Rui Zhang

W329
Westgate Building
University Park, PA 16802

rmz5227@psu.edu
https://ryanzhumich.github.io/
+1-734-741-3578

ACADEMIC EMPLOYMENT	**Penn State University**	July 2020 - Present
	Assistant Professor in Computer Science & Engineering	

EDUCATION	**Yale University**	Aug 2017 - May 2020
	Ph.D., Computer Science Advisor: Dragomir Radev	
	University of Michigan, Ann Arbor	Aug 2015 - May 2017
	Ph.D. student, Computer Science	
	University of Michigan, Ann Arbor	Sept 2013 - Apr 2015
	B.S.E., Computer Engineering. Summa Cum Laude	
	Shanghai Jiao Tong University	Sept 2011 - Aug 2015
	B.S.E., Electrical and Computer Engineering	

RESEARCH INTERNSHIPS	**Google AI**, Mountain View	June 2019 - Aug 2019
	Grammarly Research, New York	May 2018 - Aug 2018
	IBM Thomas J. Watson Research Center, New York	May 2017 - Aug 2017

HONORS AND AWARDS	Senior Area Chair Paper Award at NAACL 2025	2025
	NSF CAREER Award	2024
	Outstanding Area Chair at EMNLP 2024	2024
	Yale Conference Travel Fellowship	Yale University, 2017, 2018
	Rackham Conference Travel Grant	University of Michigan, March 2016
	CSE Departmental Fellowship	University of Michigan, 2015-2016
	Outstanding Undergraduate Research Award	University of Michigan, March 2015
	James B. Angell Scholar	University of Michigan, 2015
	Wang Chu Chien-Wen Research Scholarship	University of Michigan, May 2014
	Academic Excellence Scholarship (top 1%)	University of Michigan, Oct 2012, Oct 2013
	Bao Steel Excellence Scholarship	Shanghai Jiao Tong University, Oct 2012
	University Merit Student	Shanghai Jiao Tong University, Sept 2012
	Dean's List	Every Semester

GRANTS	NSF FDT-BioTech, NSF ($900,000)	2025
	Open Philanthropy ($166,078)	2025
	Raytheon ($287,264)	2025
	NSF CAREER Award, NSF ($560,000)	2024
	NSF NAIRR Pilot project, NSF ($10,000)	2024
	a2 Pilot Award, NIH and PennAITech ($150,000)	2024
	Microsoft Research Award ($50,000)	2024
	eBay Research Award ($100,000)	2024
	eBay Research Award ($82,000)	2024
	Cisco Research Award ($100,000)	2023
	Alexa Prize TaskBot Challenge ($250,000)	2022
	eBay Research Award ($100,000)	2022
	Amazon Research Award ($45,000 + $75,000 AWS Credits)	2021
	Microsoft Block Grant Awards ($4000)	2022
	Google Cloud Platform Research Credits ($5000)	2020

INVITED TALKS

Trustworthy and Collaborative LLM Agents
- U.S. Securities and Exchange Commission. May 28, 2025.

Trustworthy Large Language Models for Summarizing Electronic Health Records
- Penn State College of Medicine AI Grand Rounds. March 26, 2025.

Gradient-based Prompt Optimization for LLM Reasoning
- Guest Lecture at Penn State CSE/MATH 555. December 10, 2024.
- eBay Tech Talk. December 19, 2024.

How Can LLMs Help Peer Review?
- LLM for Science Birds-of-a-Feather. NAACL 2024. June 17, 2024.

Are Large Language Models Fair Summarizers?
- New Frontiers in Summarization workshop at EMNLP 2023. December 5, 2023.

Trustworthy Human-Centered Summarization
- eBay. June 13, 2023.
- Cisco. June 1, 2023.

Semantic Parsing in the Era of Large Language Models
- Penn State CSE Faculty Seminar. December 1, 2022.
- University of Pennsylvania. CLunch - Penn NLP. October 31, 2022.

Contrastive Learning for Natural Language Processing
- Amazon AWS. August 4, 2022.
- MLNLP Seminar. July 31, 2022.
- Penn State NSF REU Site: Machine Learning in Cybersecurity. July 27, 2022.
- The University of Tokyo. July 7, 2022.

FeTaQA: Free-form Table Question Answering
- Workshop on Conversational Agents with Structured Knowledge. AKBC 2021. October 7, 2021

SCoRe: Pre-Training for Context Representation in Conversational Semantic Parsing
- Google NLP Reading Group. April 19, 2021.

