

# Yuyang Bai

Address: Room 206, Pengkang Building, No. 28, West Xianning Road, Xi'an, China

Email: yuyangbai2002@gmail.com, 1206944633@stu.xjtu.edu.cn, ybai3@nd.edu

Homepage: leopoldwhite.github.io

---

## Research Interests

My research interests include interpreting and enhancing the knowledge & reasoning ability of Large Language Models, and NLP & social network analysis for fairness and common good.

## Education

**Xi'an Jiaotong University**, Xi'an, Shaanxi, China (member of the C9 league in China) 2021.09 - present

Artificial Intelligence Experimental Class

GPA: 91.88/100 [transcript]

Thesis Advisor: Prof. Minnan Luo

**University of Notre Dame**, Notre Dame, IN, United States

2023.08 - 2023.12

Non-degree Undergraduate (Exchange Student)

GPA: 3.92/4.00 [transcript]

Thesis Advisor: Prof. Meng Jiang

## Publications (\* indicates equal contribution)

1. **KGQuiz: Evaluating the Generalization of Encoded Knowledge in Large Language Models**  
Yuyang Bai\*, Shangbin Feng\*, Vidhisha Balachandran, Zhaoxuan Tan, Shiqi Lou, Tianxing He, Yulia Tsvetkov  
*arXiv preprint 2310.09725, 2023*
2. **Knowledge Card: Filling LLMs' Knowledge Gaps with Plug-in Specialized Language Models**  
Shangbin Feng, Weijia Shi, Yuyang Bai, Vidhisha Balachandran, Tianxing He, Yulia Tsvetkov.  
*arXiv preprint 2305.09955, 2023*
3. **FactKB: Generalizable Factuality Evaluation using Language Models Enhanced with Factual Knowledge**  
Shangbin Feng, Vidhisha Balachandran, Yuyang Bai, Yulia Tsvetkov.  
*In Proceedings of EMNLP, 2023*
4. **Detecting Spoilers in Movie Reviews with External Movie Knowledge and User Networks**  
Heng Wang, Wenqian Zhang, Yuyang Bai, Zhaoxuan Tan, Shangbin Feng, Qinghua Zheng, Minnan Luo.  
*In Proceedings of EMNLP, 2023*
5. **Twibot-22: Towards Graph-Based Twitter Bot Detection.**  
Shangbin Feng, ... Yuyang Bai, ... Minnan Luo.  
*In Proceedings of the NeurIPS, Datasets and Benchmarks Track, 2022.*

## Research Experience

**Research Assistant**, Luo lab Undergraduate Division (LUD) @ Xi'an Jiaotong University 2022.02 - present

- Worked on addressing the challenges in Twitter bot detection, by introducing a new, comprehensive graph-based benchmark.
  - Proposed TwiBot-22, the most extensive bot detection benchmark to date, with diverse entities, relations, and high-quality annotations, and re-implemented 35 existing bot detection methods for robust evaluation.

- Worked on improving spoiler detection in online movie reviews by integrating external movie knowledge and user behavior.
  - Developed MVSD, which leverages three interconnected heterogeneous information networks and employs a heterogeneous graph neural network for nuanced, node-level spoiler classification.
  - MVSD outperformed existing state-of-the-art spoiler detection methods on curated datasets, confirming that incorporating external knowledge and user interaction patterns leads to more robust spoiler detection.
- Authored 1\**NeurIPS'22*, 1\**EMNLP'23*.

Advisor: Prof. Minnan Luo

**Research Assistant**, TsuetShop @ University of Washington

2022.10 - present

- Worked on evaluating the factual consistency of summaries generated by automatic summarization systems, focusing on improving the ability to detect erroneous entities and relations.
  - Introduced FactKB, leveraging language models pretrained with facts from external knowledge bases and incorporating three types of complementary factuality pretraining objectives designed to enhance the model's generalization capabilities for factuality evaluation.
- Worked on enhancing large language models with current, domain-specific knowledge without the need for frequent retraining.
  - Proposed a modular framework using specialized language models trained on domain-specific corpora to provide updated information that can be integrated with the base LLM, along with content selectors to ensure the relevance, brevity, and factuality of the output.
- Worked on Evaluating the knowledge generalization abilities of large language models across various domains and task complexities.
  - Introduced KGQuiz, a scalable benchmark based on triplet knowledge, encompassing three knowledge domains with five types of tasks of increasing complexity to evaluate the knowledge abilities of LLMs.
- Authored 1\**EMNLP'23*, 2\**under review*.

Advisor: Ph.D. student Shangbin Feng, Prof. Yulia Tsvetkov

## Services

- Student Program Committee (PC) member, CSUC22@XJTU
- Web Admin of the LUD lab

## Honors and Awards

Scholarship for Outstanding Students, Second Prize, Xi'an Jiaotong University	2022
Dean's List, Xi'an Jiaotong University	2022

## Skills

- Programming Skills: C/C++, Python, PyTorch, MATLAB,  $\LaTeX$ , Git
- Language Skills: Mandarin (native), English (TOFEL 104: R 29, L 29, S 22, W 24)