Lucy Li

lucy3_li@berkeley.edu | lucy3.github.io

Research interests: natural language processing, computational sociolinguistics, computational social science.

Education

University of California, Berkeley

PhD. Information Science. Aug 2019 – Present

- Advisor: David Bamman.

Stanford University

MS. Computer Science. Depth area in Artificial Intelligence. Sept 2017 – June 2019

BS. Symbolic Systems with distinction. Concentration in Natural Language. Sept 2014 – June 2018

- Study abroad at University of Oxford, Winter 2017.

Peer-reviewed Papers

* indicates equal contribution

1. **Li Lucy** & David Bamman. Gender and Representation Bias in GPT-3 Generated Stories. To appear, *Workshop on Narrative Understanding (WNU)* at the North American Association for Computational Linguistics (NAACL), 2021.
2. **Li Lucy** & David Bamman. Characterizing English variation across social media communities with BERT. To appear, *Transactions of the Association of Computational Linguistics (TACL)*, 2021.
3. **Li Lucy***, Dora Demszky*, Patricia Bromley, & Dan Jurafsky. Content Analysis of Textbooks via Natural Language Processing: Findings on Gender, Race, and Ethnicity in Texas U.S. History Textbooks. *AERA Open,* 2020. [Best paper at American Educational Research Association (AERA) Educational Data Science Conference.]
4. Emma Lurie, **Li Lucy**, Masha Belyi, Sofia Dewar, Daniel Rincón, John Baldwin, & Rajvardhan Oak. Investigating Causal Effects of Instructions in Crowdsourced Claim Matching. *Computation + Journalism Symposium (C+J)*, 2020. [non-archival.]
5. **Li Lucy** & Julia Mendelsohn. Using sentiment induction to understand variation in gendered online communities. *Society for Computation in Linguistics (SCiL)*, 2019.
6. **Li Lucy** & Jon Gauthier. Are distributional representations ready for the real world? Evaluating word vectors for grounded perceptual meaning. *Language Grounding for Robotics (RoboNLP) Workshop* at the Association for Computational Linguistics (ACL), 2017.

Presentations

Content Analysis of Textbooks via Natural Language Processing.

Guest lecture for Stanford graduate course, "Doing Digital History." Feb 2021.
Stanford Human-Computer Interaction Lunch Seminar. Feb 2021.
Guest lecture for Stanford graduate course, "Using Data to Describe the World." May 2020.
10th Annual New Directions in Analyzing Text as Data (TADA). Stanford, CA. Oct 2019.

Grants & Awards

National Science Foundation Graduate Research Fellowships Program. April 2019
K. Jon Barwise Award for Distinguished Contributions to the Symbolic Systems Program. June 2018
Stanford Undergraduate Advising & Research (UAR) Small Grant. $1,500. May - June 2018
Symbolic Systems Grants for Education and Research (GEAR). $1,145. Aug 2017
Phi Beta Kappa, elected as junior. May 2017

Experience

Microsoft Research | Research Intern | Montreal, Canada May 2021-Aug 2021

- Fairness, Accountability, Transparency, Ethics (FATE) team.

Stanford Computer Science | Research Assistant | Stanford, CA Jan 2019 – Dec 2019

- Advisors: Dan Jurafsky, Patricia Bromley.
- Investigated the framing and representation of underrepresented groups in history textbooks with linguistics PhD student Dora Demszky.

École Polytechnique Fédérale de Lausanne | Research Intern | Lausanne, Switzerland July 2018 – Sept 2018

- Advisor: Robert West (Data Science Lab)
- Operationalized and analyzed behavioral trends in a political quote dataset using Apache Spark, emotion lexicons, Stanford CoreNLP parsers, and social networks.

Stanford Computer Science | Research Assistant | Stanford, CA April 2017 – June 2018

- Advisors: David Jurgens, Jure Leskovec (Stanford Network Analysis Project), Dan Jurafsky (NLP group)
- Used language and social network features to classify fictional and real relationships with scikit-learn, NLTK, and Keras.

Teaching & Advising

Stanford CS 224U: Natural Language Understanding | Course Assistant April 2019 – June 2019

- Awarded a bonus for being in the top 5% of course assistants in computer science

Stanford Symbolic Systems Program | Advising Fellow Sept 2016 - June 2017, Jan 2019 – June 2019

Stanford EE/CME 103: Introduction to Matrix Methods | Course Assistant Sept 2017 - Dec 2017

Professional Service

Reviewing: AERA Open (2020), NeurIPS Human and Machine in-the-Loop Evaluation and Learning Strategies Workshop (2020), ACL (2021), ACL Workshop on NLP for Positive Impact (2021).
Organizing Committee: Teaching NLP (2021) at NAACL.

Community Service & Outreach

CS Kickstart | Guest Speaker Aug 2020
UC Berkeley Girls in Engineering | Cohort Leader June - July 2020
BAIR Mentoring Program | Mentor March 2020 - Present
Berkeley AI4ALL | Project Developer & TA Aug 2019
Stanford AI4ALL | Research Mentor June 2019 – July 2019
Girls Teaching Girls to Code | NLP Track Mentor, Lead April 2018, April 2019

Skills

Computer Languages: Python, Julia, C++, SQL
Natural Languages: English, Mandarin Chinese, French.
Tools: NLTK, Stanford CoreNLP, SpaCy, scikit-learn, Apache Spark, MTurk, Figure Eight, Keras, TensorFlow, PyTorch.