

CHRIS KEDZIE

CURRICULUM VITAE

phone +1 (925) 323 1837

email kedzie@cs.columbia.edu

web www.cs.columbia.edu/~kedzie

 [LinkedIn](#) ·  [Github](#) ·  [Google Scholar](#)

Research Statement

I am interested in computational models of natural language understanding and generation. I am actively investigating methods for improving neural language models via interaction with secondary models of semantics and/or syntactic structure. Human language use exhibits tremendous productivity (in the sense of Chomsky, i.e. “making infinite use of finite means”), while neural models are frequently poor generalizers of language data, often learning non-hierarchical or otherwise idiosyncratic representations. In the hopes of improving this situation, I am also investigating topics in differentiable data structures (e.g., stacks or addressable memory) and model regularization using adversarial and cooperative learning tasks to encourage more systematic and explainable model behavior.

I also apply machine learning to natural language data to make useful predictions, such as predicting the importance of text for summarization or extracting signals from the web for social scientists.

Education

Columbia University

2014-pres. Ph.D., Computer Science

Advisor: Kathleen McKeown

Thesis: *Salience Estimation and Faithful Generation: Modeling Methods for Text Summarization and Generation*

2014-2018 M.Phil., Computer Science

2012-2014 M.S., Computer Science

Loyola Marymount University

2004-2008 B.A., Music & Recording Arts Double Major

Research Experience

Research Intern, **Facebook Research**, Summer 2017.

I interned with Umut Ozertem in the Applied Machine Learning (AML) group, focusing on click bait and hate speech detection. During my project, I developed an adversarial training method to improve performance of our convolutional neural network text classifiers. I implemented this architecture in PyTorch and Caffe2, and integrated it into the internal FB Learner system for tracking and running experiments. My project is currently in use by several teams in the AML group, and is currently under submission for a patent.

Research Intern, **Microsoft Research**, Summer 2015.

I interned with Fernando Diaz in New York City, continuing our collaboration on streaming news summarization. I developed scalable summarization systems to provide users with brief updates of news events as they were unfolding. We participated in the Temporal Summarization Track of the 2015 Text Retrieval Conference, where we were the top performer, and invited to give a talk.

Selected Publications **Chris Kedzie** and Kathleen McKeown. [A Good Sample is Hard to Find: Noise Injection Sampling and Self-Training for Neural Language Generation Models](#). *Under submission*.

Controllable Language Generation Katy Gero, **Chris Kedzie**, Jonathan Reeve, and Lydia Chilton. [Low-Level Linguistic Controls for Style Transfer and Content Preservation](#). *Under submission*.

Automatic Text Summarization **Chris Kedzie**, Kathleen McKeown, and Hal Daumé III. [Content Selection in Deep Learning Models of Summarization](#). *EMNLP 2018*.

Chris Kedzie, Fernando Diaz, and Kathleen McKeown. [Real-Time Web Scale Event Summarization Using Sequential Decision Making](#). *IJCAI 2016*.

Chris Kedzie, Kathleen McKeown, and Fernando Diaz. [Predicting Salient Updates for Disaster Summarization](#). *ACL 2015*.

NLP for Social Science Philipp Blandfort, Desmond U. Patton, William R. Frey, Svebor Karaman, Surabhi Bhargava, Fei-Tzin Lee, Siddharth Varia, **Chris Kedzie**, Michael B. Gaskell, Rossano Schifanella, Kathleen McKeown, Shih-Fu Chang. [Multimodal social media analysis for gang violence prevention](#). *ICWSM 2019*.

Serina Chang, Ruiqi Zhong, Ethan Adams, Fei-Tzin Lee, Siddharth Varia, Desmond U. Patton, William R. Frey, **Chris Kedzie**, Kathleen McKeown. [Detecting gang-involved escalation on social media using context](#). *EMNLP 2018*.

Talks 2018 Agolo.
 2016 Machine Learning and Friends Lunch. UMass. Amherst.
 2015 Text Retrieval Conference (TREC), Temporal Summarization Track.
 2014 Text Retrieval Conference (TREC), Temporal Summarization Track.
 KDD at Bloomberg: Data Frameworks Track.

Demos 2016 *Monitoring Large Scale Disasters*. Data Science Day @ Columbia University Data Science Institute.

Doctoral Consortium 2016 *Extractive and Abstractive Event Summarization over Streaming Web Text*. IJCAI.

Summer Schools	2016	Deep Learning Summer School. University of Montreal.
Community Service	2016–pres. 2014–2017	Organizer. Columbia NLP Reading Group. Organizer. NLP talks at Columbia.
Teaching Experience	Spring 2014	Teaching Assistant. Semantic Technologies in IBM Watson. Instructor: Alfio Gliozzo.