Zhisheng Jin

Rochester, NY | jinzhisheng2099@outlook.com | chihshengj.github.io |

About Me

Hi! I'm a Master's student in Computational Linguistics at the University of Rochester. My current research focuses are information extraction and transformer language modeling, but I'm also interested in psycholinguistics, formal semantics, NLP in low-resource languages, and computation theory.

Education

University of Rochester, MS in Computational Linguistics　　　　　　　　　　　　Sept 2023 – May 2025
- GPA: 4.00/4.00
- **Coursework:** Machine Learning, End to End Deep Learning, Computational Linguistics, Large Language Models

Fudan University, BA in Chinese Language　　　　　　　　　　　　　　　　　　Sept 2019 – June 2023
- GPA: 3.52/4.00
- **Coursework:** Psycholinguistics, Phonetics, Philosophy of Artificial Intelligence, Data Structures and Algorithms

Professional Experience

Research Assistant, FACTS.lab – University of Rochester　　　　　　　　　　　　Apr 2024 – Present
- Participated in the study on Cross-Document Event-keyed Machine Summarization.
- Currently building a comprehensive dataset on document-level and cross document event extraction.

Project Management Intern, Shanghai AI Laboratory　　　　　　　　　　　　　July 2022 – Sep 2022
- Graduate student data management with Python and Excel.
- Organized a one week long workshop on Large Language Models and Multimodal Learning.
- Participated in organizing the World Artificial Intelligence Conference 2022 in Shanghai.

Publications

Cross-Document Event-Keyed Summarization (pre-print)　　　　　　　　　　　　Oct 2024

William Gantt Walden, Pavlo Kuchmiichuk, Alexander Martin, *Chihsheng Jin*,
Angela Cao, Claire Sun, Curisia Allen, Aaron Steven White

arxiv.org/abs/2410.14795

Projects

Multiple Context Free Grammar Parser　　　　　　　　　　　　　　　　　　　Github link
- Developed an agenda-based parser that can parse any customized multiple context free languages under 3 degrees.
- Tools Used: Python, Pytest

Retrieval Based Document-level Event Extraction　　　　　　　　　　　　　　Github link
- Developed a framework for extracting event information from a document accompanied by an identified event trigger. The framework itself relies on a BERT-like encoder model that is trained on a contrastive learning objective.
- Tools Used: SpanBERT, LLM APIs, Contrastive Learning

Dynamics in the phonological encoding in bilingual speech production　　　　　　　PDF
- This is my senior thesis. I used a picture-naming paradigm to test whether the phonological mapping status of cognates in Mandarin and Shanghai Dialect will take an effect on the speech production process and found a cognate facilitation effect when the phonological forms of the cognates match the basic mapping rules regardless of the discrepancy of their phonetic forms.
- Tools Used: PsychoPy, Python, Praat

Professional Skills

Languages: Python, LaTeX, Haskell, HTML, CSS, R

Tools: Git, Docker, PostgreSQL, AWS, MTurk, PyTorch, HuggingFace Transformers, NLTK, PyTest, LLM APIs

Softwares: Praat, PsychoPy, Ableton Live, Photoshop

Additional skills: Audio recording, Mixing and mastering, Composition and music production, Photography.