Hanzhi Zhang

hanzhizhang@my.unt.edu (940) 208-6639 Denton, TX, US

RESEARCH INTERESTS

Responsible AI, Hallucination in LLMs, Transformers, Natural Language Processing, Deep Learning

EDUCATION

University of North Texas (UNT) Aug. 2023 - present
PhD Student, Computer Science
Advisors: Dr. Yunhe Feng

University of Birmingham (UoB) Sept. 2021 - Dec. 2022
MS, Data Science with Distinction

Xiamen University (XMU) Sept. 2017 - Aug. 2021
BS, Computer Science and Technology (Honours)

PROFESSIONAL EXPERIENCE

Responsible AI Lab Denton, TX
Graduate Research Assistant, Responsible AI Aug. 2023 - present

OPPO, Inc. Shenzhen, China
Storage Backend Intern, File System Aug. 2020 - Apr. 2021

PUBLICATIONS

- **Hanzhi Zhang**, Heng Fan, Weijian Zheng, Yan Huang, and Yunhe Feng. Investigating Biased Cross-Lingual Hallucination Detection Capabilities of LLMs: POLY-FEVER Benchmark and Mitigating Approaches. In *The 62nd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2024 [Under review].

PROFESSIONAL ACTIVITIES

Conference & Journal External Reviewer

- External referee for the WWW 2024, AAAI 2024, WACV 2024, Inscrypt 2023

Conference & Journal Reviewer

- Reviewer for the IEEE ICPADS 2023

ACADEMIC PROJECTS

Multilingual LLMs Hallucination Detection and Mitigation Nov. 2023 - Feb. 2024

- Introduced **Poly-FEVER**, a vast **multilingual dataset** with **800,000 fact claims** in **11 languages**, tailored for hallucination detection tasks
- Analyzed hallucination detection in advanced language models (ChatGPT, LLaMA 2 series) using Poly-FEVER, employing **language-wise and classification prompts**
- Investigated multilingual hallucination causes using **LDA for topic analysis** and **automated web searches** to assess resource imbalances
- Proposed **a mitigation plan** leveraging **LDA** and **RAG** strategy to address linguistic discrepancies and resource imbalances for enhanced information verification

Car Accident Prediction based on Edge Computing Jun. 2022 - Sept. 2022

- Simulated **a Vehicular Ad hoc Network (VANET)** environment by utilizing NS3 and SUMO
- Compared real-time traffic incident prediction with and without VANET using **decision tree** and **LSTM** algorithms to highlight VANET's time efficiency benefits

A Small Medical Imaging Data Vault Sept. 2021 - Dec. 2021

- Developed a Django-based medical imaging **data vault**, managing over 30,000 data points across multiple studies, with emphasis on efficient **storage** and **metadata** refinement
- Implemented a user-friendly **Django browser GUI** for data addition, deletion, analysis, and secure user authentication

OPEN SOURCE CONTRIBUTIONS

- [**CubeFS**]: Distributed File System

SKILLS & PROFICIENCY

- **Programming Languages:** Python, C, C++, Golang, Shell, SQL, Rust
- **Software & Tools:** Linux, Docker, Git, Cmake, Makefile
- **Frameworks & Libraries:** PyTorch, TensorFlow, OpenCV, Django
- **Databases:** MySQL, PostgreSQL