Building Robust Conversational Question Answering Systems Over Databases of Tabular Data
- The Center for Artificial Intelligence Foundations and Engineered Systems (CAFE). Penn State University. Oct 15, 2021.
- University of Pennsylvania. CLunch - Penn NLP. October 19, 2020.

Deep Contextual Modeling for Natural Language Understanding, Generation, and Grounding
- Department of Computer Science, Stony Brook University. April 10, 2020.
- Department of Computer Science and Engineering, The Pennsylvania State University. March 30, 2020.
- Department of Computer Science, Virginia Tech. March 19, 2020.
- Department of Computer Science and Engineering, The University of Connecticut. February 24, 2020.

Editing-based SQL Query Generation for Cross-Domain Context-Dependent Questions.
- STAR Talk 1st Place Prize. Natural Language, Dialog and Speech (NDS) Symposium, The New York Academy of Sciences, New York, November 2019.

This Email Could Save Your Life: Introducing the Task of Email Subject Line Generation.
- Invited poster at the Microsoft Research AI Breakthroughs Workshop, Redmond, September 2019.

Interleaving Thoughts of Interlocutors via RNNs for Response Selection in Multi-Turn Dialog Systems.
- Invited poster at the 5th IBM Research Cognitive Colloquium, the IBM T.J. Watson Research Center, New York, September 2016.

PUBLICATIONS

[62] VisOnlyQA: Large Vision Language Models Still Struggle with Visual Perception of Geometric Information. Ryo Kamoi, Yusen Zhang, Sarkar Snigdha Sarathi Das, Ranran Haoran Zhang, Rui Zhang. *In Conference on Language Modeling (COLM), 2025*

[61] HRScene: How Far Are VLMs from Effective High-Resolution Image Understanding? 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. *In the International Conference on Computer Vision (ICCV), 2025*

[60] GreaterPrompt: A Unified, Customizable, and High-Performing Open-Source Toolkit for Prompt Optimization. Wenliang Zheng, Sarkar Snigdha Sarathi Das, Yusen Zhang, Rui Zhang. *In the 63rd Annual Meeting of the Association for Computational Linguistics (ACL) - Demo, 2025*

[59] Enhance Multimodal Consistency and Coherence for Text-Image Plan Generation. Xiaoxin Lu, Ranran Haoran Zhang, Yusen Zhang, Rui Zhang. *In the 63rd Annual Meeting of the Association for Computational Linguistics (ACL) - Findings, 2025*

[58] TaxoAdapt: Aligning LLM-Based Multidimensional Taxonomy Construction to Evolving Re-

search Corpora. Priyanka Kargupta, Nan Zhang, Yunyi Zhang, Rui Zhang, Prasenjit Mitra, Jiawei Han. *In the 63rd Annual Meeting of the Association for Computational Linguistics (ACL), 2025*

[57] Improving Fairness of Large Language Models in Multi-document Summarization. Haoyuan Li, Rui Zhang, Snigdha Chaturvedi. *In the 63rd Annual Meeting of the Association for Computational Linguistics (ACL), 2025*

[56] From Lazy to Prolific: Tackling Missing Labels in Open Vocabulary Extreme Classification by Positive-Unlabeled Sequence Learning. Haoran Ranran Zhang, Bensu Uçar, Soumik Dey, Hansi Wu, Binbin Li, Rui Zhang. *In 2025 Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL) - Findings, 2025*

[55] Coverage-based Fairness in Multi-document Summarization. Haoyuan Li, Yusen Zhang, Rui Zhang, Snigdha Chaturvedi. *In 2025 Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL), 2025*
Senior Area Chair Paper Award

[54] SiReRAG: Indexing Similar and Related Information for Multihop Reasoning. Nan Zhang, Prafulla Kumar Choubey, Alexander Fabbri, Gabriel Bernadett-Shapiro, Rui Zhang, Prasenjit Mitra, Caiming Xiong, Chien-Sheng Wu. *In the Thirteenth International Conference on Learning Representations (ICLR), 2025*

[53] GReaTer: Gradients Over Reasoning Makes Smaller Language Models Strong Prompt Optimizers. Sarkar Snigdha Sarathi Das, Ryo Kamoi, Bo Pang, Yusen Zhang, Caiming Xiong, Rui Zhang. *In the Thirteenth International Conference on Learning Representations (ICLR), 2025*

