

# Shuguang Chen

University of Houston, 4800 Calhoun Rd, Houston, TX 77004, United States

☎ (+1) 806-459-0724 | ✉ [sgchen.cs@gmail.com](mailto:sgchen.cs@gmail.com) | 🌐 [sgchen.info](http://sgchen.info) | 🗣 [nehcgs](#) | 📺 [shuguang-chen](#)

## EDUCATION

---

### University of Houston

Ph.D. in Computer Science

- Advisor: Prof. Thamar Solorio
- Research Area: Natural Language Processing (NLP)
- Dissertation: Named Entity Recognition on Social Media

Houston, TX

Aug. 2018 - Dec. 2022

### Beijing Forestry University

B.S. in Computer Science

- Advisor: Prof. Meng Wei
- Thesis: Music Generation Using Recurrent Neural Networks

Beijing, China

Sep. 2014 - Jul. 2018

## RESEARCH INTEREST

---

My research interest lies in Natural Language Processing (NLP) with a special focus on neural sequence labeling. My research aims to overcome linguistic challenges in noisy environments, making NLP models **resilient**, **robust**, and **reliable** under various conditions.

## WORK EXPERIENCE

---

### University of Houston, RiTUAL Lab

Research Assistant, Advisor: Prof. Thamar Solorio

- Sequence labeling on user-generated text
- Data transformation for model robustness
- Knowledge transfer in linguistic code-switching

Houston, TX

Aug. 2019 - Dec. 2022

### Microsoft Research, Deep Learning Group

Research Intern, Mentor: Dr. Hao Cheng

- Research Topic: Temporal Text Modeling for Event Detection
- Investigated the impact of temporal information on dialogue data.
- Proposed methods to induce temporal information into transformers.

Redmond, WA

May. 2022 - Aug. 2022

### Melax Technologies Inc

NLP Intern, Mentor: MS Jingqi Wang

- Research Topic: Document Classification and Information Extraction
- Developed a data annotation platform for NLP tasks, including named entity recognition and question answering.
- Studied named entity recognition and relation extraction with transformers for biomedical data.

Houston, TX

May. 2021 - Aug. 2021

## SELECTED PROJECT

---

### Data Augmentation for Cross-domain Entity Recognition [[Github](#)]

Advisor: Prof. Thamar Solorio

- Research Topic: Data Augmentation for Low-resource Named Entity Recognition
- Investigated data transformation with GANs to improve model robustness.
- Proposed a new method to transfer the data from low-resource to high-resource domains.

Houston, TX

Feb. 2021 - Sep. 2021

### Multimodal Named Entity Recognition on Social Media [[Github](#)]

Advisor: Prof. Thamar Solorio

- Research Topic: Multimodal Named Entity Recognition
- Studied multimodal information representation, extraction, and fusion.
- Proposed methods to reduce performance degradation by fusing visual and textual information.

Houston, TX

Sep. 2019 - Sep. 2021

### Named Entity Recognition on Diachronic Twitter Data [[Github](#)]

Advisor: Prof. Thamar Solorio

- Research Topic: Named Entity Recognition under Temporal Drift
- Probed the impact of temporal drift on entity memorization and context generalization.
- Presented a method to efficiently update model parameters with the most informative data.

Houston, TX

Jun. 2020 - Apr. 2021

## HONORS & AWARDS

---

### AWARDS & SCHOLARSHIPS

- 2022 **Cullen Graduate Student Success Fellowship**, University of Houston *Houston, TX*
- 2018 **Outstanding Graduate Awards**, Beijing Forestry University *Beijing, China*

### ACADEMIC AND SCIENTIFIC COMPETITIONS

- 2018 **Academic Merit Scholarship**, Beijing Forestry University *Beijing, China*
- 2016 **Bronze Metal**, China Collegiate Programming Contest (CCPC) *Hangzhou, China*
- 2016 **2nd prize**, The 7th Blue Bridge Cup National Software Competition *Beijing, China*

## PROFESSIONAL SERVICE

---

### RESEARCH ACTIVITIES

- **Program Committee Member** for EMNLP 2020, ACL 2020, MCPR 2021, NAACL 2021, W-NUT 2021.
- **Co-organizer** for the fifth workshop on Computational Approaches to Linguistic Code-Switching [[CALCS](#)].
- **Webmaster** for Linguistic Code-switching Evaluation Benchmark [[LinCE](#)]

### TEACHING & MENTORING

- **COSC 3320 Algorithms and Data Structures**, Instructor: Prof. Rakesh Verma
- **COSC 3360 Fundamentals of Operating Systems**, Instructor: Prof. Jehan-Francois Paris
- **COSC 6358 Interactive Game Development**, Instructor: Prof. ChangHoon Yun

### LICENSES & CERTIFICATIONS

- **Deep Learning Specialization**, Coursera, Instructor: Andrew Ng [[Certificate](#)]
- **Build Basic Generative Adversarial Networks (GANs)**, Coursera, Instructor: Sharon Zhou [[Certificate](#)]
- **Computer Vision**, Udacity, Instructor: Suzanne Camacho [[Certificate](#)]

## PUBLICATION

---

### PRE-PRINTS

- **CALCS 2021 Shared Task: Machine Translation for Code-Switched Data**  
**Shuguang Chen**, Gustavo Aguilar, Anirudh Srinivasan, Mona Diab and Tamar Solorio.  
[arXiv:2202.09625](#)

### PEER-REVIEWED PUBLICATIONS

- **Style Transfer as Data Augmentation: A Case Study on Named Entity Recognition**  
**Shuguang Chen**, Leonardo Neves, Tamar Solorio.  
**EMNLP 2022**
- **A Simple Approach to Jointly Rank Passages and Select Relevant Sentences in the OBQA Context**  
Man Luo, **Shuguang Chen**, Chitta Baral.  
**NAACL 2022** SWR
- **Data Augmentation for Cross-Domain Named Entity Recognition**  
**Shuguang Chen**, Gustavo Aguilar, Leonardo Neves, Tamar Solorio.  
**EMNLP 2021**
- **Can images help recognize entities? A study of the role of images for Multimodal NER**  
**Shuguang Chen**, Gustavo Aguilar, Leonardo Neves, Tamar Solorio.  
**EMNLP 2021** W-NUT
- **Mitigating Temporal-Drift: A Simple Approach to Keep NER Models Crisp**  
**Shuguang Chen**, Leonardo Neves, Tamar Solorio.  
**NAACL 2021** SocialNLP
- **Proceedings of the Fifth Workshop on Computational Approaches to Linguistic Code-Switching**  
Tamar Solorio, **Shuguang Chen**, Alan W Black, Mona Diab, Sunayana Sitaram, Victor Soto, Emre Yilmaz.  
**NAACL 2021** CALCS