

Esin Durmus

Postdoctoral Scholar, Stanford University

CONTACT INFORMATION	Department of Computer Science Stanford University Stanford, CA 94305	<i>email:</i> esdurus@stanford.edu <i>website:</i> https://esdurus.github.io <i>google scholar:</i> https://shorturl.at/hnqKT
RESEARCH INTERESTS	Natural Language Generation, Evaluation of Generation System, Faithfulness of Generation Systems, Evaluating Fairness and Bias of NLP Systems, Responsible AI, Social Media Analysis, Understanding Race-related Conversations on Social Media, Computational Argumentation.	
ACADEMIC EXPERIENCE AND EDUCATION	Postdoctoral Scholar at Stanford University Affiliated with Stanford NLP group and Stanford AI lab. Mentors: Dan Jurafsky and Tatsu Hashimoto.	<i>2021 - Present</i>
	Ph.D. in Computer Science, Cornell University Advisor: Claire Cardie.	<i>2015 – 2021</i>
	Master of Science in Computer Science, Cornell University	<i>2015 – 2017</i>
	B.S.E. degree in Computer Engineering, Koç University	<i>2010 – 2015</i>
	B.S.E. degree in Industrial Engineering, Koç University	<i>2010 – 2015</i>
HONORS AND AWARDS	Selected as one of the Rising Stars in EECS 2019 by UIUC. Facebook Fellowship Finalist, 2019. Valedictorian in Department of Computer Engineering, Koç University, 2015. Full Scholarship to attend Koç University, 2010-2015. Koç University Vehbi Koç Scholar Honor Roll, 2010-2015. Full scholarship to attend Stanford Summer International Honors Program, 2014.	
PUBLICATIONS	<ol style="list-style-type: none">Improving Faithfulness by Augmenting Negative Summaries from Fake Documents Tianshu Wang, Faisal Ladhak, Esin Durmus, He He. To appear in EMNLP 2022.Spurious Correlations in Reference-Free Evaluation of Text Generation Esin Durmus^{*1}, Faisal Ladhak*, Tatsunori Hashimoto. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL), 2022.Language modeling via stochastic processes Rose Wang, Esin Durmus, Noah Goodman, Tatsunori Hashimoto. In Proceedings of ICLR, 2022.Faithful or Extractive? On Mitigating the Faithfulness-Abtractiveness Trade-off in Abtractive Summarization Faisal Ladhak*, Esin Durmus*, He He, Claire Cardie, Kathleen McKeown.	

¹Equal Contribution.

- In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (**ACL**), 2022.
5. Proceedings of the Third Workshop on Computational Modeling of People’s Opinions, Personality, and Emotion’s in Social Media
Malvina Nissim, Viviana Patti, Barbara Plank, **Esin Durmus**.
COLING 2020.
 6. Exploring the Role of Argument Structure in Online Debate Persuasion.
Jialu Li, **Esin Durmus** and Claire Cardie.
In Proceedings of the Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2020.
 7. WikiLingua: A New Benchmark Dataset for Cross-Lingual Abstractive Summarization
Faisal Ladhak*, **Esin Durmus***, Claire Cardie and Kathleen McKeown.
In Proceedings of **EMNLP Findings**, 2020.
 8. A Question Answering Evaluation Framework for Faithfulness Assessment in Abstractive Summarization
Esin Durmus, He He and Mona Diab.
In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (**ACL**), 2019.
 9. The Role of Pragmatic and Discourse Context in Determining Argument Impact
Esin Durmus, Faisal Ladhak and Claire Cardie.
In Proceedings of the Conference on Empirical Methods in Natural Language Processing (**EMNLP**), 2019.
 10. Determining Relative Argument Specificity and Stance for Complex Argumentative Structures
Esin Durmus, Faisal Ladhak and Claire Cardie.
In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (**ACL**), 2019.
 11. A Corpus for Modeling User and Language Effects in Argumentation on Online Debating
Esin Durmus and Claire Cardie.
In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (**ACL**), 2019.
 12. Persuasion of the Undecided: Language vs. the Listener
Liane Longpre, **Esin Durmus** and Claire Cardie.
In Proceedings of **Argmining 2019: The 6th Workshop on Argument Mining**.
 13. Modeling the Factors of User Success in Online Debate
Esin Durmus and Claire Cardie.
In Proceedings of the 2019 **World Wide Web Conference** (WWW ’19), May 13–17, 2019, San Francisco, CA, USA.
 14. Exploring the Role of Prior Beliefs for Argument Persuasion
Esin Durmus and Claire Cardie.
In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL**), 2018.
 15. Understanding the Effect of Gender and Stance on Opinion Expression in Debates on “Abortion”
Esin Durmus and Claire Cardie.
In Proceedings of **PEOPLES2018 workshop** on computational modeling of people’s opinions, personality, and emotions in social media.