[52] Chain of Agents: Large Language Models Collaborating on Long-Context Tasks. Yusen Zhang, Ruoxi Sun, Yanfei Chen, Tomas Pfister, Rui Zhang, Sercan Ö. Arık. *In the Thirty-Eighth Annual Conference on Neural Information Processing Systems (NeurIPS), 2024*

[51] LLMs Assist NLP Researchers: Critique Paper (Meta-)Reviewing. Jiangshu Du, Yibo Wang, Wenting Zhao, Zhongfen Deng, Shuaiqi Liu, Renze Lou, Henry Peng Zou, Pranav Narayanan Venkit, Nan Zhang, Mukund Srinath, Haoran Ranran 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. *In the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2024*

[50] FOLIO: Natural Language Reasoning with First-Order Logic. Simeng Han, Hailey Schoelkopf, Yilun Zhao, Zhenting Qi, Martin Riddell, Wenfei Zhou, James Coady, David Peng, Yujie Qiao, Luke Benson, Lucy Sun, Alex Wardle-Solano, Hannah Szabo, Ekaterina Zubova, Matthew Burtell, Jonathan Fan, Yixin Liu, Brian Wong, Malcolm Sailor, Ansong Ni, Linyong Nan, Jungo Kasai, Tao Yu, Rui Zhang, Alexander R. Fabbri, Wojciech Kryscinski, Semih Yavuz, Ye Liu, Xi Victoria Lin, Shafiq Joty, Yingbo Zhou, Caiming Xiong, Rex Ying, Arman Cohan, Dragomir Radev. *In the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2024*

[49] When Can LLMs Actually Correct Their Own Mistakes? A Critical Survey of Self-Correction of LLMs. Ryo Kamoi, Yusen Zhang, Nan Zhang, Jiawei Han, Rui Zhang. *In the Transactions of the Association for Computational Linguistics (TACL), 2024*

[48] DocMath-Eval: Evaluating Math Reasoning Capabilities of LLMs in Understanding Long and Specialized Documents. Yilun Zhao, Yitao Long, Hongjun Liu, Linyong Nan, Lyuhao Chen, Ryo Kamoi, Yixin Liu, Xiangru Tang, Rui Zhang, Arman Cohan. *In the 62nd Annual Meeting of the*

Association for Computational Linguistics (ACL), 2024

[47] FinanceMATH: Knowledge-Intensive Math Reasoning in Finance Domains. Yilun Zhao, Hongjun Liu, Yitao Long, Rui Zhang, Chen Zhao, Arman Cohan. *In the 62nd Annual Meeting of the Association for Computational Linguistics (ACL), 2024*

[46] Evaluating LLMs at Detecting Errors in LLM Responses. 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. *In the 1st Conference on Language Modeling (COLM), 2024*

[45] Veiled Pathways: Investigating Covert and Side Channels within GPU Uncore. Yuanqing Miao, Yingtian Zhang, Dinghao Wu, Danfeng Zhang, Gang Tan, Rui Zhang, Mahmut Kandemir. *In the 57th IEEE/ACM International Symposium on Microarchitecture (MICRO), 2024*

[44] Fair Abstractive Summarization of Diverse Perspectives. Yusen Zhang, Nan Zhang, Yixin Liu, Alexander Fabbri, Junru Liu, Ryo Kamoi, Xiaoxin Lu, Caiming Xiong, Jieyu Zhao, Dragomir Radev, Kathleen McKeown, Rui Zhang. *In 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2024*

[43] Pruning as a Domain-specific LLM Extractor. Nan Zhang, Yanchi Liu, Xujiang Zhao, Wei Cheng, Runxue Bao, Rui Zhang, Prasenjit Mitra, Haifeng Chen. *In 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL) - Findings, 2024*

[42] MT-Ranker: Reference-free Machine Translation Evaluation by Inter-system Ranking. Ibraheem Muhammad Moosa, Rui Zhang, Wenpeng Yin. *In the Twelfth International Conference on Learning Representations (ICLR), 2024*

[41] FaMeSumm: Investigating and Improving Faithfulness of Medical Summarization. Nan Zhang, Yusen Zhang, Wu Guo, Prasenjit Mitra, Rui Zhang. *In the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023*

