Yunyue Su

+86-176-003-13217 | yunyue.su@ia.ac.cn

OPilgrim | Google Scholar |

Haidian, Beijing - 100000, China

Objective

My research focuses on deep learning, natural language processing, and agent augmentation, particularly in event extraction, fact-checking, graph learning, tool and retrieval-augmented generation. Recently, I have extended my interest to multi-modal foundation models in AI4Science. During my PhD, I aim to advance the reasoning and continual learning capabilities of MLLMs and explore their applications in workflow planning and embodied intelligence.

Education

Computer Network Information Center, University of Chinese Academy of Sciences *Sept. 2020 - Jun. 2023*
M.S. Thesis, Computer Technology; Advised by Associate Professor Yuanping Chen and Xiaoqian Zhu Beijiing, China
GPA: 3.89/4.00 IELTS: 6.5 (2024.11)

College of Mathematics and Computer Science, Fuzhou University *Sept. 2016 - Jun. 2020*
B.S., Computer Science and Technology; Advised by Professors Xiangwen Liao and Qingyuan Bai Fujian, China
GPA: 3.62/5.0 (Equally 3.51/4.0 according to Peking University)

Publications and Patents

C=Conference, J=Journal, P=Patent, S=In Submission, T=Thesis

[J.1] Shu Wu, Zekun Li, Yunyue Su, Zeyu Cui, Xiaoyu Zhang, Liang Wang. (2025). **GraphFM: Graph Factorization Machines for Feature Interaction Modelling.** Manuscript submitted for publication in *Machine Intelligence Research*, DOI: 10.1007/s11633-024-1505-5.

[C.1] Ge Shi*, Yunyue Su*, Yongliang Ma, and Ming Zhou. (2023). **A Hybrid Detection and Generation Framework with Separate Encoders for Event Extraction.** In *the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 3163-3180. Association for Computational Linguistics. 2023, Dubrovnik, Croatia. DOI: 10.18653/v1/2023.eacl-main.231.

[P.1] Yongliang Ma, Yunyue Su, Ge Shi, and Ming Zhou. (2022). **Event extraction method, system and computer readable storage medium**. Lanzhou Technology Co ltd, Patent No. CN202210160088.8A. Registration Date: 2022-02-22, Grant Date: 2022-06-21, Publication Date: 2022-06-21.

[S.1] Yunyue Su, Qiang Liu, Liang Wang. (2025). **LToGc: Efficient Large-Scale Tokenization for Generative Tool Calling**. *Open reviewing*.

* Equal Contribution.

Experience

NLPR, Institute of Automation, CAS 2023 - Now
Advised by Professor Qiang Liu, Shu Wu, and Liang Wang. Beijing, China
Lead researcher on national key projects; core team contributor and co-author of academic publications.

Research Engineer

- Co-authored the paper *SGTC* with Prof. Qiang Liu. This paper research efficient methods for enabling LLMs to invoke large-scale toolkits (over 47,000). Developed a structure-aware semantic tokenization scheme and a post-guided iterative training strategy, unifying tool representation, retrieval, and calling within a single generative process. [Code released soon]
- Co-authored the paper *GraphFM* with Prof. Shu Wu. This work pioneered the integration of graph structures into Factorization Machines (FM) frameworks, modeling feature interactions from a topological perspective. By leveraging GNN-based aggregation strategies, it explicitly captures high-order interactions while automatically filtering noise, addressing inherent limitations of traditional FMs. [Code]

Intermediate Engineer

- **Cross-Modal Scientific Data Framework**: Developed a multimodal reasoning framework integrating spectral and mass spectrometry data to infer molecular structures (e.g., functional groups, SMILES) and predict properties under experimental conditions, enabling end-to-end spectrum–structure–property mapping for applications in drug discovery and materials science.
- **Algorithm Development Leadership**: Led R&D for industry-academia projects, including a national R&D sub-project on graph neural network-based industrial risk prediction (HuaXia Bank, 30 million CNY), Ant Group's retrieval-augmented controllable generation project (350K CNY), and Tencent's knowledge-driven LLM editing system (200K CNY).
- **Multimodal Fact-Verification Platform**: Built a RAG-based platform supporting cross-modal evidence retrieval (audio, video, text, images) for offline/online fact-checking, deployed for clients including the Chinese Academy of Sciences and Cyberspace Administration of China.

