Chao Jiang

(+1) 434-227-9935
chaojiang06@gmail.com
https://chaojiang06.github.io

Education

School of Interactive Computing, Georgia Institute of Technology Atlanta, GA
Ph.D. Candidate in Computer Science August 2020 – December 2024 *(expected)*
- Advisor: Prof. Wei Xu
- Thesis Committee: Prof. Alan Ritter and Prof. Kartik Goyal at Georgia Tech,
  Prof. Nanyun Peng at UCLA, Dr. Cheng Li at Google Research
- Research Areas: Natural Language Processing (NLP), Large Language Models (LLMs)

Dept. of Computer Science and Engineering, Ohio State University Columbus, OH
Ph.D. Student in Computer Science and Engineering (Transfer out) 2018 - 2020
- Advisor: Prof. Wei Xu

Department of Computer Science, University of Virginia Charlottesville, VA
M.S. in Computer Science 2016 - 2018
- Advisor: Prof. Kai-Wei Chang
- GPA: 3.97 / 4.00

School of Electrical and Information Engineering, Tianjin University China
B.E. in Communication Engineering (equiv. to Electrical and Computer Eng.) 2016 - 2018

Professional Summary

I am a final-year Ph.D. student at Georgia Tech NLP group, advised by Prof. Wei Xu. My dissertation is on "Studying text revision in scientific writing". My Ph.D. work covers the full spectrum of text revision, ranging from fine-grained local edits [5] to global document restructuring across multiple drafts [4,7], as well as revisions made for different purposes, e.g., improving readability [1], grammar, style, and supporting content updates [3] etc.

During the Ph.D., I gained experience across the full NLP and ML pipeline, including creating high-quality datasets [1,3,4,5,6,7,8], developing SOTA models [5,6,8] and text generation systems [4,5,6], designing evaluation metrics [1], and steering the output of language models [2], all of which are closely aligned with the latest advancements in LLMs.

Additionally, I have experience in pre-training multilingual language models from scratch on TPU and multi-node GPU clusters, which served as the backbone for a multilingual information extraction system that achieved top results in a government-funded project.

Selected Publications

1. MedReadMe: A Systematic Study for Fine-grained Sentence Readability in Medical Domain
   Chao Jiang, Wei Xu
   EMNLP 2024, long paper and **Oral Presentation**
   Received a meta-review score of 5

2. Frustratingly Easy Label Projection for Cross-lingual Transfer
   Yang Chen, **Chao Jiang**, Alan Ritter, Wei Xu
   Findings of ACL 2023, long paper

3. arXivEdits: Understanding the Human Revision Process in Scientific Writing
   Chao Jiang, Wei Xu, Samuel Stevens
   EMNLP 2022, long paper

4. Improving Large-scale Paraphrase Acquisition and Generation
   Yao Dou, **Chao Jiang**, Wei Xu
   EMNLP 2022, long paper

5. Neural semi-Markov CRF for Monolingual Word Alignment
   Wuwei Lan*, **Chao Jiang***, Wei Xu. (* indicates equal contribution)
   ACL 2021, long paper and **Oral Presentation**

SELECTED PUBLICATIONS (COUT.)	6. Neural CRF Model for Sentence Alignment in Text Simplification **Chao Jiang**, Mounica Maddela, Wuwei Lan, Yang Zhong, Wei Xu ACL 2020, long paper and **Oral Presentation** Received >100 citations
	7. Discourse Level Factors for Sentence Deletion in Text Simplification Yang Zhong, **Chao Jiang**, Wei Xu, Junyi Jessy Li AAAI 2020, long paper and **Oral Presentation**
	8. Learning Word Embeddings for Low-Resource Languages by PU Learning **Chao Jiang**, Hsiang-Fu Yu, Cho-Jui Hsieh, Kai-Wei Chang NAACL 2018, long paper and **Oral Presentation**

INTERNSHIPS

Apple AIML, Siri Information Intelligence | Seattle, WA May 2024 - August 2024

Machine Learning Engineer Intern

- Conducted research to improve retrieval performance for short queries in the Spotlight ranking on the iOS and macOS platforms.
- Delivered a pairwise ranking model achieving a 4.1% reduction in test loss on over 1 million test instances, while adhering to strict model size constraints.

AWARDS AND HONORS

Third Place, Georgia Tech College of Computing Graduate Poster Symposium	2024
Student Travel Scholarship for AAAI 2020 Conference	2020
Outstanding Milestone Project, Tianjin University	2015
Merit Student Scholarship, Tianjin University	2015
Honorable Mention Prize, Mathematical Contest in Modeling	2014
Second Place, China Undergraduate Mathematical Contest in Modeling	2013

RESEARCH INTERESTS

Natural Language Processing, Text Edits and Revision, Model-assisted Dataset Creation, Structured Models in Machine Learning, Large Language Models, and Generative AI.

ACADEMIC ACTIVITIES

Regular Reviewer for: ACL Rolling Review, ACL, EMNLP, NAACL, AAAI, COLING

Guest Lecture, CSE5525 Speech and Language Processing	2020
Teaching Assistant for: CS7650 Natural Language Processing, Georgia Tech	2022-2024
CSE2331 Algorithms, Ohio State University	2018
CS4501-003 Machine Learning, University of Virginia	2017

SKILLS

Programming: Python, C/C ++, Java, Bash, SQL, MATLAB.

Framework: PyTorch, TensorFlow, Transformers, Git, Pandas, spaCY, NLTK, Scikit-Learn CoreNLP, AWS, OpenAI Toolkits.