[40] Unified Low-Resource Sequence Labeling by Sample-Aware Dynamic Sparse Finetuning. Sarkar Snigdha Sarathi Das, Ranran Haoran Zhang, Peng Shi, Wenpeng Yin, Rui Zhang. *In the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023*

[39] Exploring the Potential of Large Language Models in Generating Code-Tracing Questions for Introductory Programming Courses. Aysa Xuemo Fan, Ranran Haoran Zhang, Luc Paquette, Rui Zhang. *In the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP) - Findings, 2023*

[38] Hermes: Unlocking Security Analysis of Cellular Network Protocols by Synthesizing Finite State Machines from Natural Language Specifications. Abdullah Al Ishtiaq, Sarkar Snigdha Sarathi Das, Syed Md Mukit Rashid, Ali Ranjbar, Kai Tu, Tianwei Wu, Zhezheng Song, Weixuan Wang, Mujtahid Al-Islam Akon, Rui Zhang, Syed Rafiul Hussain. *In the 33rd USENIX Security Symposium (USENIX Security), 2024*

[37] XSemPLR: Benchmarking Cross-Lingual Semantic Parsing in Multiple Natural Languages and Meaning Representations. Yusen Zhang, Jun Wang, Zhiguo Wang, Rui Zhang. *In the 61th Annual Meeting of the Association for Computational Linguistics (ACL), 2023*

[36] MACSum: Controllable Summarization with Mixed Attributes. Yusen Zhang, Yang Liu, Ziyi Yang, Yuwei Fang, Yulong Chen, Dragomir Radev, Chenguang Zhu, Michael Zeng, Rui Zhang. *In*

the Transactions of the Association for Computational Linguistics (TACL), 2023

[35] ConEntail: An Entailment-based Framework for Universal Zero and Few Shot Classification with Supervised Contrastive Pretraining. Ranran Haoran Zhang, Aysa Xuemo Fan, Rui Zhang. *In The 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL), 2023*

[34] Selective Annotation Makes Language Models Better Few-Shot Learners. Hongjin Su, Jungo Kasai, Chen Henry Wu, Weijia Shi, Tianlu Wang, Jiayi Xin, Rui Zhang, Mari Ostendorf, Luke Zettlemoyer, Noah A. Smith, Tao Yu. *In The Eleventh International Conference on Learning Representations (ICLR), 2023*

[33] Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models. with many authors in the BIG-bench Team. *In The Transactions on Machine Learning Research (TMLR), 2023*

[32] XRICL: Cross-lingual Retrieval-Augmented In-Context Learning for Cross-lingual Text-to-SQL Semantic Parsing. Peng Shi, Rui Zhang, He Bai, Jimmy Lin. *In the Findings of EMNLP, 2022*

[31] ReasTAP: Injecting Table Reasoning Skills During Pre-training via Synthetic Reasoning Examples. Yilun Zhao, Linyong Nan, Zhenting Qi, Rui Zhang, Dragomir Radev. *In the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022*

[30] UnifiedSKG: Unifying and Multi-Tasking Structured Knowledge Grounding with Text-to-Text Language Models. Tianbao Xie, Chen Henry Wu, Peng Shi, Ruiqi Zhong, Torsten Scholak, Michihiro Yasunaga, Chien-Sheng Wu, Ming Zhong, Pengcheng Yin, Sida I Wang, Victor Zhong, Bailin Wang, Chengzu Li, Connor Boyle, Ansong Ni, Ziyu Yao, Dragomir Radev, Caiming Xiong, Lingpeng Kong, Rui Zhang, Noah A. Smith, Luke Zettlemoyer, Tao Yu. *In the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022*

[29] Summ^N: A Multi-Stage Summarization Framework for Long Input Dialogues and Documents. Yusen Zhang, Ansong Ni, Ziming Mao, Chen Henry Wu, Chenguang Zhu, Budhaditya Deb, Ahmed H. Awadallah, Dragomir Radev, Rui Zhang. *In the 60th Annual Meeting of the Association for Computational Linguistics (ACL), 2022*

[28] CONTaiNER: Few-Shot Named Entity Recognition via Contrastive Learning. Sarkar Snigdha Sarathi Das, Arzoo Katiyar, Rebecca J. Passonneau, Rui Zhang. *In the 60th Annual Meeting of the Association for Computational Linguistics (ACL), 2022*

