Yunze Xiao

2080-O,Carnegie Mellon University Qatar, Education City, Doha, Qatar

☎ +974 39968706 | ✉ lyxiao@cmu.edu | ⊙ algoroxyolo | in yunzexiao | 🐦 @lrzneedresearch

Education

Carnegie Mellon University

BS COMPUTER SCIENCE   *July 2022 -May 2025*

- Minors in Philosophical Computation
- GPA 3.4/4.0

Research Experience

Carnegie Mellon University Qatar   *Doha, Qatar*

ADVISOR: PROF. HOUDA BOUAMOR   *May. 2023 - Now*

- Conducted an in-depth survey on offensive language detection in Chinese, analyzing current benchmarks, approaches, and tools.
- Identified and evaluated specific models and tools for offensive speech detection in Chinese, proposing potential research avenues to address the linguistic and cultural nuances of the language.
- Developed a novel task for labeling subversive Chinese phrases and curated a dataset of 6000+ entries, enhancing offensive language detection and promoting a respectful online environment.

Qatar Computing Research Institute   *Doha, Qatar*

ADVISOR: DR. FIROJ ALAM   *May. 2023 - Oct. 2023*

- Actively participated in the ArAIEval 2023 shared task, focusing on the advancement of Arabic language models for detecting propaganda and disinformation, directly addressing societal issues related to misinformation.
- Fine-tuned state-of-the-art BERT-based transformer models and applied zero- and few-shot learning techniques with GPT-4, showcasing the ineffefctiveness of Large Language Model in detecting disinformation.
- Conducted extensive experimentation with transformer models, resulting in improved performance metrics, underscoring the ability to effectively handle imbalanced datasets and optimize model parameters.

Carnegie Mellon University - Language Technology Institute   *Pittsburgh,PA*

ADVISOR: PROF.DAVID MORTENSEN   *Sep. 2023 - Oct.2023*

- Conducted a study on lexical-syntactic flexibility, assessing the capacity of five prominent language models (including GPT-3.5, GPT-4, Mistral 7B, Falcon 40B, and Llama 2 70B) to adapt words to non-prototypical grammatical contexts, enhancing the understanding of English morphology.
- Developed and implemented a testing methodology within a natural language inference framework to systematically evaluate the ability of language models to handle conversion or zero-derivation.
- Analyzed and interpreted complex data from 3,069 prompts, which revealed GPT-4's superior performance and challenged assumptions about model size and lexical-syntactic flexibility.

Singapore University of Technology and Design   *Singapore*

ADVISOR: PROF. ROY KA-WEI LEE   *Jan. 2024 - Sep. 2024*

- Developed advanced detection techniques for offensive language in Chinese, focusing on homophone and emoji perturbations, enhancing robustness against real-world adversarial attacks.
- Implemented and evaluated various Large Language Models (LLMs) with a focus on macro-F1 metrics, contributing to a deeper understanding of model performance under linguistic perturbations.
- Analyzed the impact of homophone and emoji perturbations on offensive language detection, revealing substantial limitations in existing models' capabilities.

Carnegie Mellon University - Language Technology Institute   *Pittsburgh,PA*

ADVISOR: PROF.DAPHNE IPPOLITO   *Aug.2024 - Dec. 2024*

- Designed and implemented a robust data transformation pipeline that pre-processes data, applies stylistic transformations, and ensures the transformed data maintains semantic coherence with the original content.
- Evaluated the impact of styled datasets on LLM performance, conducting experiments to assess how these datasets influence the generation of text that adheres to specific stylistic requirements.
- Explored the use of styled datasets for addressing challenges in text-style transfer, including experiments on multi-style transfer and dynamically-tailored content generation, leveraging models fine-tuned on synthetic data.

Professional Experience

2022-2024	**Curriculum Support Developer**, School of Computer Science, Carnegie Mellon University Qatar
2023-2024	**Student Instructor**, Carnegie Mellon University Qatar
2024	**AI research Intern**, GreenDynamics
2024	**AI research Intern**, Squirrel Learning AI

Presentations

Non-archival Presentations

Yunze Xiao, Malak Ibrahim, Madhavi Ganapathiraju. Classical and Modern Language Modeling Techniques for Transmembrane Helix Prediction. Extended Abstract Poster presentation in ISMB 2024

Teaching

F2023	**15-110:Principles of Computing**, Course Assistant
S2024	**98-031:NLP Ethics in a Nutshell**, Student Instructor
S2024	**15-112:Fundamentals of Programming and Computer Science**, Course Assistant

Publications

Published

Yunze Xiao. 2022. A Transformer-based Attention Flow Model for Intelligent Question and Answering Chatbot. In *the proceeding of 14th International Conference on Computer Research and Development* (ICCRD), Dec 2022.

Yunze Xiao, Firoj Alam. 2023. Nexus at ArAIEval shared task: Persuasion techniques detection: an interdisciplinary approach to identifying manipulative strategies. In *proceeding of The First Arabic Natural Language Processing Conference* (ArabicNLP 2023), Dec 2023.

David R. Mortensen*, Valentina Izrailevitch*, **Yunze Xiao**, Hinrich Schütze and Leonie Weissweiler. Verbing Weirds Llamas: Evaluation of English Zero-Derivation in Large Language Models. In *The proceedings of 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*(LREC-COLING 2024)

Xintao Wang, **Yunze Xiao**, Quan Tu, Yaying Fei, Ziang Leng, Wei Wang, Jianjie Chen, Cheng Li, Rui Xu, Haoran Guo, and Yanghua Xiao. ***InCharacter***: Do Role-Playing Agents Accurately Capture Characters' Personalities? In *the proceedings of 62nd Annual Meeting of the Association for Computational Linguistics(ACL 2024)*

Pre-prints

Yunze Xiao, Houda Bouamor,and Wajdi Zaghouani, 2024. Chinese Offensive Language Detection: Current Status and Future Directions.

Qingyang Wu, Ying Xu, Tingsong Xiao, **Yunze Xiao**, Yitong Li, Tianyang Wang, Yichi Zhang, Shanghai Zhong, Yuwei Zhang, Wei Lu, Yifan Yang. 2024. Surveying Attitudinal Alignment Between Large Language Models Vs. Humans Towards 17 Sustainable Development Goals.

Yunze Xiao, Yujia Hu,Kenny Tsu Wei Choo,Roy Ka-wei Lee. ToxiCloakCN: Evaluating Robustness of Offensive Language Detection in Chinese with Homophonic and Emoji Perturbations

Outreach & Professional Development

Service and Outreach

2023	**Student Majilis**, Head of Academics	*Doha,Qatar*
2024-	**ACL Rolling Review**, Reviewer	*Remote*
2024	**The Web Conference 2024**, Volunteer	*Singapore*