

Ranran Haoran Zhang

W343 Westgate Building
University Park, PA, 16802
hzz5361@psu.edu
windchimeran.github.io

RESEARCH INTERESTS

My research interests lie in making better use of **available** and **unavailable** data to develop better model, e.g.,

- With millions of data, millions of annotators, is it exhaustive enough to train a strong specialized model?
- What strategies can be employed to unify diverse labeled datasets for training universal models across multiple tasks?
- When combining heterogeneous datasets, how can we maintain and verify data quality and consistency?

I'm also interested in the applications of LLMs in Computer Science Education (CSEdu), and use the CSEdu theory to develop better CodeLLMs datasets.

EDUCATION

	<i>Ph.D. Computer Science and Engineering</i> Penn State University Advisor: Dr. Rui Zhang	Jan. 2022 - Present
	<i>M.S. Information Management</i> University of Illinois Urbana-Champaign Advisor: Dr. Heng Ji	Aug. 2019 - May. 2021
	<i>B.S. Computer Science and Technology</i> Changsha University of Science & Technology Thesis: Intent Classification Based on Pattern Engine Advisor: Dr. Daojian Zeng	Sep. 2014 - Jun. 2018

EXPERIENCE

	<i>Co-OP Researcher</i> eBay, Remote Project: Advertising Keyphrase Generation <ul style="list-style-type: none">• Low-latency Seq2Seq model to generate infinite number of keyphrases from a finite annotated dataset.• Paper & Product-driven development.	Sep. 2023 - Apr. 2025
	<i>Team Lead</i> Amazon, Remote Project: Task-Oriented Multimedia Chatbot <ul style="list-style-type: none">• Team lead of 12 members for 8 months, working on NLP, Multidemia and MLOps.	Jan. 2023 - Aug. 2023
	<i>R & D Intern</i> Benten Tech, Remote Project: Chatbot for Alzheimer's Caregiver <ul style="list-style-type: none">• Information Retrieval based FAQ module.• Neural based chitchat module with long-term memory.	Aug. 2021 - Nov. 2021

- Deployed on Alexa Skill.

MANUSCRIPT

- [P2]. **Ranran Haoran Zhang**, Soumik Dey, Ashirbad Mishra, Hansi Wu, Binbin Li, Rui Zhang. “Batch Speculative Decoding Done Right”. arXiv preprint 2025.
- [P1]. **Ranran Haoran Zhang**, Aysa Xuemo Fan, Xiaoxin Lu, Wenpeng Yin, Rui Zhang. “AlignGuide: Resolving Cross-Dataset Annotation Conflicts through Guideline-Data Co-Reconciliation”. Under-review by TACL-2025