[27] MultiHiertt: Numerical Reasoning over Multi Hierarchical Tabular and Textual Data. Yilun Zhao, Yunxiang Li, Chenying Li, Rui Zhang. *In the 60th Annual Meeting of the Association for Computational Linguistics (ACL), 2022*

[26] DYLE: Dynamic Latent Extraction for Abstractive Long-Input Summarization. Ziming Mao, Chen Henry Wu, Ansong Ni, Yusen Zhang, Rui Zhang, Tao Yu, Budhaditya Deb, Chenguang Zhu, Ahmed H. Awadallah, Dragomir Radev. *In the 60th Annual Meeting of the Association for Computational Linguistics (ACL), 2022*

[25] FeTaQA: Free-form Table Question Answering. Linyong Nan, Chiachun Hsieh, Ziming Mao, Xi Victoria Lin, Neha Verma, Rui Zhang, Wojciech Kryściński, Nick Schoelkopf, Riley Kong, Xiangru Tang, Murori Mutuma, Ben Rosand, Isabel Trindade, Renusree Bandaru, Jacob Cunningham, Caiming Xiong, Dragomir Radev. *In the Transactions of the Association for Computational Linguistics (TACL), 2022*

[24] Structured in Space, Randomized in Time: Leveraging Dropout in RNNs for Efficient Training. Anup Sarma, Sonali Singh, Huaipan Jiang, Rui Zhang, Mahmut T Kandemir, Chita R Das. *In the Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS), 2021*

[23] TURINGBENCH: A Benchmark Environment for Turing Test in the Age of Neural Text Generation. Adaku Uchendu, Zeyu Ma, Thai Le, Rui Zhang, Dongwon Lee. *In the Findings of EMNLP, 2021*

[22] An Exploratory Study on Long Dialogue Summarization: What Works and What's Next. Yusen Zhang*, Ansong Ni*, Tao Yu, Rui Zhang, Chenguang Zhu, Budhaditya Deb, Asli Celikyilmaz, Ahmed Hassan Awadallah, Dragomir Radev. *In the Findings of EMNLP, 2021*

[21] Cross-language Sentence Selection via Data Augmentation and Rationale Training. Yanda Chen, Chris Kedzie, Suraj Nair, Petra Galuščáková, Rui Zhang, Douglas Oard, Kathleen McKeown. *In the 59th Annual Meeting of the Association for Computational Linguistics (ACL), 2021*

[20] Logic-Consistency Text Generation from Semantic Parses. Chang Shu*, Yusen Zhang*, Xiangyu Dong, Peng Shi, Tao Yu, Rui Zhang. *In the Findings of ACL, 2021*

[19] SCoRe: Pre-Training for Context Representation in Conversational Semantic Parsing. Tao Yu, Rui Zhang, Oleksandr Polozov, Christopher Meek, Ahmed Hassan Awadallah. *In the Ninth International Conference on Learning Representations (ICLR), 2021*

[18] DART: Open-Domain Structured Data Record to Text Generation. Linyong Nan, Dragomir Radev, Rui Zhang, Amrit Rau, Abhinand Sivaprasad, Chiachun Hsieh, Nazneen Fatema Rajani, Xiangru Tang, Aadit Vyas, Neha Verma, Pranav Krishna, Yangxiaokang Liu, Nadia Irwanto, Jessica Pan, Faiaz Rahman, Ahmad Zaidi, Murori Mutuma, Yasin Tarabar, Ankit Gupta, Tao Yu, Yi Chern Tan, Xi Victoria Lin, Caiming Xiong, Richard Socher. *In 2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2021*

[17] ESPRIT: Explaining Solutions to Physical ReasonIng Tasks. Nazneen Fatema Rajani*, Rui Zhang*, Yi Chern Tan, Stephan Zheng, Jeremy Weiss, Aadit Vyas, Abhijit Gupta, Caiming Xiong, Richard Socher, Dragomir Radev. *In the 58th Annual Meeting of the Association for Computational Linguistics (ACL), 2020*

[16] Editing-based SQL Query Generation for Cross-Domain Context-Dependent Questions. Rui Zhang, Tao Yu, He Yang Er, Sungrok Shim, Eric Xue, Xi Victoria Lin, Tianze Shi, Caiming Xiong, Richard Socher and Dragomir Radev. *In the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019*

