

# Shuguang Chen

610 Purdue Mall, West Lafayette, IN 47907, 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

Houston, TX

Aug. 2018 - Dec. 2022

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

### Beijing Forestry University

B.S. in Computer Science

Beijing, China

Sep. 2014 - Jul. 2018

## RESEARCH INTEREST

I specialize in Natural Language Processing (NLP) with a primary focus on **cross-source information extraction**. I work to facilitate comprehensive information access across a broad spectrum of data sources, ranging from cross-modal, cross-domain, and cross-lingual data to diachronic and low-resource data. My research seeks to enable machines to extract accurate and trustworthy information embedded in data originating from heterogeneous sources, advancing their capability and adaptability in diverse and dynamic environments.

## WORK EXPERIENCE

### Purdue University, PMML Lab

Postdoctoral Researcher, Advisor: Prof. Guang Lin

West Lafayette, IN

Aug. 2023 - Present

- Cross-source Information Extraction
- Efficient/Low-Resource Methods for NLP
- NLP for Science

### Microsoft Research, Deep Learning Group

Research Intern, Mentor: Dr. Hao Cheng

Redmond, WA

May. 2022 - Aug. 2022

- 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.

### Melax Technologies Inc

NLP Intern, Mentor: Dr. Jingqi Wang

Houston, TX

May. 2021 - Aug. 2021

- 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.

## SELECTED PROJECT

### Style Transfer as Data Augmentation [[Github](#)]

Advisor: Prof. Thamar Solorio

Houston, TX

Nov. 2021 - Aug. 2022

- Research Topic: Data Augmentation for Low-resource Named Entity Recognition
- Proposed a novel approach to establish a style mapping of text between different domains.
- Presented a new method to transfer data across domains by changing stylistic attributes.

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

Advisor: Prof. Thamar Solorio

Houston, TX

Feb. 2021 - Sep. 2021

- 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.

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

Advisor: Prof. Thamar Solorio

Houston, TX

Sep. 2019 - Sep. 2021

- 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.

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

Advisor: Prof. Thamar Solorio

Houston, TX

Jun. 2020 - Apr. 2021

- 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.

## 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**, ACL, EMNLP, NAACL, W-NUT, and ARR.
- **Co-organizer**, The Fifth workshop on Computational Approaches to Linguistic Code-Switching [[CALCS](#)].
- **Webmaster**, Linguistic Code-switching Evaluation Benchmark [[LinCE](#)]

### 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

---

### PEER-REVIEWED PUBLICATIONS

- **Style Transfer as Data Augmentation: A Case Study on Named Entity Recognition**  
Shuguang Chen, Leonardo Neves, Thamar 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, Thamar Solorio.  
EMNLP 2021
- **Can images help recognize entities? A study of the role of images for Multimodal NER**  
Shuguang Chen, Gustavo Aguilar, Leonardo Neves, Thamar Solorio.  
EMNLP 2021 W-NUT
- **Mitigating Temporal-Drift: A Simple Approach to Keep NER Models Crisp**  
Shuguang Chen, Leonardo Neves, Thamar Solorio.  
NAACL 2021 SocialNLP

### PRE-PRINTS

- **Context-aware Adversarial Attack on Named Entity Recognition**  
Shuguang Chen, Leonardo Neves, Thamar Solorio.  
arXiv:2309.08999
- **CALCS 2021 Shared Task: Machine Translation for Code-Switched Data**  
Shuguang Chen, Gustavo Aguilar, Anirudh Srinivasan, Mona Diab, Thamar Solorio.  
arXiv:2202.09625