EMILY ALLAWAY

eallaway@cs.columbia.edu
https://emilyallaway.github.io/

EDUCATION

Columbia University, New York, NY — Sept. 2018 – Present
Ph.D. Student, Computer Science – Natural Language Processing
Advisor: Professor Kathleen McKeown
Expected: Spring 2024
M.S. – Computer Science
Received Spring 2021

University of Washington, Seattle, WA — Sept. 2014 – June 2018
Graduated cum laude, GPA: 3.82/4.0
Bachelor of Science (B.S.) in Computer Science
Bachelor of Science (B.S.) in Mathematics
Minor in Ancient Greek

RESEARCH INTERESTS

- **Generalizations and Pragmatics.** Modeling generalizations and their implications to develop NLP methods based in linguistic theory. Understanding the role of generalizations in reasoning, especially in situations with incomplete information, and in argumentation and stancetaking.
- **Reasoning.** Studying how models can reason, how conclusions can be updated or changed, and how automatic reasoning compares to human behavior. Particularly interested in nonmonotonic reasoning.

RESEARCH EXPERIENCE

Graduate Research
Department of Computer Science, Columbia University

- **Generalizations and Pragmatics.** Automatically generating examples of where a generic does and does not hold using linguistic-theory-guided decoding (EACL 2023). Ongoing work on further developing the theoretical framework, improving the methodology, and investigating the connection to nonmonotonic reasoning. Applying generic exceptions to counter stereotypes (Workshop at EMNLP 2022).
- **Stance Detection.** Investigating zero-shot stance detection both with a new dataset and a model that improves performance on a number of challenging linguistic phenomena (EMNLP 2020) and with adversarial learning (NAACL 2021). Using a new lexicon for noun and adjective connotations (EACL 2021) and human rationales (EMNLP 2021) to improve stance detection in low-resource settings. Analysis of methodological paradigms and challenges for zero-shot stance detection (Frontiers of AI 2023) and stance detection for legal language (Workshop at EMNLP 2022). Ongoing work on improving multilingual zero-shot stance detection.
- **Information Extraction.** Developing multilingual relation extraction methods using sentiment and implicit knowledge for a range of languages (TAC 2018, TAC 2019). Using events to denoise unsupervised relation training data to produce strong results with minimized computation cost (COLING 2020). Developed an efficient model for cross-document event coreference (EMNLP 2021). Methods to cluster documents to expert-provided taxonomies (Findings EMNLP 2022).

Undergraduate Research
Paul G. Allen School of Computer Science and Engineering, University of Washington

- **Commonsense Knowledge.** Data on motivations for and reactions to events (ACL 2018), extended the knowledge based to form *ATOMIC*, a resource of almost 900k textual descriptions of inferential knowledge about events (AAAI 2019). Work with Professor Yejin Choi, Hannah Rashkin, and Maarten Sap.
- **Lexical Connotations.** Developed a new multi-task neural architecture to predict verb connotations from dictionary definitions (2017). Work with Professor Yejin Choi and Hannah Rashkin. 2017

PUBLICATIONS

- *Penguins don't Fly: Reasoning about Generics through Instantiations and Exceptions.*
  Emily Allaway, Jena D. Hwang, Chandra Bhagavatula, Kathleen McKeown, Doug Downey, Yejin Choi. In EACL 2023.
- *Zero-Shot Stance Detection: Paradigms and Challenges.*
  Emily Allaway, Kathleen McKeown. In Frontiers of Artificial Intelligence, vol. 5, 2023.
- *Towards Countering Essentialism Through Social Bias Reasoning.*
  Emily Allaway, Nina Taneja, Sarah-Jane Leslie, Maarten Sap. In Workshop on NLP for Positive Impact, EMNLP 2022.
- *Legal and Political Stance Detection of SCOTUS Language.*
  Noah Bergam, **Emily Allaway**, Kathleen McKeown. In NLLP Workshop, EMNLP 2022.
- *Seeded Hierarchical Clustering for Expert-Crafted Taxonomies.*
  Anish Saha, Amith Ananthram, **Emily Allaway**, Heng Ji, Kathleen McKeown. In Findings EMNLP 2022.
- *SafeText: A Benchmark for Exploring Physical Safety in Language Models.*
  Sharon Levy, **Emily Allaway**, Melanie Subbiah, Lydia Chilton, Desmond Patton, Kathleen McKeown, William Yang Wang. In EMNLP 2022.
- *Mitigating Covertly Unsafe Text within Natural Language Systems.*
  Alex Mei, Anisha Kabir, Sharon Levy, Melanie Subbiah, **Emily Allaway**, John Judge, Desmond Patton, Bruce Bimber, Kathleen McKeown, William Yang Wang. In Findings EMNLP 2022.
- *Mapping the Multilingual Margins: Intersectional Biases of Sentiment Analysis Systems in English, Spanish, and Arabic.*
  António Câmara, Nina Taneja, Tamjeed Azad, **Emily Allaway,** Richard Zemel. In Proceedings of Second Workshop on Language Technology for Equality, Diversity, Inclusion, ACL 2022.
- *Does Putting a Linguist in the Loop Improve NLU Data Collection?*
  Alicia Parrish, William Huang, Omar Agha, Soo-Hwan Lee, Nikita Nangia, Alex Warstadt, Karmanya Aggarwal, **Emily Allaway**, Tal Linzen, and Samuel R. Bowman. In Findings of EMNLP 2021.
- *Human Rationales as Attribution Priors for Explainable Stance Detection.*
  Sahil Jayaram and **Emily Allaway**. In EMNLP 2021.
- *Sequential Cross-Document Coreference Resolution.*
  Emily Allaway, Shuai Wang, and Miguel Ballesteros. In EMNLP 2021.
- *A Unified Feature Representation for Lexical Connotations.*
  Emily Allaway and Kathleen McKeown. In EACL 2021.

