XIAONAN LI

2005 Songhu Rd., Shanghai, China 200438

(+86) 181 5803 7912 ⋄ lixiaonan1208@gmail.com

PROFILE

- 4th-year **Ph.D.** candidate at Fudan University, advised by Xipeng Qiu. One-year research **intern at Microsoft**. My research focuses on **NLP & LLM**, especially on **Retrieval-Augmented Generation**. 8 first-author papers were accepted by **ACL, EMNLP & NAACL**. Citation ∼ **1300**.

- Main research includes using various external-information to augment **LLM (Factuality, Reasoning Abilities, Specialized Abilities, etc)**. I also have experiences on **Dense Retriver Building**.

- Developing skills for independently **leading an NLP research project** and **mentoring juniors**.

EDUCATION

Fudan University   *2020.9 - 2025.6* (Expected)
Ph.D. Candidate, Natual Language Processing   Shanghai, China
Advisor: Prof. Xipeng Qiu   School of Computer Science

Xidian University   *2016 - 2020*
Bachelor of Engineering, Software Engineering   Xi'an, China
GPA: 3.9/4.0 (rank 3/400)   School of Computer Science and Technology

SKILLS

Technical Skills

- Retrieval Augmented Generation (RAG)

- Information Retrieval (Document, Code, Task Demonstration)

- LLM Fine-Tuning

- Verifiable LLM Generation (with Citation)

- In-Context Learning

General Skills

- Problem Abstraction and Definition

- Experiment Design

- Efficient Method Verification & Iteration

- Independent Paper Writing

- Mentorship for Juniors

HONORS & AWARDS

National Scholarship for Undergraduate Student (three times)   *2016 - 2019*
National Scholarship for Graduate Student   *2021*

INTERNSHIP

Microsoft Research Asia, NLP Research Intern   *2020.6 - 2021.6*
Supervisor: Yeyun Gong; **Department:** Natural Language Computing   Beijing, China

Summary: Research on code pre-training for better code search, two papers accepted by EMNLP.

RESEARCH OVERVIEW

My research focuses on **NLP & LLM**, specially **LLM Augmentation with External Information**.

- **Augmenting LLM's Sepcialized Abilities with External Demonstration**

  - Retrieving helpful demonstrations for a specific test query
    Unified Demonstration Retriever for In-Context Learning [*ACL* 2023 Oral]
    Xiaonan Li*, Kai Lv*, Hang Yan, Tianyang Lin, ..., Xipeng Qiu
    Citation ∼ 50

  - Select representative task demonstrations
    Finding Supporting Examples for In-Context Learning [*EMNLP findings* 2023]
    Xiaonan Li, Xipeng Qiu
    Citation ∼ 50

- **Augmenting LLM's Factual Correctness and Verifiability with External Document**

  LLatrieval: LLM-Verified Retrieval for Verifiable Generation [*NAACL* 2024]
  Xiaonan Li*, Changtai Zhu*, Linyang Li, Zhangyue Yin, Tianxiang Sun, Xipeng Qiu

  Unified Active Retrieval for Retrieval Augmented Generation [*Arxiv* 2024.6]
  Qinyuan Cheng*, Xiaonan Li*, ..., Tianxiang Sun, Hang Yan, Xipeng Qiu

- **Augmenting LLM with Low-Cost External-Information**

  - Augmenting LLM with Unlabeled Data
    MoT: Memory-of-Thought Enables ChatGPT to Self-Improve [*EMNLP* 2023]
    Xiaonan Li, Xipeng Qiu

  - Augmenting PLM with Lexicon Information
    FLAT: Chinese NER Using Flat-Lattice Transformer [*ACL* 2020]
    Xiaonan Li, Hang Yan, Xipeng Qiu, Xuanjing Huang
    Citations ∼ 400

- **Dense Retriever Building**

  - Unsupervised Query-Positive Building
    CodeRetriever: Large-scale Contrastive Pre-training for Code Search [*EMNLP* 2022]
    Xiaonan Li, Yeyun Gong, Yelong Shen, Xipeng Qiu, ..., Daxin Jiang, Weizhu Chen, Nan Duan

  - False-Negative Filtering
    Soft-Labeled Contrastive Pretraining for Code Representation [*EMNLP findings* 2022]
    Xiaonan Li*, Daya Guo*, Yeyun Gong, ..., Xipeng Qiu, Daxin Jiang, Weizhu Chen, Nan Duan

- **The Necessity of External-Information Augmentation**

  Scaling Laws for Fact Memorization of Large Language Models [*Arxiv* 2024.6]
  Xingyu Lu*, Xiaonan Li*, Qinyuan Cheng, Kai Ding, Xuanjing Huang, Xipeng Qiu

- Other First-Author Papers

  Accelerating BERT Inference for Sequence Labeling via Early-Exit [*ACL* 2021 Oral]
  Xiaonan Li, Yunfan Shao, Tianxiang Sun, Hang Yan, Xipeng Qiu, Xuanjing Huang

- Co-Author Papers

  A unified Token-Pair Classification Architecture for Information Extraction [*ACL* 2022]
  Hang Yan, Yu Sun, Xiaonan Li, Yunhua Zhou, Xuan-Jing Huang, Xipeng Qiu

  An Embarrassingly Easy But Strong Baseline for Nested NER [*ACL* 2023]
  Hang Yan, Yu Sun, Xiaonan Li, Xipeng Qiu

Backdoor Attacks on Pre-trained Models by Layerwise Weight Poisoning [*EMNLP* 2022]
Linyang Li, Demin Song, Xiaonan Li, Jiehang Zeng, Ruotian Ma, Xipeng Qiu

BERT for Monolingual and Cross-Lingual Reverse Dictionary [*EMNLP findings* 2021]
Hang Yan, Xiaonan Li, Xipeng Qiu

Learning Sparse Sharing Architectures for Multiple Tasks [*AAAI* 2021]
Tianxiang Sun, Yunfan Shao, Xiaonan Li, Pengfei Liu, Hang Yan, Xipeng Qiu, Xuanjing Huang

TENER: Adapting Transformer Encoder for Name Entity Recognition [*Arxiv*]
Hang Yan, Bocao Deng, Xiaonan Li, Xipeng Qiu

Does It Make Sense? And Why? A Pilot Study for Sense Making and Explanation
Cunxiang Wang, Shuailong Liang, Yue Zhang, Xiaonan Li, Tian Gao [ACL 2019]