Computer Network Information Center, CAS 2020 - 2023
Graduate Student Researcher, Advised by Associate Professor Yuanping Chen and Xiaoqian Zhu.
- Designed a *document-level generative event extraction framework for financial risk management*, integrating domain knowledge-enhanced multi-task training (recursive event path expansion, multi-dimensional knowledge matrix fusion) to address cross-sentence argument extraction and multi-event argument sharing in low-resource long-text scenarios, significantly improving event detection and argument extraction accuracy. [Report]

Baidu, Inc. 2022 - 2023
Research Intern, Mentored by Tianfu Zhang.
- Developed a *probe-based* method to automatically generate syntactic dependency trees by extracting structured features from language models, enabling high-quality annotation for new texts and enhancing long-distance argument association modeling and extraction accuracy in event processing.

Langboat Technology 2021 - 2022
Research Intern, Mentored by Yongliang Ma, Assistant Professor Ge Shi, and Professor Ming Zhou.
- Developed a *generative event extraction framework* for financial data, integrating constrained decoding and adaptive template mechanisms (soft prompt/dynamic prefix) to resolve error-prone template generation and few-shot adaptation challenges; optimized event pre-screening on banking data, achieving above 10% accuracy improvement in argument extraction; deployed in production with a patent application (CN*) filed.
- Designed a separate-encoder framework with joint bridging training to mitigate feature interference between event detection and argument extraction, research published as a long paper in EACL 2023. [Code]

Ruijie Networks 2019 - 2020
Audio processing intern, Mentored by Honghe Wu
- Responsible for collecting and labeling audio data for training speech separation models; re-engineered the popular audio processing framework Kaldi using Python to achieve operator acceleration.
- Implemented an OCR recognition algorithm for PDF documents, using multimodal similarity calculation algorithms to match recognized document blocks and enable document plagiarism detection, deployed online via Docker.

Supercomputing Center, Department of Computer Science, FuZhou University 2016 - 2020
Undergraduate Student Researcher, Advised by Professor Xiangwen Liao and Qingyuan Bai.
- Designed and implemented an *end-to-end music recommendation* algorithm by leveraging temporal convolutional networks to extract deep features from audio Mel-spectrograms, integrating user behavior sparse features and audio semantic features via DeepFM model for user interest prediction, and generating personalized recommendations based on ranking.
- Under the supervision of Prof. Dong Zhang, collaborated with a team of 9 peers to design and develop Jarvis for Chat, an intelligent conversational assistant powered by large language models, as part of a software engineering capstone project. [Blogs, Github]

Graduate Coursework

UCAS: The Design and Analysis of Computer Algorithm, Advanced Artificial Intelligence, Human-Computer Interaction, Deep learning for natural language processing, Big Data Analysis, Knowledge Graph, Advanced Software Engineering, Advanced Data Management, Internet Finance, The Introduction and Frontiers of Social Computing.

FZU: Artificial Intelligence, Pattern Recognition, Software Engineering, Algorithms and Data Structures, Computer Operating Systems, Principles of Computer Organization, Computer Networks, Compiler Design, Principles of Database Systems, Higher Mathematics, Linear Algebra, Discrete Mathematics, University Physics, Digital Circuits.

Awards and Honors

Government-funded Full Scholarships *2020 - 2023*
University of Chinese Academy of Sciences

Outstanding Undergraduate Thesis *2020*
Fuzhou University

(5x) Merit-based Scholarship Scholarships *2016 - 2020*
Fuzhou University

Outstanding Student Leader *2017 - 2019*
Fuzhou University

Extracurricular Activities

Pattern recognition Lab Health Club *2023–Pres.*
Volunteer | Beijing, China
- Responsible for organizing badminton activities twice a week and group swimming once a week in the laboratory.

Personnel education division *2022*
Defense Secretary | Beijing, China
- Coordinated full-cycle academic defenses for 7 students and schedules for 17 faculty evaluators. Led official defense proceedings.
- Ensured comprehensive archival submission through real-time documentation and verification of all defense materials.

Computer Science Student Council *2016–2019*
Director | Fuzhou, China

- Managed council's physical assets including office facilities and equipment inventory, while operating weekly consultation desks providing academic/administrative support to 300+ peers.
- Spearheaded logistics for major campus events including Freshman Orientation Week (600+ attendees) , School Sports Meet, etc.
- Initiated peer-led study groups weekly and improved participants' GPA across 10+ core courses.

Fuzhou University Library Association *2016–2019*

Volunteer | Fuzhou, China

- Managed cataloging systems and patron services (book loans, research assistance), maintaining study-ready environments through noise regulation and facility upkeep.
- Conducted STEM-focused tours for K-12 groups at Technology Exhibition Hall and led university history presentations for official visitors.

INTERESTS

Cycling (can go beyond 40km — if only time allowed!), swimming, badminton, HIIT, strength training (can do 20+ proper push-ups!), and karaoke.