XIAOCHUANG HAN (Han)

Email: xhan77@cs.washington.edu ◇ Homepage: xhan77.github.io ◇ Google Scholar ◇ Twitter

EDUCATION

University of Washington September 2021 - Present (Exp. 2025)
Ph.D. in Computer Science and Engineering
Advisor: Yulia Tsvetkov

Carnegie Mellon University August 2019 - August 2021
M.S. in Language Technologies
Advisor: Yulia Tsvetkov

Georgia Institute of Technology August 2015 - May 2019
B.S. in Computer Science, Minor in Mathematics
Advisor: Jacob Eisenstein

SELECTED PUBLICATIONS

In-Context Alignment: Chat with Vanilla Language Models Before Fine-Tuning
Xiaochuang Han
arXiv preprint

SSD-2: Scaling and Inference-time Fusion of Diffusion Language Models
Xiaochuang Han, Sachin Kumar, Yulia Tsvetkov, Marjan Ghazvininejad
Under review

Trusting Your Evidence: Hallucinate Less with Context-aware Decoding
Weijia Shi*, **Xiaochuang Han***, Mike Lewis, Yulia Tsvetkov, Luke Zettlemoyer, Scott Wen-tau Yih
Under review

SSD-LM: Semi-autoregressive Simplex-based Diffusion Language Model for Text Generation and Modular Control
Xiaochuang Han, Sachin Kumar, Yulia Tsvetkov
ACL 2023

Understanding In-Context Learning via Supportive Pretraining Data
Xiaochuang Han, Daniel Simig, Todor Mihaylov, Yulia Tsvetkov, Asli Celikyilmaz, Tianlu Wang
ACL 2023

ORCA: Interpreting Prompted Language Models via Locating Supporting Evidence in the Ocean of Pretraining Data
Xiaochuang Han and Yulia Tsvetkov
Under review

Toward Human Readable Prompt Tuning: Kubrick's The Shining is a good movie, and a good prompt too?
Weijia Shi*, **Xiaochuang Han***, Hila Gonen, Ari Holtzman, Yulia Tsvetkov, Luke Zettlemoyer
Under review

Influence Tuning: Demoting Spurious Correlations via Instance Attribution and Instance-Driven Updates
Xiaochuang Han and Yulia Tsvetkov
Findings of EMNLP 2021

Fortifying Toxic Speech Detectors Against Veiled Toxicity
Xiaochuang Han and Yulia Tsvetkov
EMNLP 2020

Explaining Black Box Predictions and Unveiling Data Artifacts through Influence Functions
Xiaochuang Han, Byron C. Wallace, Yulia Tsvetkov
ACL 2020

Unsupervised Domain Adaptation of Contextualized Embeddings for Sequence Labeling
Xiaochuang Han and Jacob Eisenstein
EMNLP 2019

No Permanent Friends or Enemies: Tracking Dynamic Relationships between Nations from News
Xiaochuang Han, Eunsol Choi, Chenhao Tan
NAACL 2019

Mind Your POV: Convergence of Articles and Editors Towards Wikipedia's Neutrality Norm
Umashanthi Pavalanathan, **Xiaochuang Han**, Jacob Eisenstein
CSCW 2018

RESEARCH EXPERIENCE

Meta AI, FAIR Labs October 2022 - Present
Visiting Researcher, with Marjan Ghazvininejad and Omer Levy

- Explore the scaling and inference-time collaboration of diffusion-based language models.
- Work on controllable text generation to enhance generation diversity and tool-using abilities.

UW NLP / CMU LTI TsvetShop August 2019 - Present
Graduate Research Assistant, with Yulia Tsvetkov

- Develop a performant diffusion language model based on vocabulary simplexes for modular control.
- Interpret prompted language models by finding evidence in the pretraining data.
- Demote spurious correlations in models by instance attribution and instance-driven updates.
- Fortify toxic language classifiers against veiled toxicity using interpretable ML methods.
- Explore the interpretability of NLP models through the lens of training examples.

Meta AI, FAIR Labs June 2022 - September 2022
Research Intern, with Tianlu Wang

- Interpreted mechanisms of in-context learning by extracting data evidence from the pretraining data.

Georgia Tech Computational Linguistics Lab August 2017 - May 2019
Undergraduate Research Assistant, with Jacob Eisenstein

- Improved unsupervised domain adaptation of contextualized embeddings for sequence labeling.
- Explored variational methods for geo-entity resolution.

- Analyzed the effect of Wikipedia's neutrality norm.
- Worked on stance classifiers in a quantitative model of stancetaking in online forums.

University of Colorado Boulder NLP and CSS Lab May 2018 - August 2018
Research Intern, with Chenhao Tan

- Built an unsupervised model to explore entity-to-entity relations in world news.

TEACHING ASSISTANTSHIPS

UW CSE 447 / M 547: Natural Language Processing Spring 2022
Head Teaching Assistant, with Yulia Tsvetkov

- Adapted and redesigned homework assignments, gave tutorials on structured prediction methods, designed quiz questions and hosted weekly office hours.

CMU 11-711: Algorithms for NLP Fall 2020
Graduate Teaching Assistant, with Emma Strubell, Yulia Tsvetkov, and Robert Frederking

- Adapted and redesigned homework assignments, gave a lecture on natural language inference and interpretability in neural NLP, led recitations and hosted office hours.

ACADEMIC SERVICE

Reviewer (*outstanding reviewer)

- NeurIPS 2023, ACL 2023, EMNLP 2022, NeurIPS 2022, ICLR 2022, DistShift 2022, ACL 2021*, ARR 2021, NAACL 2021, EACL 2021*, CSUR 2021, EMNLP 2020, W-NUT 2020

STUDENT ORGANIZATIONS

Georgia Tech Big Data Club August 2015 - May 2019
President and Lecturer

- Organized weekly meetings and gave lectures on machine learning and database tools and algorithms.