Kawin Ethayarajh

✉ kawin@stanford.edu ethayarajh kawine.github.io

Education

Stanford University
Ph.D., Computer Science — 2019 – 2023 (expected)
Advisor: Dan Jurafsky

University of Toronto
M.Sc., Computer Science — 2017 – 2019
Advisor: Graeme Hirst

University of Toronto, Victoria College
B.Sc. Hons., Computer Science — 2013 – 2017

Awards

ICML Outstanding Paper — 2022

Facebook PhD Fellowship: $84,000 USD — 2021
Awarded to only two people in the area of Natural Language Processing (NLP).

NSERC Postgraduate Scholarship - Doctoral: $63,000 CAD — 2019

NSERC Canada Graduate Scholarship - Doctoral: $105,000 CAD (declined) — 2019

Best Paper – Repl4NLP, ACL 2018 — 2018

Rhodes Scholarship Finalist — 2017

University of Toronto Fellowship: $11,200 CAD — 2017

John H. Moss Scholarship: $16,650 CAD — 2017
Given to the top graduating student, for academics and leadership.

Chancellor Northrop Frye Gold Medal — 2017
For the graduating student with the highest academic standing at Victoria College.

NSERC Undergraduate Student Research Award: $4,500 CAD — 2015
Awarded by NSERC (Canadian NSF) to undergraduate researchers.

Bank of Montreal National Scholarship: $75,000 CAD — 2013
Merit–based university scholarship granted to 8 Canadians.

Former Positions

Allen Institute for Artificial Intelligence, Research Scientist Intern — Summer 2021
Hosts: Yejin Choi and Swabha Swayamdipta (Mosaic Team)
Project: Understanding dataset difficulty with information theory.

Google, SWE Intern — Summer 2019
Hosts: AdsAI Team
Project: Embed hypergraphs at scale using autoencoders.

Google, SWE Intern — Summer 2018
Hosts: Research & Machine Intelligence Team
Project: Zero-shot relation extraction using pre-trained QA models.

University of Toronto, Graduate Research Assistant — 2017 – 2019
Hosts: Graeme Hirst and David Duvenaud
Project: Theoretical analysis of word embeddings and sentence embeddings in NLP.

University of Toronto, Undergraduate Research Assistant — 2015 – 2016
Hosts: Michalis Famelis and Marsha Chechik
Project: Transferability across domain-specific languages.

14. **The Authenticity Gap in Human Evaluation.**
Kawin Ethayarajh and Dan Jurafsky.
Empirical Methods in Natural Language Processing (EMNLP), 2022.

13. **Understanding Dataset Difficulty with $\mathcal{V}$-Usable Information.**
Kawin Ethayarajh, Yejin Choi, and Swabha Swayamdipta.
International Conference on Machine Learning (ICML), 2022. 🏆 **Outstanding Paper.**

12. **Richer Countries and Richer Representations.**
Kaitlyn Zhou, Kawin Ethayarajh, and Dan Jurafsky.
Findings of ACL, 2022.

11. **Problems with Cosine as a Measure of Embedding Similarity for High Frequency Words.**
Kaitlyn Zhou, Kawin Ethayarajh, Dallas Card, and Dan Jurafsky.
Association for Computational Linguistics (ACL), 2022.

10. **Dynaboard: An Evaluation-As-A-Service Platform for Holistic Next-Generation Benchmarking.**
Zhiyi Ma*, Kawin Ethayarajh*, Tristan Thrush*, Somya Jain, Ledell Wu, Robin Jia, Christopher Potts, Adina Williams, and Douwe Kiela.
Neural Information Processing Systems (NeurIPS), 2021.

9. **Conditional Probing: Measuring Usable Information Beyond a Baseline.**
John Hewitt, Kawin Ethayarajh, Percy Liang, and Chris Manning.
Empirical Methods in Natural Language Processing (EMNLP), 2021.

8. **On the Opportunities and Risks of Foundation Models.**
Rishi Bommasani, ..., Kawin Ethayarajh, ..., Percy Liang.
whitepaper.

7. **Attention Flows are Shapley Value Explanations.**
Kawin Ethayarajh and Dan Jurafsky.
Association for Computational Linguistics (ACL), 2021.

6. **Utility is in the Eye of the User: A Critique of NLP Leaderboards.**
Kawin Ethayarajh and Dan Jurafsky.
Empirical Methods in Natural Language Processing (EMNLP), 2020.

5. **Is Your Classifier Actually Biased? Measuring Fairness under Uncertainty with Bernstein Bounds.**
Kawin Ethayarajh.
Association for Computational Linguistics (ACL), 2020.

4. **How Contextual are Contextualized Word Representations? Comparing the Geometry of BERT, ELMo, and GPT-2 Embeddings.**
Kawin Ethayarajh.
Empirical Methods in Natural Language Processing (EMNLP), 2019.

3. **Understanding Undesirable Word Embedding Associations.**
Kawin Ethayarajh, David Duvenaud, and Graeme Hirst.
Association for Computational Linguistics (ACL), 2019.

2. **Towards Understanding Linear Word Analogies.**
Kawin Ethayarajh, David Duvenaud, and Graeme Hirst.
Association for Computational Linguistics (ACL), 2019.

1. **Unsupervised Random Walk Sentence Embeddings: A Strong but Simple Baseline.**
Kawin Ethayarajh.
Repl4NLP Workshop at ACL 2018. 🏆 **Best Paper.**

Invited Talks

Understanding Dataset Difficulty

- RIKEN Center for Advanced Intelligence Project (Japan) — 09/2022
- Stanford NLP — 08/2022

Utilitarian Model Evaluation

- IBM Research (Zurich) — 05/2022
- Stanford HCI — 11/2020

Teaching

Stanford CS221: Artificial Intelligence, Teaching Assistant — Fall 2022

Foundational principles and practice implementing various AI systems. Specific topics include machine learning, search, Markov decision processes, game playing, constraint satisfaction, graphical models, and logic.

Service

Review of Undergraduate Computer Science, Founding Editor-in-Chief — 2015 – 2016

Started non-archival publication dedicated to CS undergrad research, helping students write and edit their first scientific articles and publishing research from from UToronto, Cornell, and MIT.

Governing Council of the University of Toronto, Board Member — 2015 – 2016

Appointed to a board of the university's highest governing body to shape student affairs, where I debated and voted on several key issues, including student privacy and data collection.

Professional Service

Reviewing

- EMNLP — 2020, 2021, 2022
- ACL — 2020, 2021, 2022
- ACL Rolling Review — 2021, 2022