[15] CoSQL: A Conversational Text-to-SQL Challenge Towards Cross-Domain Natural Language Interfaces to Databases. Tao Yu, Rui Zhang, He Yang Er, Suyi Li, Eric Xue, Bo Pang, Xi Victoria Lin, Yi Chern Tan, Tianze Shi, Zihan Li, Youxuan Jiang, Michihiro Yasunaga, Sungrok Shim, Tao Chen, Alexander Fabbri, Zifan Li, Luyao Chen, Yuwen Zhang, Shreya Dixit, Vincent Zhang, Caiming Xiong, Richard Socher, Walter Lasecki and Dragomir Radev. *In the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019*

[14] This Email Could Save Your Life: Introducing the Task of Email Subject Line Generation. Rui Zhang, Joel Tetreault. *In the 57th Annual Meeting of the Association for Computational Linguistics (ACL), 2019*

[13] Improving Low-Resource Cross-lingual Document Retrieval by Reranking with Deep Bilingual Representations. Rui Zhang, Caitlin Westerfield, Sungrok Shim, Garrett Bingham, Alexander Fabbri, Neha Verma, William Hu, Dragomir Radev. *In the 57th Annual Meeting of the Association for*

Computational Linguistics (ACL), 2019

[12] SParC: Cross-Domain Semantic Parsing in Context. Tao Yu, Rui Zhang, Michihiro Yasunaga, Yi Chern Tan, Xi Victoria Lin, Suyi Li, Heyang Er, Irene Li, Bo Pang, Tao Chen, Emily Ji, Shreya Dixit, David Proctor, Sungrok Shim, Jonathan Kraft, Vincent Zhang, Caiming Xiong, Richard Socher and Dragomir Radev. *In the 57th Annual Meeting of the Association for Computational Linguistics (ACL), 2019*

[11] ScisummNet: A Large Annotated Corpus and Content-Impact Models for Scientific Paper Summarization with Citation Networks. Michihiro Yasunaga, Jungo Kasai, Rui Zhang, Alexander Fabbri, Irene Li, Dan Friedman, Dragomir Radev. *In the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI), 2019*

[10] Surprise Languages: Rapid-Response Cross-Language IR. Douglas Oard, Petra Galuščáková, Kathleen McKeown, Marine Carpuat, Mohamed Elbadrashiny, Ramy Eskander, Kenneth Heafield, Efsun Kayi, Chris Kedzie, Smaranda Muresan, Suraj Nair, Xing Niu, Dragomir Radev, Anton Ragni, Han-Chin Shing, Yan Virin, Weijia Xu, Rui Zhang, Elena Zotkina, Joseph Barrow, Mark Gales. *In the Ninth International Workshop on Evaluating Information Access (EVIA), 2019*

[9] SyntaxSQLNet: Syntax Tree Networks for Complex and Cross-Domain Text-to-SQL Task. Tao Yu, Michihiro Yasunaga, Kai Yang, Rui Zhang, Dongxu Wang, Zifan Li, Dragomir Radev. *In the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2018*

[8] Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Semantic Parsing and Text-to-SQL Task. Tao Yu, Rui Zhang, Kai Yang, Michihiro Yasunaga, Dongxu Wang, Zifan Li, James Ma, Irene Li, Qingning Yao, Shanelle Roman, Zilin Zhang, Dragomir Radev. *In the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2018*

[7] Neural Coreference Resolution with Deep Biaffine Attention by Joint Mention Detection and Mention Clustering. Rui Zhang, Cícero Nogueira dos Santos, Michihiro Yasunaga, Bing Xiang, Dragomir Radev. *In the 56th Annual Meeting of the Association for Computational Linguistics (ACL), 2018*

[6] Improving Text-to-SQL Evaluation Methodology. Catherine Finegan-Dollak, Jonathan K. Kummerfeld, Li Zhang, Karthik Ramanathan, Sesh Sadasivam, Rui Zhang, Dragomir Radev. *In the 56th Annual Meeting of the Association for Computational Linguistics (ACL), 2018*

[5] TypeSQL: Knowledge-based Type-Aware Neural Text-to-SQL Generation. Tao Yu, Zifan Li, Zilin Zhang, Rui Zhang, Dragomir Radev. *In the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT), 2018*

[4] Addressee and Response Selection in Multi-Party Conversations with Speaker Interaction RNNs. Rui Zhang, Honglak Lee, Lazaros Polymenakos, Dragomir Radev. *In The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI), 2018*

[3] Graph-based Neural Multi-Document Summarization. Michihiro Yasunaga, Rui Zhang, Kshitijh Meelu, Ayush Pareek, Krishnan Srinivasan, Dragomir Radev. *In the Conference on Computational Natural Language Learning (CoNLL), 2017*

[2] Effects of Text Corpus Properties on Short Text Clustering Performance. Catherine Finegan-Dollak, Reed Coke, Rui Zhang, Xiangyi Ye, Dragomir Radev. *In the 54th Annual Conference of the Association for Computational Linguistics (ACL), 2016*

[1] Dependency Sensitive Convolutional Neural Networks for Modeling Sentences and Documents.

