Yuhan Liu

lyh6560@stu.xjtu.edu.cn，yuhanliu6560@gmail.com
personal website
Pengkang Building 206, NO. 28 Xianning W Rd, Xi'an, China

RESEARCH INTERESTS

Her research mainly lies in **NLP for social good**, specifically in natural language generation, knowledge graphs, graph neural networks, and social network analysis

EDUCATION BACKGROUND

Xi'an Jiaotong University (XJTU), Xi'an, China 09/2020-06/2024

- School of Automation Science and Engineering, Honors Engineering Program
  Degree: Bachelor of Engineering
- **Major:** Automation Science and Informatics **GPA:** 93.8/100 **Rank:** 1/33
- Related courses: Linear Algebra and Geometry, Mathematical Analysis for Engineering ,Program Design Method and Practice, Data Structures and Algorithms, Modern Approaches of Artificial Intelligence, Machine Learning, Operations Research

The Special Class for the Gifted Young / Honors Youth Program of XJTU 09/2018-Present

PUBLICATIONS

- Shangbin Feng, Chan Young Park, **Yuhan Liu**, and Yulia Tsvetkov. **From Pretraining Data to Language Models to Downstream Tasks: Tracking the Trails of Political Biases Leading to Unfair NLP Models,** in *Proceedings of ACL, 2023.* **Best Paper**
- **Yuhan Liu**, Zhaoxuan Tan, Heng Wang, Shangbin Feng, Qinghua Zheng, Minnan Luo. **BotMoE: Twitter Bot Detection with Community-Aware Mixtures of Modal-Specific Experts,** in *Proceedings of SIGIR, 2023*.
- Shangbin Feng*, ... , **Yuhan Liu**, ... , Minnan Luo. **TwiBot-22: Towards Graph-Based Twitter Bot Detection.,** in *Proceedings of the NeurIPS Datasets and Benchmarks Track, 2022.*

RESEARCH EXPERIENCES

Research Assistant, **TsvetShop** 11/2022-present

Supervised by Professor Yulia Tsvetkov from University of Washington.

- Explored and assessed the political biases in large language models and their unfair impact on downstream tasks under the guidance of Shangbin Feng(University of Washington)
  - Analyzed and present the inherent biases of models from BERT-series ,GPT-series to state of art LLMs like Codex and LLaMa.
  - Examined how biases change over time (pre-Trump and post-Trump) and with variations in pretraining epochs and corpus size.
  - Illustrated the sensitivity of LLMs to groups with opposing biases and their protective tendencies towards those with similar biases.

- Co-authored the paper *"From Pretraining Data to Language Models to Downstream Tasks: Tracking the Trails of Political Biases Leading to Unfair NLP Models,"* which received the **Best Paper Award at ACL 2023** (as the third author).
- Developed methods to mitigate bias shifts in natural language generation.
  - Demonstrated in text summarization tasks that LLMs can alter the political bias of the original context during generation.
  - Implemented control in the decoding process of diffusion models to ensure continuations align with the original biases using classifier-guided diffusion models.
  - Jointly controlled political leaning and factuality of continuations through gradient-descent-based methods.

Research Assistant, **XLang Lab** 7/2023-present

Supervised by Professor Tao Yu from The University of Hong Kong.

- Collaborated with Weijia Shi (PhD at the University of Washington) to develop personalized language models.
  - Employed reinforcement learning with large language models (e.g., T5) using our synthetic datasets.
  - Introduced persona-info sensitivity as an evaluation metric for personalizing large language models.

Director, Research Assistant, **XJTU LUD Lab** 02/2022-present

Supervised by Professor Minnan Luo from Xi'an Jiaotong University

- Served as the director of the LUD lab and a reviewer for the NeurIPS Datasets and Benchmarks Track in 2022-2023.
  - Focused on graph-based Twitter bot detection, utilizing a mixture-of-experts approach to detect social bots within specific communities.
  - Developed models to jointly reason across graph, text, and metadata modalities of users.
  - Conducted extensive experiments to evaluate adaptability to different communities, robustness against manipulated features, and generalization to unseen accounts.
  - Published a paper as the first author: *"BotMoE: Twitter Bot Detection with Community-Aware Mixtures of Modal-Specific Experts*" at SIGIR 2023.

Vice President, **XJTU VTOL club** 06/2022-06/2023

Supervised by Tonghui Wu from Xi'an Jiaotong University

- Led a team focusing in 3D object detection and tracking
  - Implemented yolov3,yolov4 and Karman filter algorithms.
  - Deployed the model on edge-computing device Jetson Nano.

Research Assistant, **Xi'an Jiaotong University** 10/2021-02/2022

Supervised by Shaoyi Du from Xi'an Jiaotong University

- Contributed to the semantic segmentation part of a multi-modal project on medical image processing.
- Implemented U-Net, DeepLabv3, and DeepLabv3++ for the project.

Researcher, **National Undergraduate Training Programs** 05/2022-05/2023

Supervised by Professor Minnan Luo from Xi'an Jiaotong University

- Conducted research on knowledge graph representation learning based on pretrained language models
  - Implemented KG T5 for link prediction task, and designed special attention method in transformers based on the model to capture the multi-hop graph structure of knowledge graphs.

Researcher, **National Undergraduate Training Programs** 03/2021-07/2021

Supervised by Yijun Yang from Xi'an Jiaotong University

- Conducted research on fast boolean operations of triangular network models based on GPU.

HONOURS & AWARDS

- National Scholarship for the Academic Year of 2020-2021 and 2021-2022 09/2021,09/2022
- Dean's List of XJTU of the Academic Year of 2019-2020 and 2020-2021 09/2020,09/2021
- Outstanding Award of National Engilish Competition for College Students 05/2020
- First Prize of the Chinese Mathematics Competitions 09/2019

SKILLS

- **Programming Languages:** Proficient in Python, MATLAB, C/C++
- **Technical Tools:** PyTorch, LATEX, Visual Studio, MobaXTerm, Git, ssh, Vim