Lucy Li

Natural language processing, computational sociolinguistics, & computational social science

lucy3_li@berkeley.edu | lucy3.github.io

Education

University of California, Berkeley *Berkeley, CA*

PhD Information Science *Aug 2019 - present*

- Advisor: David Bamman
- Berkeley AI Research (BAIR)

Stanford University *Stanford, CA*

BS Symbolic Systems, MS Computer Science *Sept 2014 - June 2019*

- A coterminal degree, w/ a concentration in language and depth in artificial intelligence.
- Study abroad at University of Oxford, Winter 2017.

Experience

Allen Institute for Artificial Intelligence *Seattle, Washington*

Research Intern *May 2022 - Aug 2022*

- Mentors: Katie Keith, Jesse Dodge
- Mapping scientific domains on the Semantic Scholar and AllenNLP teams.

Microsoft Research *Montreal, Canada*

Research Intern *May 2021 - Aug 2021*

- Mentors: Alexandra Olteanu, Su Lin Blodgett
- Auditing natural language generation systems on the Fairness, Accountability, Transparency, and Ethics (FATE) team.

Stanford Computer Science *Stanford, CA*

Research Assistant *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 *Lausanne, Switzerland*

Research Intern *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 *Stanford, CA*

Research Assistant *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.

Publications

*indicates equal contribution

Li Lucy, David Bamman. Gender and Representation Bias in GPT-3 Generated Stories. *Workshop on Narrative Understanding (WNU) at the North American Association for Computational Linguistics (NAACL)*, 2021.

Li Lucy, David Bamman. Characterizing English variation across social media communities with BERT. *Transactions of the Association of Computational Linguistics (TACL)*, 2021.

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.]

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.]

Li Lucy, Julia Mendelsohn. Using sentiment induction to understand variation in gendered online communities. *Society for Computation in Linguistics (SCiL)*, 2019.

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.

Awards, Fellowships, & Grants

Human-Centered Artificial Intelligence Seed Grant, Stanford HAI (PI: Patricia Bromley) *2021*
Graduate Research Fellowship, National Science Foundation *2019*
K. Jon Barwise Award for Distinguished Contributions, Stanford Symbolic Systems *2018*
Undergraduate Advising & Research (UAR) Small Grant, $1500, Stanford University *2018*
Grants for Education and Research, $1145, Stanford Symbolic Systems *2017*
Phi Beta Kappa, Stanford University *2017*

Presentations

Social NLP.
April 2022. Guest lecture, "Natural Language Processing," University of California, Berkeley.

Characterizing English variation across social media communities with BERT.
June 2021. Guest lecture, "Computational Text Analysis," Barnard College.
Nov 2021. Guest lecture, "Practical Approaches to Data Science with Text," Emory University.

Content Analysis of Textbooks via Natural Language Processing.
Oct 2021. 103rd Anniversary of the School of Information, Berkeley.
Feb 2021. Guest lecture, "Doing Digital History", Stanford.
Feb 2021. Stanford Human-Computer Interaction Lunch Seminar.
May 2021. Guest lecture, "Using Data to Describe the World," Stanford.
May 2020. Guest lecture, "Using Data to Describe the World," Stanford.
Oct 2019. 10th Annual New Directions in Analyzing Text as Data (TADA). Stanford, CA.

Teaching Experience

Stanford CS 224U, Course Assistant, top 5% in computer science *Spring 2019*
Symbolic Systems Program, Advising Fellow *2016 - 2017, 2019*
Stanford EE/CME 103, Course Assistant *Fall 2017*

Service

Professional

Reviewer: ACL Rolling Review (2021-Present), NAACL Student Research Workshop (2022), NAACL Workshop on Understanding Implicit and Underspecified Language (2022), SCiL (2022), CHI (2022), CSCW (2022), ACL (2021), ACL Workshop on NLP for Positive Impact (2021), EMNLP (2021), AERA Open (2020), NeurIPS Human and Machine in-the-Loop Evaluation and Learning Strategies Workshop (2020).

Organizing committee: Teaching NLP (2021) at NAACL.

Community

General: Diaries of Social Data Research (Podcast Host, 2021-Present), Sociologists of Digital Things (Admin, 2021).

Students: Berkeley Undergraduate Research Apprentice Program (2021-Present), BAIR Mentoring Program (2020-2021), CS Kickstart (Speaker; 2020), UC Berkeley Girls in Engineering (Leader; 2020), Berkeley AI4ALL (Mentor; 2019), Stanford AI4All (Mentor; 2019), Girls Teaching Girls to Code (Mentor/Lead; 2018, 2019).

Skills

Computer Languages: Python, Julia, C++, SQL

Natural Languages: English, Mandarin Chinese

Tools: NLTK, Stanford CoreNLP, SpaCy, scikit-learn, Apache Spark, MTurk, Figure Eight, Keras, TensorFlow, PyTorch.