Rui Zhang, Honglak Lee, Dragomir Radev. *In the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT), 2016*

WORKSHOP
PUBLICATIONS

[8] Demystify Verbosity Compensation Behavior of Large Language Models. Yusen Zhang, Sarkar Snigdha Sarathi Das, Rui Zhang. *In the Second Workshop on Uncertainty-Aware NLP at EMNLP 2025, 2025*

[7] Large Language Models for Mathematical Reasoning: Progresses and Challenges. Janice Ahn, Rishu Verma, Renze Lou, Di Liu, Rui Zhang, Wenpeng Yin. *In the Student Research Workshop at EACL 2024, 2024*

[6] MIA 2022 Shared Task: Evaluating Cross-lingual Open-Retrieval Question Answering for 16 Diverse Languages. Akari Asai, Shayne Longpre, Jungo Kasai, Chia-Hsuan Lee, Rui Zhang, Junjie Hu, Ikuya Yamada, Jonathan H Clark, Eunsol Choi. *In the Workshop on Multilingual Information Access (MIA) at NAACL 2022, 2022*

[5] Cross-Lingual Training with Dense Retrieval for Document Retrieval. Peng Shi, Rui Zhang, He Bai, Jimmy Lin. *In the First Workshop on Multilingual Representation Learning at EMNLP 2021, 2021*

[4] Prefix-to-SQL: Text-to-SQL Generation from Incomplete User Questions. Naihao Deng, Shuaichen Chang, Peng Shi, Tao Yu, Rui Zhang. *In the First Workshop on Natural Language Processing for Programming (NLP4Prog) at ACL 2021, 2021*

[3] EVOQUER: Enhancing Temporal Grounding with Video-Pivoted Back Query Generation. Yanjun Gao, Lulu Liu, Jason Wang, Xin Chen, Huayan Wang, Rui Zhang. *In the Visually Grounded Interaction and Language (ViGIL) at NAACL 2021, 2021*

[2] Did You Ask a Good Question? A Cross-Domain Question Intention Classification Benchmark for Text-to-SQL. Yusen Zhang, Xiangyu Dong, Shuaichen Chang, Tao Yu, Peng Shi, Rui Zhang. *In the First workshop on Interactive and Executable Semantic Parsing at EMNLP 2020, 2020*

[1] MATERIALizing Cross-Language Information Retrieval: A Snapshot. Petra Galuščáková, Douglas Oard, Joe Barrow, Suraj Nair, Shing Han-Chin, Elena Zotkina, Ramy Eskander, Rui Zhang. *In the Workshop on Cross-Language Search and Summarization of Text and Speech at LREC 2020, 2020*

PATENT

[1] Modeling Multiparty Conversation Dynamics: Speaker, Response, Addressee Selection Using a Novel Deep Learning Approach. Rui Zhang, Lazaros Polymenakos, Dragomir Radev, David Nahamoo, Honglak Lee. *US20190341036*, 2019

TUTORIALS

EMNLP 2024 tutorial: Enhancing LLM Capabilities Beyond Scaling Up

NAACL 2022 tutorial: Contrastive Data and Learning for Natural Language Processing

WORKSHOPS

EMNLP 2025 workshop: New Frontiers in Summarization (NewSumm)

IJCAI 2024 workshop: AI for Research (AI4Research)

NAACL 2022 workshop: Structured and Unstructured Knowledge Integration (SUKI)

NAACL 2022 workshop: Multilingual Information Access (MIA)

EMNLP 2020 workshop: Interactive and Executable Semantic Parsing (IntEx-SemPar)

