

ISABEL CACHOLA

isabelcachola.com \diamond icachola@cs.jhu.edu

EDUCATION

Johns Hopkins University

August 2020 – Present

PhD Student in Computer Science at the Center for Speech and Language Processing (CLSP).

Advised by Prof. Mark Dredze.

The University of Texas at Austin

August 2015 – May 2019

Bachelor of Science in Mathematics. Dean's Honored Graduate with Distinction in Research.

Advised by Prof. Junyi Jessy Li and Prof. Greg Durrett.

Relevant Classes: Natural Language Processing, Machine Learning, Self-Supervised Statistical Models, Machine Translation, Intro to Human-Computer Interaction, Information Retrieval and Web Agents.

RESEARCH PROJECTS & PUBLICATIONS

**Denotes Equal Contribution*

Interests: Text Generation, Summarization, Question Answering, Language Modelling, Natural Language Understanding, Computational Social Science.

Text Generation in the Scientific Domain, *The Allen Institute for AI & Johns Hopkins*

2019 – Present

Collected a dataset for extreme scientific summarization using a novel annotation protocol. Designed a training strategy that exploits an auxiliary training signal for improved performance on a BART-based model.

- **TLDR: Extreme Summarization of Scientific Documents.** Isabel Cachola, Kyle Lo, Arman Cohan, and Daniel S. Weld. *Proceedings of 2020 Conference on Empirical Methods in Natural Language Processing Findings (EMNLP)*. November 2020.
- **Explaining Relationships Between Scientific Documents.** Kelvin Luu*, Xinyi Wu*, Rik Koncel-Kedziorski, Kyle Lo, Isabel Cachola, Noah A. Smith. *The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP)* August 2021.

Interpretability of Black Box Models, *Johns Hopkins*

2020 – 2022

Applied model distillation techniques to train an inherently interpretable student model, capable of producing both global and local explanations for large, black box models. Showed that this method is faithful to the teacher model.

- **Model Distillation for Faithful Explanations of Medical Code Predictions.** Zach Wood-Doughty*, Isabel Cachola*, Mark Dredze. *Proceedings of 2020 Conference on Empirical Methods in Natural Language Processing Findings (EMNLP)*. May 2022.
- **Proxy Model Explanations for Time Series RNNs.** Zach Wood-Doughty, Isabel Cachola, Mark Dredze. *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA). Processing Findings (EMNLP)*. December 2021.

Accessibility of Scientific Documents, *The Allen Institute for AI*

2020 – 2021

Built and implemented a parser that converts low accessibility scientific papers into higher accessibility HTML renders and evaluated the system with blind or low vision researchers. Conducted a large-scale, empirical study of the state of accessibility in scientific papers and suggested recommendations for improvement.

- **SciA11y: Converting Scientific Papers to Accessible HTML (Best Artifact Award).** Lucy Lu Wang*, Isabel Cachola*, Jonathan Bragg, Evie Yu-Yen Cheng, Chelsea Haupt, Matt Latzke, Bailey Kuehl, Madeleine N van Zuylen, Linda Wagner, and Daniel S. Weld. *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*. October 2021.
- **Improving the accessibility of scientific documents.** Lucy Lu Wang*, Isabel Cachola*, Jonathan Bragg, Evie Yu-Yen Cheng, Chelsea Haupt, Matt Latzke, Bailey Kuehl, Madeleine Van Zuylen, Linda Wagner, and Daniel S. Weld. *ArXiv:2105.00076*. April 2021.

Vulgarity on Social Media, *The University of Texas at Austin* 2017 – 2018
Conducted a large-scale, empirical analysis of vulgar word usage on social media and leveraged vulgar features to improve sentiment detection. Trained a model to predict the pragmatic function of vulgar word usage and then proved that this task can be useful in downstream tasks such as hate speech detection.

- **Expressively vulgar: The socio-dynamics of vulgarity and its effects on sentiment analysis in social media (Area Chair Favorite).** Isabel Cachola*, Eric Holgate*, Daniel Preotiuc-Pietro and Junyi Jessie Li. *Proceedings of the 27th International Conference on Computational Linguistics (COLING)*. August 2018.
- **Why Swear? Analyzing and Inferring the Intentions of Vulgar Expressions** Eric Holgate, Isabel Cachola, Daniel Preotiuc-Pietro and Junyi Jessie Li. *Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. November 2018.

PROFESSIONAL & TEACHING EXPERIENCE

Computers as Authors Course Instructor, <i>Johns Hopkins University</i>	Fall 2022
Intro to Algorithms Teaching Assistant, <i>Johns Hopkins University</i>	Fall 2021
Pre-Doctoral Young Investigator , <i>The Allen Institute for AI</i>	2019 – 2020
Data Science Intern, <i>Atipica</i>	Spring 2019
Data Science Intern, <i>CDK Global</i>	Summer 2018

PRESENTATIONS

Guest Lecturer, <i>AI Ethics for Healthcare Applications at Johns Hopkins University</i>	October 2022
Guest Lecturer, <i>Intro to Machine Learning at Johns Hopkins University</i>	September 2022
Poster Session, <i>BioNLP at ACL 2022</i>	May 2022
New Voices in AI Interview, <i>AIHub</i>	February 2022
Invited Speaker, <i>KungFu.AI Engineering</i>	February 2022
Keynote, <i>NLP Summit</i>	October 2021
Interview, <i>Hanselminutes Podcast</i>	November 2020
Interview, <i>DAIR.AI</i>	June 2020
Oral Presentation, <i>COLING 2018</i>	August 2018

AWARDS

NSF Graduate Research Fellowship Program Honorable Mention	Fall 2020
National Center for Women & Information Technology Collegiate Award Honorable Mention	Fall 2020
Dean's Honored Graduate, Highest graduating honor in The College of Natural Science	Spring 2019
College of Natural Sciences Aspire Award for Research Excellence	Spring 2019
Gold Medalist in Texas Collegiate Powerlifting	Fall 2018
UT Austin Undergraduate Research Fellowship	Spring 2018

SERVICE

Women Mentoring Whiting, Mentor, <i>Johns Hopkins University</i>	Fall 2021 – Present
Graduate Student Mentorship Initiative, Mentor, <i>Científico Latino</i>	Fall 2020 – Spring 2021
Recruitment Visit Committee, Member, <i>Center for Speech and Language Processing</i>	Spring 2021
Committee on Diversity, Equity, and Inclusion, Member, <i>Allen Institute for AI</i>	Fall 2019 – Summer 2020
Austin Community Financial Center, Spanish Translator, <i>Foundation Communities</i>	Spring 2019
UTeach Science Outreach Teacher, <i>UT Austin</i>	Spring 2017 & Fall 2015
Volunteer Deputy Registrar, <i>City of Austin</i>	Fall 2015 – Fall 2016

SKILLS

Technical: Python, Linux, Git, AWS, ML Packages (PyTorch, HuggingFace Transformers, scikit-learn, etc.).
Natural Language: Native English Speaker, Proficient in Spanish.