- *Adversarial Learning for Zero-Shot Stance Detection on Social Media.*
  Emily Allaway*, Malavika Srikanth*, Kathleen McKeown. In NAACL 2021.
  * denotes equal contribution

- *Event-Guided Denoising for Multilingual Relation Learning.*
  Amith Ananthram, **Emily Allaway**, Kathleen McKeown. In COLING 2020.

- *Zero-Shot Stance Detection: A Datset and Model using Generalized Topic Representations.*
  Emily Allaway and Kathleen McKeown. In EMNLP 2020.

- *GAIA at SM-KBP2019 - A Multi-media Multi-lingual Knowledge Extraction and Hypothesis Generation System.*
  Manling Li, Ying Lin, Ananya Subburathinam, Spencer Whitehead, Xiaoman Pan,Di Lu1, Qingyun Wang, Tongtao Zhang, Lifu Huang, Heng Ji, Alireza Zareian, Hassan Akbari, Brian Chen, Bo Wu, **Emily Allaway**, Shih-Fu Chang, Kathleen McKeown, Yixiang Yao, Jennifer Chen, Eric Berquist, Kexuan Sun, Xujun Peng, Ryan Gabbard, Marjorie Freedman, Pedro Szekely, T.K. Satish Kumar, Arka Sadhu, Ram Nevatia, Miguel Rodriguez, Yifan Wang, Yang Bai, Ali Sadeghian, Daisy Zhe Wang. In TAC 2019.

- *ATOMIC: An Atlas of Machine Commonsense for If-Then Reasoning.*
  Maarten Sap, Ronan Le Bras, **Emily Allaway**, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A. Smith, Yejin Choi. In AAAI 2019.

- *GAIA - A Multi-media Multi-lingual Knowledge Extraction and Hypothesis Generation System.*
  Tongtao Zhang, Ananya Subburathinam, Ge Shi, Lifu Huang, Di Lu, Xiaoman Pan, Manling Li, Boliang Zhang, Qingyun Wang, Spencer Whitehead, Heng JiAlireza Zareian, Hassan Akbari, Brian Chen, Ruiqi Zhong, Steven Shao, **Emily Allaway**, Shih-Fu Chang, Kathleen McKeown, Dongyu Li, Xin Huang, Kexuan Sun, Xujun Peng, Ryan Gabbard, Marjorie Freedman, Mayank Kejriwal, Ram Nevatia, Pedro Szekely, T.K. Satish Kumar, Ali Sadeghian, Giacomo Bergami, Sourav Dutta, Miguel Rodriguez, Daisy Zhe Wang. In TAC 2018.

- *Event2Mind: Commonsense Inference on Events, Intents, and Reactions.*
  Hannah Rashkin, Maarten Sap, **Emily Allaway**, Noah A. Smith and Yejin Choi. In ACL 2018.

INVITED TALKS

- Lecture "Unsupervised Constrained Decoding with NeuroLogic". For Language Generation and Summarization Course at Columbia University. September 2022.
- Lecture "Neural Networks for Natural Language Processing". For Natural Language Processing Course at Columbia University. September 2021.
- Research talk "Implicit Meaning and Domain Adaptation in Stance Detection". Columbia University NLP Seminar. March 2021.
- Research talk "Stance Detection". For Columbia University Department of Computer Science Alumni Day. February 2021.
- Research talk "Lexical Connotations & Stance Detection". For demystifying the dissertation talk series, Columbia University Department of Computer Science. August 2020.
- Research talk "Word Connotations & NLP". For Emerging Scholars Program at Columbia University. November 2019.

