

# SIJIA GE

📍 Cincinnati, OH ◇ 📩 sijiage007@gmail.com ◇ 💻 konic-nlp.github.io ◇ 💬 LinkedIn

## EXPERIENCE

**ML/AI Engineer** 01/2024 - 02/2025  
Rippey.ai *Louisville, CO*

- **Doc-Indexing&Invoice AI:** Developed an end-to-end pipeline with SOTA NLP models for the logistics industry to classify structured documents and unstructured emails, extract specified information, and generate quotes.
- **MLOps:** Developed a seamless workflow that integrates data querying from MongoDB, model development with Hugging Face and PyTorch, deployment as endpoints on SageMaker, and real-time monitoring for efficient, scalable operations.
- **Model Training:** Leveraged large models through Ollama, Huggingface and LangChain, LayoutLM, and Transformers, achieving 87% F1 in classification with 17 types and 85% in information extraction.

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

- **Event Similarity**[code]: Contributed to a **DARPA**-funded project on graph-based event similarity computation. Leveraged integer linear programming and **Wikidata** to map event arguments and calculate similarity metrics.
- **UMR-Writer**[code][paper]: Redesigned and updated an annotation tool with improved UI/UX to support multilingual semantic role labeling. Enhanced user interaction by replacing click-based inputs with keyboard shortcuts and implementing multiple event listeners for streamlined usability. Added a project management module, resulting in a **60%** increase in annotation efficiency.
- **Ancient Chinese NER**[code][paper]: Developed and optimized state-of-the-art models for Classical Chinese processing, focusing on sequence labeling tasks such as Named Entity Recognition, and sentence segmentation. Led model development and validation across historical corpora, achieving **87%** NER F-1 and **92%** sentence segmentation F-1 through cross-era corpus and label integration.
- **Machine Learning Course Manager:** Facilitated assignments and student support for a **60-student** class. Developed scripts for assignments, automated grading, and feedback processes. Provided personalized support via office hours and online platforms, resolving student inquiries efficiently and maintaining a high level of engagement.

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

- **ShuYuBao Product Design & Management:** Designed and iterated a terminology extraction and management product, addressing user pain points and analyzing competitors. Led product prototyping and feature design, introducing a differentiated user monetization model. Achieved a **60%** user retention rate in the first week and exceeded revenue expectations by **120%**.

## EDUCATION

**Ph.D. in Computer Science& Engineering, University of Cincinnati** 08/2025 - present  
**MS in Computational Linguistics, University of Colorado-Boulder |GPA: 3.972/4.0** 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

## SKILLS

**Programming Languages & Web Development:** Python, Java, JavaScript, React, Django, Flask, HTML, Heroku, FastAPI, CSS, Jquery

**Machine Learning:** Pytorch, Transformers, Tensorflow, NLTK, Scikit-learn, Optuna, Spacy, Scipy, DVC, MLflow

**Large Language Models:** LangChain, Ollama, LlamaIndex

**Data Science:** SQL, SPSS, R, Tableau, Numpy, Pandas, Polaris

**Others:** Axure RP, Bash Shell, LaTex, Git, PlantUML, Neo4j, Google Cloud Platform, Docker, Amazon Web Services