li, Chenyan Xiong*
Preprint. Under Review.

On the Feasibility of In-Context Probing for Data Attribution
***Cathy Jiao**, Gary Gao, Aditi Raghunathan, Chenyan Xiong*
NAACL 2025

Examining Prosody in Spoken Navigation Instructions for People with Disabilities.
Cathy Jiao, Aaron Steinfeld, Maxine Eskenazi.
NAACL 2024, Third Workshop on Bridging Human-Computer Interaction and Natural Language Processing

Understanding the Effectiveness of Very Large Language Models on Dialog Evaluation.
*Jessica Huynh, **Cathy Jiao**, Prakhar Gupta, Shikib Mehri, Payal Bajaj, Vishrav Chaudhary, Maxine Eskenazi.*
IWSDS 2023

Improving Zero and Few-shot Generalization in Dialogue through Instruction Tuning
*Prakhar Gupta, **Cathy Jiao**, Yi-Ting Yeh, Shikib Mehri, Maxine Eskenazi, and Jeffrey P. Bigham*
EMNLP 2022

Improving compositional generalization for multi-step quantitative reasoning in question answering
*Armineh Nourbakhsh, **Cathy Jiao**, Sameena Shah, and Carolyn Rosé*
EMNLP 2022

The DialPort Tools
Cathy Jiao, Jessica Huynh*, Shikib Mehri*, Maxine Eskenazi*
SIGDIAL 2022

ET tu, CLIP? Addressing Common Object Errors for Unseen Environments
Cathy Jiao, Ye Won Byun*, Shahriar Noroozizadeh*, Jimin Sun*, and Rosa Vitiello**
CVPR 2022, Embodied AI Workshop

PATENTS

Natural language text conversion and method therefor
*Roger Donaldson, **Cathy Jiao***
U.S. Patent 11275906, 2022

TEACHING/MENTORING

Teaching Assistant

Fall 2024	Large Language Models: Methods and Applications (11-667), CMU
Spring/Fall 2024	Computer Networking (CPSC 317), UBC

Student Research Mentoring

2025- Present	Yijun Pan	University of Michigan Undergraduate
2025- Present	Emily Xiao	CMU Master
2025- Present	Daisy Sheng	CMU Undergraduate (SCS)
2025- Present	Ishita Dasgupta	CMU Master
2025- Present	Niket Jain	CMU Master
2024	Gary Gao	CMU Undergraduate (Mathematical Sciences)

AWARDS

2021 –	CMU Graduate Research Fellowship
2020	Motorola Solutions Merit Award *Awarded for significant contributions in developing novel NLP applications*
2019	Motorola Solutions President's Project Award
2018	Motorola Solutions AI/ML Project Award *Awarded once per year for best project with ML applications*
2018	University of Waterloo, David R. Cheriton School of Computer Science Undergraduate Research Conference Award *Full funding provided to selected Canadian undergraduate students to attend UROC*
2014	University of British Columbia Chancellor Scholar Award *Given to incoming undergraduates with GPA >95% (A+)*

ACADEMIC SERVICE

Conference Reviewer**:** ACL, EMNLP, NAACL (2022 – present), ICTIR 2024
Workshop Reviewer: ICLR DATA-FM 2025, ICML ToM 2023

VOLUNTEERING

2025	*Committee Member*, Student Committee for Faculty Hiring, CMU LTI
2022	*Organizer*, Student Research Symposium, CMU LTI
2022	*Peer Mentor*, CMU LTI Mentoring Program
2022	*Peer Mentor*, CMU A.I. Undergraduate Mentoring Program
2017	*Peer Mentor*, UBC Computer Science Department Tri-Mentoring Program
2016	*Orientation Leader*, UBC Faculty of Science