PREPRINTS

1. Holistic Evaluation of Language Models
Preprint, 2022.
2. Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale
Federico Bianchi*, Pratyusha Kalluri*, **Esin Durmus***, Faisal Ladhak*, Myra Cheng*, Debora Nozza, Tatsunori Hashimoto, Dan Jurafsky, James Zou, Aylin Caliskan
Preprint, 2022.
3. On the opportunities and risks of foundation models
Preprint, 2021.

MANUSCRIPTS

1. An Intersectional Analysis of Stereotype Amplification in Stories Generated by Large Language Models.
Myra Cheng, Tatsu Hashimoto, Dan Jurafsky, **Esin Durmus**.
Under Preparation.
2. When Do Pre-Training Biases Propagate to Downstream Tasks? A Case Study in Text Summarization
Faisal Ladhak*, **Esin Durmus***, Mirac Suzgun*, Tianyi Zhang, Dan Jurafsky, Kathleen McKeown, Tatsunori Hashimoto.
Under Submission.
3. Tracing and Removing Annotation Errors in Natural Language Generation Datasets
Faisal Ladhak, **Esin Durmus**, Tatsunori Hashimoto
Under Submission.
4. Evaluating the Interactability of Language Models
Mina Lee, Megha Srivastava, Amelia Hardy, **Esin Durmus**, Ashwin Paranjape, John Thickstun, Ines Gerard-Ursin, Faisal Ladhak, Frieda Rong, Rose E. Wang, Xiang Lisa Li, Minae Kwon, Joon Sung Park, Hancheng Cao, Tony Lee, Rishi Bommasani, Michael Bernstein, Percy Liang.
Under Preparation.
5. User-Centric Evaluation of Dialogue Capabilities of Large Language Models
Amelia Hardy, Ashwin Paranjape, **Esin Durmus**, Ines Gerard-Ursin, Faisal Ladhak, Percy Liang
Under Preparation.

TEACHING
EXPERIENCE

Co-Instructor for Introduction to Natural Language Processing, Cornell University. *Fall 2020*

Teaching Assistant for Introduction to Natural Language Processing, Cornell University. *Fall 2016, Fall 2017, Fall 2018*

Teaching Assistant for Machine Learning for Data Science, Cornell University. *Spring 2016*

Teaching Assistant for Introduction to Web Design, Cornell University. *Fall 2015*

PUBLISHED
DATASETS

WikiLingua Dataset. The dataset includes 770k article and summary pairs in 18 languages from WikiHow. We extracted gold-standard article-summary alignments across languages by aligning the images that are used to describe each how-to step in an article.

Kialo Dataset. We extracted argument trees for 741 controversial topics from www.kialo.com. The dataset includes diverse set of controversial topics. Each controversial topic is represented by a thesis and tagged to be related to pre-defined generic categories such as *Politics*, *Ethics*, *Society* and *Technology*. The dataset includes 95,312 unique claims.

DDO Dataset. We collected 77,655 debates and extensive information about 45,348 user participating in or voting for these debates. The dataset includes debate rounds, comments and votes for the debates as well as debater and voter profile information such as their religion, ideology, and opinions about controversial topics.

ACTIVITIES

Research Mentor

Mentored several students to help them conduct NLP research and publish research papers.

Reviewer

EMNLP 2019, UIST 2019, Widening NLP 2019, ACL 2020, SRW 2021, ACL 2021.

Publicity and Publication Chair

PEOPLES2020 workshop co-located with COLING 2020.

Organizing Committee

GEM workshop at ACL-IJCNLP 2021.

Student Research Workshop at NAACL 2021.

New Student Mentor, Cornell University.

Mentoring students to provide academic and social assistance.

INDUSTRY EXPERIENCE

Research Intern at Google LLC.

Summer 2020

Worked at the Google AI language team on improving text generation systems with human feedback.

Applied Scientist Intern at Amazon.com Services, Inc.

Summer 2019

Worked at the AWS team on evaluating faithfulness of abstractive summarization systems.

Applied Scientist Intern at Amazon Corporate LLC.

Summer 2017

Worked at the Alexa Natural Language Understanding team. Built n-gram, DNN and LSTM based systems for dialog state tracking. Applied personalization methods to adapt the model using user-specific information.