

# EMILY SHENG

ewsheng at gmail • <https://ewsheng.github.io>  
Los Angeles, CA

## EDUCATION

### PH.D. CANDIDATE IN COMPUTER SCIENCE

Aug 2015 - present

*University of Southern California*

Advisors: Dr. Prem Natarajan and Dr. Nanyun Peng

Research interests: fairness in natural language processing, language generation, dialogue systems

### M.S. IN COMPUTER SCIENCE

Aug 2015 - May 2017

*University of Southern California*

### B.A. IN COMPUTER SCIENCE, COGNITIVE SCIENCE

Aug 2010 - May 2014

*University of California, Berkeley*

## PROFESSIONAL EXPERIENCE

### RESEARCH INTERN

May 2019 – April 2020

*Google*

*Mountain View, CA*

Evaluate and mitigate biases towards different demographics in a poetry composition system.

### SOFTWARE ENGINEERING INTERN

May 2018 – Aug 2018

*Google*

*Mountain View, CA*

Evaluate textual similarity across text lengths for bag-of-words and convolutional neural network model variants.

### SOFTWARE ENGINEER

July 2014 - July 2015

*Expect Labs*

*San Francisco, CA*

Prototype classifier for domain-specific named entity recognition to improve a natural language understanding system, and full-stack development of developer platform tools.

### SOFTWARE ENGINEERING INTERN

May 2013 - Aug 2013

*Samsung Telecommunications America*

*San Jose, CA*

Build back end of an analytics prototype project.

## PUBLICATIONS

Sheng, E., Chang, K.-W., Natarajan, P., Peng, N. (2020) "Nice Try, Kiddo": Ad Hominems in Dialogue Systems. To appear in *Proceedings of NAACL 2021*.

Sheng, E., Chang, K.-W., Natarajan, P., Peng, N. (2020). Towards Controllable Biases in Language Generation. In *Proceedings of EMNLP-Findings 2020*.

Sheng, E., & Uthus, D. (2020). Investigating Societal Biases in a Poetry Composition System. In *Proceedings of the 2<sup>nd</sup> Gender Bias in NLP Workshop*.

Sheng, E., Chang, K.-W., Natarajan, P., Peng, N. (2019). The Woman Worked as a Babysitter: On Biases in Language Generation. In *Proceedings of EMNLP 2019*.

Sheng, E., & Natarajan, P. (2018). A Byte-sized Approach to Named Entity Recognition. *arXiv preprint arXiv:1809.08386*.

Sheng, E., Miller, S., Ambite, J. L., Natarajan, P. (2017). A Neural Named Entity Recognition Approach to Biological Entity Identification. In *Proceedings of the BioCreative VI Workshop*.

Sheng, E., Natarajan, P., Gordon, J., & Burns, G. (2017). An Investigation into the Pedagogical Features of Documents. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 109-120).

Gordon, J., Aguilar, S., Sheng, E., & Burns, G. (2017). Structured generation of technical reading lists. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 261-270).

### Oral presentations

Sheng, E., & Uthus, D. (2020). Investigating Societal Biases in a Poetry Composition System. *2<sup>nd</sup> Gender Bias in NLP workshop*.

Sheng, E. (2020). Fairness in Natural Language Processing. *Presented at USC / ISI Research Day 2020*.

Sheng, E., Chang, K.-W., Natarajan, P., Peng, N. (2019). The Woman Worked as a Babysitter: On Biases in Language Generation. *Presented at EMNLP 2019*.

Sheng, E., Miller, S., Ambite, J. L., Natarajan, P. (2017). A Neural Named Entity Recognition Approach to Biological Entity Identification. *Presented at the BioCreative VI Workshop*.

### Poster presentations

Sheng, E., Chang, K.-W., Natarajan, P., Peng, N. (2020) "Nice Try, Kiddo": Ad Hominems in Dialogue Systems. *SoCal NLP Symposium 2021*.

Sheng, E., Chang, K.-W., Natarajan, P., Peng, N. (2020). Towards Controllable Biases in Language Generation. *BlackboxNLP 2020 workshop*.

Sheng, E., Natarajan, P., Gordon, J., & Burns, G. (2017). An Investigation into the Pedagogical Features of Documents. *12th Workshop on Innovative Use of NLP for Building Educational Applications*.

Sheng, E., and Natarajan, P. (2016). An Investigation into the Pedagogical "Value" of Documents. *CRA-W Grad Cohort Workshop and ISI Graduate Student Symposium*.

## RESEARCH EXPERIENCE

**USC / ISI: Natural Language Processing** Aug 2015 - present  
*Graduate Research Assistant at University of Southern California / Information Sciences Institute*  
Recent work on evaluating and mitigating biases in language generation, prior work on information extraction.

**UC Berkeley / ICSI: Resolving prepositional phrase attachment ambiguity** Jan 2014 - May 2014  
*Research project at University of California, Berkeley / International Computer Science Institute*  
Survey of lexical, semantic, and contextual methods to resolve ambiguity (adv. Dr. Jerome Feldman).

**UC Berkeley / Walker Lab: Sleep study** June 2012 - May 2013  
*Research Assistant at University of California, Berkeley*  
EEG, MRI, and behavioral tests to study effect of sleep on adolescents.

**UC Berkeley / Concepts and Cognition Lab: Yahoo Answers study** Aug 2011 - May 2013  
*Research Assistant at University of California, Berkeley*  
Extract features of up-voted Yahoo Answers to find those favored in "good" explanations.

## PROFESSIONAL ACTIVITY

**SEMINAR COORDINATOR** Sept 2019 – Aug 2020  
*USC Information Sciences Institute Natural Language Seminar* Marina del Rey, CA

**STUDENT CO-CHAIR**  
*2019 SoCal NLP Symposium*

Sept 2019  
*Los Angeles, CA*

**REVIEWER**

*\*SEM 2019, SoCal NLP Symposium 2019, ACL 2020, EMNLP 2020 (Outstanding Reviewer), NAACL 2021, ACL 2021*

**TEACHING EXPERIENCE**

**TEACHING ASSISTANT**  
*Introduction to Computing course*

Aug 2015 – May 2016  
*University of Southern California*