JIAJUN SHEN

(+86) 131-8522-6852 ◇ Lausanne, Switzerland
jiajun.shen@epfl.ch

EDUCATION

University of Chinese Academy of Sciences Sept 2021 - June 2025
Bachelor of Engineering in Artificial Intelligence **GPA: 3.96/4.0** | **Grade : 91.7/100** | **Ranking: 2/70**
- Coursework: Image Processing and Computer Vision: 93/100, Pattern Recognition and Machine Learning: 96/100, Natural Language Processing: 90/100, Math Courses Average: 94.5/100

University of California, Berkeley Jan 2024 - May 2024
Visiting student in Data Science **GPA: 4.0/4.0**
- Coursework: Introduction to Mathematical Economics: A, Efficient Algorithms and Intractable Problems: A+, The Structure and Interpretation of Computer Programs: A+

École Polytechnique Fédérale de Lausanne Sept 2025 - Present
Master of Science in Digital Humanities

PUBLICATION

Jiajun Shen, Tong Zhou, Yubo Chen, Delai Qiu, Shengping Liu, Kang Liu, and Jun Zhao. 2025. Transparentize the Internal and External Knowledge Utilization in LLMs with Trustworthy Citation. In Findings of the Association for Computational Linguistics: ACL 2025, pages 17858–17877, Vienna, Austria. Association for Computational Linguistics.

Jiajun Shen, Tong Zhou, Yubo Chen, Kang Liu, and Jun Zhao. 2025. CiteLab: Developing and Diagnosing LLM Citation Generation Workflows via the Human-LLM Interaction. In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations), pages 490–501, Vienna, Austria. Association for Computational Linguistics.

INTERNSHIP

Institute of Automation, Chinese Academy of Sciences, Beijing Feb 2024 - Jul 2025
Student Intern, The Key Laboratory of Cognition and Decision Intelligence for Complex Systems
Supervisor: Jun Zhao
Conducted research as an intern at CASIA, contributing to the publication of a paper on LLM Citation and Information Retrieval.

RESEARCH EXPERIENCE

Modular Toolkit for Large Language Model Citation Generation Mar 2024 - Jul 2024
Student Intern, advised by Prof. Jun Zhao, Institute of Automation, Chinese Academy of Sciences *Beijing, China*

- Proposed a framework for LLM citation tasks, consisting of 4 main modules and 14 components, with 16 functions to manage citation generation processes.
- Introduced Citekit, a toolkit with 11 predefined recipes to support various citation-generation methods.
- Evaluated 11 baselines on 2 state-of-the-art LLMs and proposed a new method.

Survey on Robustness of Retrieval Augmented Generation Feb 2024 - Jun 2024
Research Practice, advised by Prof. Jiajun Zhang, University of Chinese Academy of Sciences *Beijing, China*

- Surveyed 18 RAG experiments, 20 state-of-the-art RAG methods, and 8 evaluation metrics or frameworks.
- Summarized 8 important aspects that affect the robustness of RAG across its 4 main processes.

Transparentize the Internal and External Knowledge Utilization in LLMs with Trustworthy Citation Jul 2024 - Feb 2025
Student Intern, advised by Prof. Jun Zhao, Institute of Automation, Chinese Academy of Sciences *Beijing, China*

- Introduced a new task for citation generation, evaluating helpfulness, faithfulness, and trustworthiness with five complementary metrics.
- Developed a paradigm incorporating multi-scenario data generation and interpretability-focused alignment for trustworthy reference citation.
- Evaluated six baselines and three LLMs across four scenarios, showing significant improvements in reference quality and trustworthiness.

PROJECTS

Neural Dynamics Modeling and Simulation Oct 2023 - Dec 2023
Cognitive Neuroscience, University of Chinese Academy of Sciences *Beijing, China*

- Modeled neural systems using BrainPy, implementing the LIF neuron model and continuous attractor networks.
- Reproduced a research paper by developing a continuous attractor network with feedback inhibition for input tracking.

Remote Sensing Image Segmentation Oct 2022 - Jan 2023
Pattern Recognition and Machine Learning, University of Chinese Academy of Sciences *Beijing, China*

- Designed segmentation models in multiple categories, including 7 categories such as construction, arable land, woodland, water body, road, grassland, and others.
- used and combine methods including CNN(PSPNet, FCN) and K-Means (Fuzzy K-means, NaiveBayes-based and CNN-based K-means)
- maximum 51.97% mIoU and 81.84% accuracy

Relation Extraction and Knowledge Modelling Dec 2023 - Jan 2024
Knowledge Engineering, University of Chinese Academy of Sciences *Beijing, China*

- Designed and implemented an ontology and knowledge graph modeling system, leveraging BERT for named entity recognition and CNN for relation classification.
- Achieved a 0.787 F1-score in end-to-end relation extraction from unstructured natural language texts.

SKILLS

Programming Languages	**Python** (Working Language), **C**, **C++**, **MySQL**, **html5**, LaTeX
Languages	**Mandarin** (native), **English** (fluent, IELTS 8.0/9.0), **French** (intermediate)
Certificate	**GRE** (V: 161, Q: 170, AW: 3.5)