

SIJIA GE

☎ +1(720) 530-7370 ◊ 📍 Boulder, CO ◊ ✉ sjiage007@gmail.com ◊ 🔗 [konic-nlp.github.io](https://github.com/konic-nlp) ◊ 🌐 [LinkedIn](#) ◊ 📄 [Google Scholar](#)

EDUCATION

MS in Computational Linguistics, University of Colorado-Boulder | **GPA: 3.969/4.0** 08/2021 - 12/2023
Courses: Natural language Processing, Machine Learning, Neural Networks and Deep Learning, Design Analysis of Algorithm, Semantics and Pragmatics, Object-Oriented Analysis and Design, Computational Models of Discourse, Methods in Cognitive Science
MA in Computational Linguistics, Nanjing Normal University | **GPA: 86.3/100** 09/2016 - 06/2019
Courses: Language Statistics Method, Corpus Linguistics, Chinese Information Processing, Database&Programming

EXPERIENCE

ML/AI Engineer 01/2024 - Present
Rippey.ai *Louisville, CO*

- **ML End-to-End Workflow:** Established an end-to-end ML workflow using Data Version Control, FastAPI, Docker, and AWS Sagemaker, encapsulating data preprocessing, service wrapping, and deployment as an inference endpoint on Sagemaker for production into one command.
- **Email & Document Automation**[\[demo\]](#): Deployed lightweight classification and extraction models for unstructured email and multi-modality document automation. The classification model achieved a performance of **84** compared to the large language model's performance of 89. Improved the extraction model for emails, increasing the F1 score for shipment details from **69% to 82%** despite limited data; leveraged document layout features and parsing table data to enhance document extraction.
- **Data Management:** Assisted in designing a data management system and data annotation processes.

Student Research Assistant 05/2022 - 12/2023
CLEAR Lab, supervised by Martha Palmer, University of Colorado-Boulder *Boulder, CO*

- **Event Similarity**[\[code\]](#): Engaged in research funded by **DARPA** aimed at calculating event similarity across participant and event mappings within the **WikiData** ontology knowledge graph.
- **UMR-Writer**[\[code\]](#)[\[paper\]](#): Developed an innovative keyboard interface for Semantic Role Labeling annotation, enhancing user experience and simplifying the annotation process from five steps to a single editing command. This enhancement led to a noteworthy **60%** increase in annotation efficiency. The interface accommodates multiple languages.
- **Ancient Chinese NER**[\[code\]](#)[\[paper\]](#): Utilized the siku-BERT pre-trained model to perform NER and sentence segmentation as multi-task sequence labeling on ancient Chinese classics. Achieved comparable results in a single round of training, with NER F-1 score reaching **87%** and sentence segmentation F-1 score at **92%**.
- **Machine Learning Course Manager:** Built scripts for automated grading and code testing, developed comprehensive assignment rubrics, skillfully addressed student concerns, managed logistical matters, and maintained the Canvas course page.

Product Manager Intern 03/2021 - 06/2021
Beijing Lingosail Tech Co., Ltd. *Beijing, China*

- **ShuYuBao Product Design & Management:** Assumed a leadership role in designing a prototype for a computer-assisted translation product and overseeing the development lifecycle. Formulated a robust profit model that contributed to a remarkable **40%** profit margin, surpassing initial projections and resulting in an outstanding **20%** increase in active users compared to anticipated figures.

SKILLS

Programming Languages & Web Development: Python, Java, JavaScript, Django, Flask, HTML, Heroku, FastAPI, CSS, JQuery
Machine Learning: NLTK, Scikit-learn, Transformers, Tensorflow, Optuna, Spacy, Scipy, DVC
Data Science: SQL, SPSS, R, Tableau, Numpy, Pandas
Others: Axure RP, Bash Shell, LaTeX, Git, PlantUML, Neo4j, Google Cloud Platform, Docker, Amazon Web Services