

# SIMENG SUN

CS Building 266  
Amherst, MA, 01002  
☎ (215) 594 7676  
✉ simengsun@umass.edu  
🏠 /~simengsun

---

## Education

- 2019–present **Ph.D. Student**, University of Massachusetts, Amherst.  
Advisor: Mohit Iyyer
- 2017–2019 **M.S.E. in Computer and Information Science**, University of Pennsylvania.  
Advisor: Ani Nenkova
- 2013–2017 **B.E. in Computer Science and Technology**, Beihang University.  
Advisor: Hui Zhang
- 2015 **Visiting student**, Trinity College Dublin, Ireland.

---

## Research Experience

- 2022 **Research Intern**, *Facebook Research, FAIR Accel*.  
Summer Topic to be decided.
- 2021 **Research Intern**, *Facebook Research, FAIR Accel (Remote)*.  
Summer Work on alternative input signals to increase the transfer in multilingual machine translation.
- 2020 **Research Intern**, *Adobe Research (Remote)*.  
Summer Proposed a new intent-guided text-infilling model and constructed an intent-labeled dataset. Conducted intrinsic and extrinsic human evaluation for a joint human-AI story generation task.
- 2019–present **Research Assistant**, *University of Massachusetts, Amherst*.  
Work on language modeling and machine translation.
- 2017-2019 **Research Assistant**, *University of Pennsylvania*.  
Worked on evaluations of text summarization. Proposed new automatic metrics for better system comparison. Worked on document-level named entity recognition.
- 2018 **NLP Intern**, *Educational Testing Service*.  
Summer Worked on determiner error correction and native language identification for automatic essay scoring. Refreshed old rule-based framework with machine learning models.
- 2016-2017 **Undergraduate Research Assistant**, *Beihang University*.  
Worked on author name disambiguation for Chinese scientific search engine *Kejso*. Proposed an unsupervised method to disambiguate names of Chinese scholars.

---

## Publications

Simeng Sun, Katherine Thai, Mohit Iyyer **ChapterBreak: A Challenge Dataset for Long-Range Language Models**. Short, NAACL 2022.

Simeng Sun, Brian Dillon, Mohit Iyyer **How Much Do Modifications to Transformer Language Models Affect Their Ability to Learn Linguistic Knowledge?** Short, Insights workshop @ ACL 2022.

Simeng Sun, Angela Fan, James Cross, Vishrav Chaudhary, Chau Tran, Philipp Koehn, Francisco Guzman **Alternative Input Signals Ease Transfer in Multilingual Machine Translation.** Long, ACL 2022.

Simeng Sun, Kalpesh Krishna, Mohit Iyyer **Do Long-Range Language Models Actually Use Long-Range Context?** Long, EMNLP 2022.

Simeng Sun, Wenlong Zhao, Varun Manjunatha, Rajiv Jain, Vlad Morariu, Franck Dernoncourt, Balaji Vasan Srinivasan, Mohit Iyyer **IGA: An Intent-Guided Authoring Assistant.** Long, EMNLP 2022.

Sumanta Bhattacharyya, Amirmohammad Rooshenas, Subhajit Naskar, Simeng Sun, Mohit Iyyer, Andrew McCallum. **Energy-Based Reranking: Improving Neural Machine Translation Using Energy-Based Models.** Long, ACL 2021.

Simeng Sun and Mohit Iyyer **Revisiting Simple Neural Probabilistic Language Models.** Short, NAACL 2021.

Weiqiu You\*, Simeng Sun\*, Mohit Iyyer. **Hard-Coded Gaussian Attention for Neural Machine Translation.** Long, ACL 2020. (\* Equal contribution)

Simeng Sun, Ani Nenkova. **The Feasibility of Embedding Based Automatic Evaluation for Single Document Summarization.** Short, EMNLP 2019

Simeng Sun, Ori Shapira, Ido Dagan, Ani Nenkova. **How to Compare Summarizers without Target Length? Pitfalls, Solutions and Re-Examination of the Neural Summarization Literature.** NeuralGen Workshop, NAACL 2019

Simeng Sun, Hui Zhang, Ning Li, Yong Chen. **Name Disambiguation for Chinese Scientific Authors with Multi-Level Clustering.** IEEE International Conference on Computational Science and Engineering (CSE) 2017

---

## Awards

- 2020 The Victor Lesser Graduate Scholarship in Artificial Intelligence
- 2019 UMass CICS PhD Fellowship
- 2019 NeuralGen workshop 2019 travel grant
- 2019 UPenn 2019 SEAS Outstanding Research Award
- 2014–2016 Beihang SCSE Merit Student Scholarship
- 2013 Outstanding Freshman Scholarship

---

## Teaching Experience

- UMass COMPSCI685 Advanced Natural Language Processing (TA), 2020 Fall
- UMass COMPSCI585 Introduction to NLP (TA), 2019 Fall
- UPenn CIS520 Machine Learning (TA, Head TA), 2018–2019

UPenn CIS521 Introduction to Artificial Intelligence (TA), 2018 Fall

---

## Service

- 2022 ARR 2022 January, February
- 2021 ARR 2021 November, December
- 2021 Review for ACL2021, NLPCC2021, EMNLP2021
- 2021 FAIR Accel intern social hour co-organizer
- 2021-present OTW Open Doors tech volunteer
  - 2021 UMass CICS 2021 PhD Candidate Friday volunteer
  - 2020 UMass CICS mentor for new PhD student
  - 2020 UMass CICS 2020 PhD Candidate Friday host
  - 2020 Review for WNGT 2020
  - 2020 PC member: EvalNLGEval INLG2020
- 2014 - 2016 Beihang University Piano Club publicity chair
- 2015 IAAF Beijing 2015 volunteer

---

## Technical Skills

- Coding Python, Java, C/C++, Matlab,  $\LaTeX$
- Libraries PyTorch, Scikit-Learn, Numpy, Scipy, TensorFlow