WORK EXPERIENCE

Allen Institute for Artificial Intelligence, MOSAIC Intern Summer 2021 – Fall 2021
Worked with Jena Hwang and Chandra Bhagavatula on automatically generating instantiations and exceptions to generics (published at EACL 2023).

Amazon Science Research Intern Summer 2020 – Fall 2020
Worked with Miguel Ballesteros as a graduate research intern on cross-document sequential coreference resolution (published at EMNLP 2020).

TEACHING EXPERIENCE

Teaching Assistant, NLP (Advanced Undergrad/Grad) Sept. 2019 – Dec. 2019
Columbia University
Designed and graded homework assignments, held office hours and graded exams.

Teaching Assistant Student Coordinator, CSE 14X (Intro CS) Dec. 2016 – Dec. 2017
Paul G. Allen School of Computer Science and Engineering, University of Washington
Interviewed and selected new TAs for CSE 142/143 each quarter, ran training for new TAs, visited weekly TA class sections and provided feedback.

Teaching Assistant, CSE 14X (Intro CS) Mar. 2015 – Dec. 2016
Paul G. Allen School of Computer Science and Engineering, University of Washington
Developed criteria for grading homework assignments, reviewed other TA grading, taught a weekly 20-25 person class section and graded assignments, held office hours, and graded exams.

STUDENT ADVISING

- Lauren Hong (Aug. 2022 - Dec. 2022)
- Nina Taneja (Jan. 2022 - Aug. 2022) [co-advised with Maarten Sap] → Google
  Paper at NLP4PosImp @ EMNLP 2022: *Towards Countering Essentialism Through Social Bias Reasoning.*
- Noah Bergam (Jan. 2022 - Sept. 2022)
  Paper at NLLP @ EMNLP 2022: *Legal and Political Stance Detection of SCOTUS Language.*
- António Câmara (Sept. 2021 - May 2022) → PhD @ Harvard
- Anish Saha (Sept. 2021 - Sept. 2022) [co-advised with Amith Ananthram]
  Paper at EMNLP Findings 2022: *Seeded Hierarchical Clustering for Expert-Crafted Taxonomies.*
- Lucy Wang (Sept. 2021 - Dec. 2022)
- Sahil Jayaram (Sept. 2020 - May 2021) → Apple
  Paper at EMNLP 2021: *Human Rationales as Attribution Priors for Explainable Stance Detection.*
- William Chiu (Jan. 2021 - May 2021) [co-advised with Amith Ananthram and Elsbeth Turcan]
- Amith Ananthram (Jan. 2020 - Dec. 2020) → PhD @ Columbia
  Paper at COLING 2020: *Event-Guided Denoising for Multilingual Relation Learning.*
- Malavika Srikanth (Jan. 2020 - Dec. 2020) → Amazon
  Paper at NAACL 2021: *Adversarial Learning for Zero-Shot Stance Detection on Social Media.*
- Yuncong Yang (Jan. 2020 - May 2020)

OTHER EXPERIENCE

Advanced Language Processing School, Online (Grenoble, France) Jan. 2021
Week-long series of lectures, seminars, and labs on a range of advanced topics in NLP

AthNLP Summer School, Athens, Greece Sept. 2019
Week-long series of lectures, seminars, and labs on a range of topics in NLP.

SERVICE

Program Committee/Reviewer:
EACL 2023; ACL 2023; EMNLP 2021, 2022; ARR June 2021, Jan., Feb., Mar., 2022; ACL-IJCNLP 2021; DeepLo 2022; LREC 2022; CSRR 2022; W-NUT 2021, 2022;
Co-Organizer/Organizer: Columbia University NLP Seminar Winter 2020 - Present
Organizer: NLP Classics Reading Group at Columbia University Fall 2020

ACADEMIC AWARDS

NSF Graduate Research Fellowship Spring 2019 – Present
National Science Foundation
Denice Dee Denton Scholars Endowment Sept. 2016
Paul G. Allen School of Computer Science and Engineering, University of Washington
Jim Greenfield Scholarship Sept. 2016
Department of Classics, University of Washington
Harvey Densmore Scholarship Sept. 2016
Department of Classics, University of Washington
Funding to attend: Grace Hopper Celebration of Women in Computing Sept. 2016
Paul G. Allen School of Computer Science and Engineering, University of Washington