Kawin Ethayarajh

kawin@stanford.edu   kawine.github.io

EDUCATION	**Stanford University**	
	Ph.D., Computer Science (advisor: Dan Jurafsky)	Fall 2019 -
	University of Toronto	
	M.Sc., Computer Science (advisor: Graeme Hirst)	2019
	B.Sc. Hons., Computer Science	2017
AWARDS	NSERC Postgraduate Scholarship - Doctoral: $63,000 CAD	2019
	Canadian equivalent to NSF Fellowship.	
	NSERC Canada Graduate Scholarship - Doctoral: $105,000 CAD (declined)	2019
	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.	
	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 scholarship granted to 8 Canadians.	
	Governor General's Academic Medal (Bronze)	2013

RESEARCH & ENGINEERING	**Google**	

INTERN, ADSAI   Summer 2019
- created a novel method for embedding heterogeneous hypergraphs using autoencoders
- achieved 10% improvement on F_1 score over previous state-of-the-art for multi-label node classification
- developed a new method of training hypergraph embeddings at scale using sharding

INTERN, RESEARCH & MACHINE INTELLIGENCE   Summer 2018
- conceived and built a pipeline for zero-shot relation extraction using pre-trained QA models
- increased precision by 12% and $F_{0.5}$ score by 0.023 over baseline

University of Toronto

RESEARCH ASST, NLP GROUP   2017 – 2019
- derived an unsupervised sentence embedding approach (Best Paper, Repl4NLP at ACL 2018)
- published proof of why analogies (e.g., *king is to queen as man is to woman*) exist in word vector spaces
- published theoretical analysis of social biases in word embedding spaces

RESEARCH ASST, SIGNAL PROCESSING & ORAL COMMUNICATION LAB   2016 – 2017
- used psycholinguistics to study seasonal changes in mood across 100K Reddit users (published)
- found that a small cohort was acutely sensitive to seasonal changes, supporting mainstream hypothesis

RESEARCH ASST, FACULTY OF LAW   Summer 2016
- made the first citation prediction model for a common law system, using 52K legal decisions (published)
- used network theory (HITS) and machine learning (SVMs) to predict citations with 93.8% accuracy

VOLUNTEERING   **Review of Undergraduate Computer Science (RUCS)**

FOUNDER & EDITOR-IN-CHIEF   2015 – 2016
- started first publication dedicated to CS undergrad research; built readership of several thousand
- RUCS has been active for 5+ years and has published work from UToronto, Cornell, and MIT

Governing Council of the University of Toronto

UNIVERSITY AFFAIRS BOARD MEMBER   2015 – 2016
- appointed to a board of the university's highest governing body to shape student affairs
- debated and voted on several key issues, including student privacy and data collection

The Artisan Toolkit (not-for-profit)

REPORTING OFFICER   2013 – 2015
- managed $600K in funds to teach business practices to traditional Afghan artisans
- helped create multimedia content in English, Dari & Pashto for literate and illiterate users

PUBLICATIONS

1. Ethayarajh, K. (2020). Is your classifier actually biased? measuring fairness under uncertainty with bernstein bounds. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2914–2919, Online. Association for Computational Linguistics.

2. Ethayarajh, K. (2019b). Rotate king to get queen: Word relationships as orthogonal transformations in embedding space. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3503–3508, Hong Kong, China. Association for Computational Linguistics.

3. Ethayarajh, K. (2019a). How contextual are contextualized word representations? comparing the geometry of BERT, ELMo, and GPT-2 embeddings. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 55–65, Hong Kong, China. Association for Computational Linguistics.

4. Ethayarajh, K., Duvenaud, D., and Hirst, G. (2019b). Understanding undesirable word embedding associations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1696–1705, Florence, Italy. Association for Computational Linguistics.

5. Ethayarajh, K., Duvenaud, D., and Hirst, G. (2019a). Towards understanding linear word analogies. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3253–3262, Florence, Italy. Association for Computational Linguistics.

6. Ethayarajh, K. (2018). Unsupervised random walk sentence embeddings: A strong but simple baseline. In *Proceedings of The Third Workshop on Representation Learning for NLP*, pages 91–100, Melbourne, Australia. Association for Computational Linguistics. **Best Paper.**