PUBLICATIONS * indicates equal contribution

- [14]. Jiangshu Du, Yibo Wang, Wenting Zhao, Zhongfen Deng, Shuaiqi Liu, Renze Lou, Henry Peng Zou, Pranav Narayanan Venkit, Nan Zhang, Mukund Sri-nath, **Ranran Haoran Zhang**, Vipul Gupta, Yinghui Li, Tao Li, Fei Wang, Qin Liu, Tianlin Liu, Pengzhi Gao, Congying Xia, Chen Xing, Jiayang Cheng, Zhaowei Wang, Ying Su, Raj Sanjay Shah, Ruohao Guo, Jing Gu, Haoran Li, Kangda Wei, Zihao Wang, Lu Cheng, Surangika Ranathunga, Meng Fang, Jie Fu, Fei Liu, Ruihong Huang, Eduardo Blanco, Yixin Cao, Rui Zhang, Philip S. Yu, Wenpeng Yin. “LLMs Assist NLP Researchers: Critique Paper (Meta-)Reviewing”. EMNLP 2024.
- [13]. Xiaoxin Lu, **Ranran Haoran Zhang**, Yusen Zhang, Rui Zhang. “Enhance Multimodal Consistency and Coherence for Text-Image Plan Generation”. ACL 2025 Findings
- [12]. Ryo Kamoi, Yusen Zhang, Sarkar Snigdha Sarathi Das, **Ranran Haoran Zhang**, Rui Zhang. “VisOnlyQA: Large Vision Language Models Still Struggle with Visual Perception of Geometric Information”. COLM 2025
- [11]. **Ranran Haoran Zhang**, Bensu Uçar, Soumik Dey, Hansi Wu, Binbin Li, Rui Zhang. “From Lazy to Prolific: Tackling Missing Labels in Open Vocabulary Extreme Classification by Positive-Unlabeled Sequence Learning”. NAACL 2025 Findings
- [10]. Yusen Zhang, Wenliang Zheng, Aashrith Madasu, Peng Shi, Ryo Kamoi, Hao Zhou, Zhuoyang Zou, Shu Zhao, Sarkar Snigdha Sarathi Das, Vipul Gupta, Xiaoxin Lu, Nan Zhang, **Ranran Haoran Zhang**, Avitej Iyer, Renze Lou, Wenpeng Yin, Rui Zhang. “HRScene: How Far Are VLMs from Effective High-Resolution Image Understanding?”. ICCV 2025.
- [9]. Ryo Kamoi, Sarkar Snigdha Sarathi Das, Renze Lou, Jihyun Janice Ahn, Yilun Zhao, Xiaoxin Lu, Nan Zhang, Yusen Zhang, **Ranran Haoran Zhang**, Sujeeth Reddy Vummanthala, Salika Dave, Shaobo Qin, Arman Cohan, Wenpeng Yin, Rui Zhang. “Evaluating LLMs at Detecting Errors in LLM Responses”. COLM 2024
- [8]. **Ranran Haoran Zhang**, Parker Sell , Yusen Zhang , Liwei Che , Allen Gao, Sathiyajith K S , Rishabh Bhatt , Piyush Nagasubramaniam , Sujeeth Vummanthala, Salika Dave , Harsh Maniar , Vishnu Dasu , Rui Zhang. “EvoquerBot: A Multimedia Chatbot Leveraging Synthetic Data for Cross-Domain Assistance”. Alexa Prize TaskBot Challenge 2 Proceedings.
- [7]. Sarkar Snigdha Sarathi Das, **Ranran Haoran Zhang**, Peng Shi, Wenpeng Yin, Rui Zhang. “Unified Low-Resource Sequence Labeling by Sample-Aware Dynamic Sparse Finetuning”. EMNLP 2023.
- [6]. Aysa Xuemo Fan, **Ranran Haoran Zhang**, Luc Paquette, Rui Zhang. “Exploring the Potential of Large Language Models in Generating Code-Tracing Questions for Introductory Programming Courses”. EMNLP 2023 Findings.

- [5]. **Ranran Haoran Zhang**, Aysa Xuemo Fan, Rui Zhang. “ConEntail: An Entailment-based Framework for Universal Zero and Few Shot Classification with Supervised Contrastive Pretraining”. EACL 2023.
- [4]. Qingyun Wang, Manling Li, Xuan Wang, Nikolaus Parulian, Guangxing Han, Jiawei Ma, Jingxuan Tu, Ying Lin, **Ranran Haoran Zhang**, Weili Liu, Aabhas Chauhan, Yingjun Guan, Bangzheng Li, Ruisong Li, Xiangchen Song, Heng Ji, Jiawei Han, Shih-Fu Chang, James Pustejovsky, David Liem, Ahmed Elsayed, Martha Palmer, Jasmine Rah, Cynthia Schneider, Boyan Onyshkevych. “COVID-19 Literature Knowledge Graph Construction and Drug Repurposing Report Generation“. NAACL’2021 Demo. Best Demo Award.
- [3]. **Ranran Haoran Zhang***, Qianying Liu*, Aysa Xuemo Fan, Heng Ji, Daojian Zeng, Fei Cheng, Daisuke Kawahara and Sadao Kurohashi. “Minimize Exposure Bias of Seq2Seq Models in Joint Entity and Relation Extraction”. EMNLP 2020-Findings.
- [2]. Daojian Zeng*, **Ranran Haoran Zhang***, Qianying Liu. “CopyMTL: Copy Mechanism for Joint Extraction of Entities and Relations with Multi-Task Learning”. AAAI 2020.
- [1]. Daojian Zeng, **Ranran Haoran Zhang**, Lingyun Xiang, Jin Wang, Guoliang Ji. “User-oriented paraphrase generation with keywords controlled network”. IEEE Access 2019

HONORS AND AWARDS	NAACL Best Demo Paper EMNLP Outstanding Reviewer	2021 2024
--------------------------	-----------------------------------------------------	--------------

SERVICES	Reviewer, ACL-ARR Reviewer, ICLR. Reviewer, NeurIPS. Reviewer, EMNLP. Reviewer, COLING. Reviewer, JMLC.	2023-2025 2024 2023 2022, 2023 2022 2021
-----------------	------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------

PROFESSIONAL REFERENCE	Rui Zhang, Ph.D. Assistant Professor, Computer Science and Engineering Department, Penn State University. Address: W329 Westgate Building, University Park, PA 16802 Email: rmz5227@psu.edu Relationship: Ph.D. Advisor
-------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------