Chihsheng Jin

Shanghai, China | chihsheng3421@gmail.com | chihshengj.github.io |

Research Interests

Large Language Models, Encoded Reasoning, Steganography, Information Extraction, Formal Semantics, Theory of Computation, Psycholinguistics

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, Language Typology, Data Structures and Algorithms

Research Experience

Research Assistant, HLTCOE – Johns Hopkins University   Aug 2025 – Present
- Developed a novel LLM-based linguistic steganography system that supports flexible stego generation in various genres.
- Annotated factual claims from videos for multimodal model retrieval augmented generation evaluation.

Research Assistant, FACTS.lab – University of Rochester   Apr 2024 – May 2025
- Improved the annotation workflow for Cross-Document Event-keyed Machine Summarization.
- Conducted experiments on using semantic representations of summaries to retrieve related documents.
- Maintained a codebase for the data collection for a project on cross-document event extraction and upgraded it with a cleaner workflow and relational database support.

Professional Experience

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

Publications

Seeing Through the MiRAGE: Evaluating Multimodal Retrieval Augmented Generation   Oct 2025

Alexander Martin, William Gantt Walden, Reno Kriz, Dengjia Zhang, Kate Sanders, Eugene Yang, *Chihsheng Jin*, Benjamin Van Durme

Arxiv pre-print

Cross-Document Event-Keyed Summarization   Oct 2024
(Presented at XLLM@ACL2025 and PEER2025)

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

ACL Anthology

Projects

A Better Tree-of-Thought Prompting scheme   Github link
- A fork of the original Tree-of-Thought project. A new framework was implemented where a symbolic solver will

be auto-generated based on the task to help cut down branches during the BFS, which drastically improve the efficiency of the original framework.

Multiple Context Free Grammar Parser   Github link

- I developed an agenda-based parser that can parse any customized multiple context free languages under 3 degrees.

Event Argument Retrieval with LLMs   PDF

- Master's thesis. I evaluated the argument retrieval capabilities of most powerful LLMs against the FAMuS dataset on two variants of the CEAF metric. To mitigate the inaccurate extractions from LLMs, I developed a retrieval inspired pipeline that significantly improved the quality of the extractions.

Phonological Mapping influences bilingual speech activation   PDF

- Senior thesis. A Chinese character-naming experiment was used to test whether the phonological mapping status of cognates in Mandarin and Shanghai Dialect has an effect on the speech production process. Experiments show that the phonological mapping interferes with the common cognate facilitation effect observed by previous studies regardless of how similar the phonetic forms of the cognates are.

Professional Skills

Languages: Python, Javascript, HTML/CSS, LaTeX

Tools: Linux, Git, Slurm, PostgreSQL, Docker, PyTorch, Lightning, Amazon MTurk

Softwares: Praat, PsychoPy, Ableton Live, Photoshop

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