

Weiqiu You

✉ weiqiuy@seas.upenn.edu

🌐 <http://fallcat.github.io/>

Last updated: Jun. 1, 2025

Research Interests

I build trustworthy machine learning models with faithful and verifiable explanations.

Education

- 2020 – ▶ **Ph.D. Computer and Information Science, University of Pennsylvania**, Philadelphia, PA
Advisor: Prof. Eric Wong
Expected Graduation: May. 2026
- 2018 – 2020 ▶ **M.S. Computer Science, University of Massachusetts Amherst**, Amherst, MA
Advisor: Prof. Mohit Iyyer
- 2014 – 2018 ▶ **B.S. Computer Science and Mathematics, Gordon College**, Wenham, MA
Advisor: Prof. Jonathan Senning, Prof. Russell Bjork
Double major. Honors Thesis title: *Predict Media Interestingness*.

Internship & Employment History

- 2024 ▶ **Research Intern** Okinawa Institute of Science and Technology. Okinawa, Japan.
- 2022 ▶ **Research Intern** IBM Research Yorktown Heights. Yorktown Heights, NY.
- 2020 ▶ **Research Assistant** University of Southern California, Information Sciences Institute. Los Angeles, CA.
- 2018 ▶ **Research Intern** NLP Center, Meituan-Dianping Inc. Beijing, China.

Publications

Conferences and Journals

- 1 Helen Jin*, Shreya Havaldar*, Chaehyeon Kim*, Anton Xue*, **Weiqiu You***, Helen Qu, Marco Gatti, Daniel A Hashimoto, Bhuvnesh Jain, Amin Madani, Masao Sako, Lyle Ungar, and Eric Wong (2025). “The FIX Benchmark: Extracting Features Interpretable to eXperts”. In: *Journal of Data-centric Machine Learning Research (DMLR)*. 🔗 URL: <https://arxiv.org/abs/2409.13684>.
- 2 **Weiqiu You**, Helen Qu, Marco Gatti, Bhuvnesh Jain, and Eric Wong (2025). “Sum-of-Parts: Self-Attributing Neural Networks with End-to-End Learning of Feature Groups”. In: *International Conference on Machine learning (ICML)*. 🔗 URL: <https://openreview.net/forum?id=r6y9TEdLMh>.
- 3 Chaehyeon Kim, **Weiqiu You**, Shreya Havaldar, and Eric Wong (2024). “Evaluating Groups of Features via Consistency, Contiguity, and Stability”. In: *The Second Tiny Papers Track at ICLR 2024*. 🔗 URL: <https://openreview.net/forum?id=IP2etbIEuC>.
- 4 Shreya Havaldar*, **Weiqiu You***, Lyle Ungar, and Eric Wong (2023). “Visual Topics via Visual Vocabularies”. In: *XAI in Action: Past, Present, and Future Applications*. 🔗 URL: <https://openreview.net/forum?id=h60T5pzMgc>.
- 5 Youngja Park and **Weiqiu You** (Dec. 2023). “A Pretrained Language Model for Cyber Threat Intelligence”. In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track*. Singapore: Association for Computational Linguistics, pp. 113–122. 🔗 DOI: 10.18653/v1/2023.emnlp-industry.12.

- 6 Li Zhang, Hainiu Xu, Yue Yang, Shuyan Zhou, **Weiqiu You**, Manni Arora, and Chris Callison-Burch (May 2023). “Causal Reasoning of Entities and Events in Procedural Texts”. In: *Findings of the Association for Computational Linguistics: EACL 2023*. Dubrovnik, Croatia: Association for Computational Linguistics, pp. 415–431. [DOI: 10.18653/v1/2023.findings-eacl.31](https://doi.org/10.18653/v1/2023.findings-eacl.31).
- 7 Artemis Panagopoulou, Manni Arora, Li Zhang, Dimitri Cugini, **Weiqiu You**, Yue Yang, Liyang Zhou, Yuxuan Wang, Zhaoyi Hou, Alyssa Hwang, Lara Martin, Sherry Shi, Chris Callison-Burch, and Mark Yatskar (2022). “QuakerBot: A household dialog system powered by large language models”. In: *Alexa Prize TaskBot Challenge 1 Proceedings*. [URL: https://www.amazon.science/alexaprize/proceedings/quakerbot-a-household-dialog-system-powered-by-large-language-models](https://www.amazon.science/alexaprize/proceedings/quakerbot-a-household-dialog-system-powered-by-large-language-models).
- 8 Thamme Gowda, **Weiqiu You**, Constantine Lignos, and Jonathan May (June 2021). “Macro-Average: Rare Types Are Important Too”. In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Online: Association for Computational Linguistics, pp. 1138–1157. [DOI: 10.18653/v1/2021.naacl-main.90](https://doi.org/10.18653/v1/2021.naacl-main.90).
- 9 **Weiqiu You***, Simeng Sun*, and Mohit Iyyer (July 2020). “Hard-Coded Gaussian Attention for Neural Machine Translation”. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Online: Association for Computational Linguistics, pp. 7689–7700. [DOI: 10.18653/v1/2020.acl-main.687](https://doi.org/10.18653/v1/2020.acl-main.687).

