XIAOCHUANG HAN (Han)

xiaochuang.han@gmail.com ◇ https://xhan77.github.io/

EDUCATION

University of Washington September 2021 - Present
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

PUBLICATIONS

ORCA: Interpreting Prompted Language Models via Locating Supporting Evidence in the Ocean of Pretraining Data
Xiaochuang Han and Yulia Tsvetkov.
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, and 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, and Chenhao Tan.
NAACL 2019

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

Interactional Stancetaking in Online Forums
Scott Kiesling, Umashanthi Pavalanathan, Jim Fitzpatrick, **Xiaochuang Han**, and Jacob Eisenstein.

Computational Linguistics, September 2018

RESEARCH EXPERIENCE

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

· Interpret mechanisms of in-context learning by extracting or generating data evidence from or based on the pretraining data, using instance attribution and controlled text generation techniques.

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

· 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.

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)

· EMNLP 2022, NeurIPS 2022, ICLR 2022, ARR 2021, CSUR 2021, ACL 2021*, NAACL 2021, EACL 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.