

Xuemei Tang

NATURAL LANGUAGE PROCESSING · DIGITAL HUMANITIES

☎ (+852) 56179108 | ✉ xuemei.tang@polyu.edu.hk | 🏠 tangxuemei1995.github.io/

Education

The Hong Kong Polytechnic University

POSTDOCTORAL FELLOW

- Research Interests: Natural Language Processing, Digital Humanities

Hong Kong, China

Nov. 2024 - Present

Peking University

PHD INFORMATION SCIENCE

- Research Interests: Natural Language Processing, Digital Humanities

Beijing, China

Sept. 2020 - Jun. 2024

Beijing Normal University

MA LINGUISTICS AND APPLIED LINGUISTICS

- Research Interests: Natural Language Processing, Digital Humanities

Beijing, China

Sept. 2017 - Jun. 2020

Tianjin Normal University

BA INFORMATION ENGINEERING & CHINESE LANGUAGE AND LITERATURE

- Main course: C Programming; Computer Network; Computer Graphics; Signals and Systems
- Minor course: Classical Chinese; Introduction to Literature; Classical Chinese Literature; Modern Chinese Literature; History of Foreign Literature.

Tianjin, China

Sept. 2012 - Jun. 2016

Publications

CONFERENCE PUBLICATIONS

- **Xuemei Tang**, Xufeng Duan, Zhenguang G. Cai. 2025. *Large Language Models for Automated Literature Review: An Evaluation of Reference Generation, Abstract Writing, and Review Composition*. In EMNLP 2025 Main.
- **Xuemei Tang**, Jun Wang, Qi Su, Chu-Ren Huang, and Jinghang Gu. 2025. *An Effective Incorporating Heterogeneous Knowledge Curriculum Learning for Sequence Labeling*. In ACL 2025 Main.
- **Xuemei Tang**, Zekun Deng, Qi Su, Jun Wang. 2024. *CHisIEC: An Information Extraction Corpus for Ancient Chinese History*. In LREC-COLING 2024.
- **Xuemei Tang**, Qi Su. 2022. *That Slepen Al the Nyght with Open Ye! Cross-era Sequence Segmentation with Switch-memory*. In ACL 2022 Main.

JOURNAL PUBLICATIONS

- **Xuemei Tang**, Linxu Wang, Jun Wang. 2025. *Language model collaboration for relation extraction from classical Chinese historical documents*. Information Processing & Management 63 (1).
- **Xuemei Tang**, Zekun Deng, Jun Wang, Qi Su. 2025. *ChisNERE: A premodern Chinese corpus with named entity and relation annotation*. Digital Scholarship in the Humanities, 40(2), 617-638.
- **Xuemei Tang**, Qi Su, Jun Wang. 2024. *Chinese word segmentation with heterogeneous graph convolutional network*. Natural Language Processing, 31(5), 1285 - 1307.

PREPRINTS

- **Xuemei Tang**, Chengxi Yan, Jinghang Gu, Chu-Ren Huang. 2025. *Joint Information Extraction Across Classical and Modern Chinese with Tea-MOELoRA*. <https://arxiv.org/abs/2509.01158>
- Xufeng Duan, Bei Xiao, **Xuemei Tang**, Zhenguang G. Cai. 2024. *HLB: Benchmarking LLMs' Humanlikeness in Language Use*. <https://arxiv.org/abs/2409.15890>

Participated Projects

Evaluating the Syntax-Semantics Knowledge in Large Language Models

The Hong Kong Polytechnic University, China

Nov. 2024 - PRESENT

- In this project, I evaluated differences between LLMs and humans in language use and academic writing, analyzing how models replicate human-like writing patterns and identifying areas where they diverge.

The Construction of the Knowledge Graph for the History of Chinese Confucianism

Natural Science Foundation of China (NSFC), China

Jan. 2020 - July 2024

- In this project, I was responsible for collecting and cleaning classical Chinese texts and developing models for automatic segmentation, punctuation, and relation extraction on these documents.

Personal Statement

- My research primarily focuses on information extraction, LLM evaluation, and sequence labeling.

IN INFORMATION EXTRACTION

- I study how to extract structured knowledge from Classical Chinese historical texts.
- **CHisIEC: An Information Extraction Corpus for Ancient Chinese History.** Constructed a high-quality dataset for NER and RE in ancient Chinese to advance Digital Humanities research.
- **Language model collaboration for relation extraction from classical Chinese historical documents.** Proposed SLCoLM, a “Training-Guide-Predict” framework where a small, task-specific PLM guides an LLM to improve RE on low-resource data.
- **Joint Information Extraction Across Classical and Modern Chinese with Tea-MoELORA.** Developed a parameter-efficient multi-task MoE-LoRA framework with task-era-aware routing to balance task- and era-specific knowledge.

IN LLM EVALUATION

- I explore differences between LLMs and human language behavior.
- **Large Language Models for Automated Literature Review.** Evaluated LLMs’ abilities in reference generation, abstract writing, and review composition via a multidimensional hallucination and factuality framework.
- **HLB: Benchmarking LLMs’ Humanlikeness in Language Use.** Built a benchmark with 10 psycholinguistic experiments (phonology–discourse) comparing 20 LLMs to 2,000 human responses, revealing that higher accuracy does not always mean greater humanlikeness.

IN SEQUENCE LABELING

- I focus on cross-era and knowledge-enhanced Chinese modeling.
- **That Slepem Al the Nyght with Open Ye! Cross-era Sequence Segmentation with Switch-memory.** Proposed CROSSWISE, introducing era-specific knowledge via a Switch-memory module for diachronic Chinese segmentation.
- **Incorporating Heterogeneous Knowledge via Curriculum Learning for Sequence Labeling.** Designed a two-stage curriculum learning framework to progressively integrate heterogeneous external knowledge.
- **Chinese Word Segmentation with Heterogeneous Graph Convolutional Network.** Leveraged syntactic and semantic structures via GCNs to enhance Chinese sequence labeling.

ONGOING RESEARCH

- **Event Ontology Construction for Classical Chinese Historical Texts.** This project leverages LLMs to construct an event ontology from classical Chinese historical documents, aiming to systematically organize and extract structured knowledge for digital humanities research.
- **Text Restoration and Information Extraction from Chinese Epitaphs.** This research focuses on restoring damaged tombstone inscriptions and extracting structured information using LLMs and VLMs, supporting cultural heritage preservation and historical analysis.

Honors & Awards

2023	Best Technology Award (Project Team) , Chinese Digital Humanities Annual Conference	<i>Wuhan, China</i>
2022	Merit Student , Peking University	<i>Beijing, China</i>
2022	Leo KoGuan Scholarship , Peking University	<i>Beijing, China</i>
2019	First-Class Scholarship , Beijing Normal University	<i>Beijing, China</i>

Skills

Deep Learning Framework	Pytorch, Tensorflow
NLP Technique	Graph Neural Networks, Pre-trained Language Model, Large Language Model, Curriculum Learning
NLP Task	Sequence Labeling, Text Classification, Text Generation, Information Extraction
Others	Gephi, Ontology, Semantic Network, Neo4j, Protege, Topic Model

Other Experiences

- 2025 **Area Chair**, ACL ARR Oct. 2025
- 2025 **Reviewer**, Information Processing & Management (IPM)
- 2025 **Reviewer**, Natural Language Processing
- 2023-2025 **Reviewer**, ACL, EMNLP, EMNLP Industry Track
- 2023 **Reviewer**, Journal of Neural Computing and Applications (NCAA)