Preprints

- 1 Helen Jin*, Anton Xue*, **Weiqiu You**, Surbhi Goel, and Eric Wong (2025). *Probabilistic Stability Guarantees for Feature Attributions*. arXiv: 2504.13787 [cs.LG]. [URL: https://arxiv.org/abs/2504.13787](https://arxiv.org/abs/2504.13787).
- 2 Delip Rao*, **Weiqiu You***, Eric Wong, and Chris Callison-Burch (2025). *NSF-SciFy: Mining the NSF Awards Database for Scientific Claims*. arXiv: 2503.08600 [cs.CL]. [URL: https://arxiv.org/abs/2503.08600](https://arxiv.org/abs/2503.08600).
- 3 Siqi Zeng, Yifei He, **Weiqiu You**, Yifan Hao, Yao-Hung Hubert Tsai, Makoto Yamada, and Han Zhao (2025). *Efficient Model Editing with Task Vector Bases: A Theoretical Framework and Scalable Approach*. arXiv: 2502.01015 [cs.LG]. [URL: https://arxiv.org/abs/2502.01015](https://arxiv.org/abs/2502.01015).
- 4 **Weiqiu You** and Youngja Park (2024). *Cyber-Attack Technique Classification Using Two-Stage Trained Large Language Models*. arXiv: 2411.18755 [cs.LG]. [URL: https://arxiv.org/abs/2411.18755](https://arxiv.org/abs/2411.18755).

Teaching Experience

2021	<ul style="list-style-type: none"> ► Computational Linguistics UPenn CIS530, Teaching Assistant, Spring 2021, Fall 2021
Spring 2020	<ul style="list-style-type: none"> ► Advanced Natural Language Processing UMass COMPSCI685, Grader
Spring 2018	<ul style="list-style-type: none"> ► Data Structures and Algorithms Gordon CPS222, Teaching Assistant
Spring 2017	<ul style="list-style-type: none"> ► Calculus II Gordon MAT122, Teaching Assistant
Fall 2016	<ul style="list-style-type: none"> ► Differential Equations Gordon MAT225, Teaching Assistant
2016 – 2018	<ul style="list-style-type: none"> ► Biostatistics Gordon, SPSS Help Session Tutor

*Equal contribution.

Teaching Experience (continued)

- ▶ **Calculus**
Gordon, Tutor

Invitations

- 2024 ▶ **Panalist**
Women in CS Panel, Computers and Society class. Gordon College, MA.
- ▶ **Speaker**
Artificial Intelligence Week Alumni Forum. High School Affiliated to Renmin University of China, Beijing, China.
- 2022 ▶ **Panalist**
Women in CS Panel, Computers and Society class. Gordon College, MA.

Awards

- 2024 ▶ **AWS-AI ASSET Fellow.**

Academic Services

- 2025 ▶ **ICML.**
Reviewer.
- 2024 ▶ **ICLR.**
Reviewer.
- 2022 – 2023 ▶ **ACL Rolling Review.**
Reviewer.
- 2023 ▶ **ACL.**
Reviewer.
- 2022 ▶ **CLunch, a weekly NLP research seminar run by PennNLP.**
Organizer
- 2021 – 2023 ▶ **EMNLP.**
Reviewer.