SERVICE

Panelist, NSF Panel, 2025

Panelist, NSF Panel, 2024

Panelist, NSF Panel, 2022

Tutorial Chair, NAACL 2025

Tutorial Chair, NAACL 2024

Area Chair, ICLR 2026

Area Chair, NeurIPS 2025

Area Chair, ACL 2025

Area Chair for Demo, ACL 2025

Best System Demonstration Award Committee, ACL 2025

Area Chair, COLING 2025

Area Chair, NeurIPS 2024

Area Chair, EMNLP 2024

Area Chair, NeurIPS 2023

Area Chair, EMNLP 2023

Area Chair, AACL 2023

Area Chair (Summarization Track), EMNLP 2022

Area Chair (Question Answering Track), NLPCC 2022

Area Chair (Summarization Track), EMNLP 2021

Area Chair (Summarization Track), NAACL 2021

Area Chair (Summarization and Generation Track), NLPCC 2021

Action Editor, Transactions on Machine Learning Research (TMLR), 2024-2025

Journal Reviewer, Transactions of the Association for Computational Linguistics (TACL), 2022-2025

Journal Reviewer, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2024

Journal Reviewer, Journal of Artificial Intelligence Research (JAIR), 2024

Journal Reviewer, Transactions on Knowledge and Data Engineering (TKDE), 2019

Journal Reviewer, Transactions on Asian and Low-Resource Language Information Processing (TAL-LIP), 2019, 2023

Faculty Mentor, IJCNLP-AACL 2025 Student Research Workshop

Organizing Committee member, The 12th Mid-Atlantic Student Colloquium on Speech, Language and Learning (MASC-SLL 2025), 2025

Panelist, KnowledgeNLP: Workshop on Knowledge Augmented Methods for NLP at AAAI 2023

Mentor, Applying for a PhD in NLP at ACL Mentorship Program 2022

Group Mentoring, the Annual Meeting of the Association for Computational Linguistics (ACL), 2020

Birds-of-a-Feather Meetup Host, the Annual Meeting of the Association for Computational Linguistics (ACL), 2020

Program Committee member, The 4th Workshop on New Frontiers in Summarization at EMNLP 2023

Program Committee member, Workshop for Summarization of Creative Text at COLING 2022

Program Committee member, The 3rd Workshop on New Frontiers in Summarization at EMNLP 2021

Program Committee member, The First Workshop on Natural Language Processing for Programming (NLP4Prog) at ACL 2021

Program Committee member, The 2nd Workshop on Data Science with Human-in-the-loop: Language Advances (DaSH-LA) at NAACL 2021

Program Committee member, The Workshop on Cross-Language Search and Summarization over Text and Speech (CLSSTS) at LREC 2020

Program Committee member, The 3rd Workshop on Natural Language Generation from the Semantic Web (WebNLG+ 2020)

Program Committee member, The 13th Workshop on Graph-Based Natural Language Processing (TextGraphs 2019) at EMNLP 2019

Program Committee member, The Workshop on e-Commerce and NLP (ECNLP 2019) at WWW 2019

Reviewer, Conference on Language Modeling (COLM), 2024

Reviewer, ACL Rolling Review, 2024

Reviewer, the 2022 China National Conference on Computational Linguistics (CCL), 2022

Reviewer, ACL Rolling Review, 2022

Reviewer, ACL Rolling Review, 2021

Reviewer, 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021

Reviewer, the Thirty-Fifth Conference on Neural Information Processing Systems (NeurIPS), 2021

Reviewer, the Ninth International Conference on Learning Representations (ICLR), 2021

Reviewer, the Thirty-seventh International Conference on Machine Learning (ICML), 2021

Reviewer, the Thirty-seventh International Conference on Machine Learning (ICML), 2020

Program Committee member, the Annual Meeting of the Association for Computational Linguistics (ACL), 2021

Program Committee member, the Annual Meeting of the Association for Computational Linguistics (ACL), 2020

Program Committee member, the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019

Program Committee member, the Annual Meeting of the Association for Computational Linguistics (ACL), 2019

Program Committee member, the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2019

Program Committee member, the AAAI Conference on Artificial Intelligence (AAAI), 2019

Student Volunteer, the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2016

Ph.D. Students

Yusen Zhang (Dr. Tse-Yun Feng Graduate Student Award)
Sarkar Snigdha Sarathi Das (Dr. Tse-Yun Feng Graduate Student Award)
Nan Zhang
Ranran Haoran Zhang
Xiaoxin Lu
Ryo Kamoi
Hieu Ngoc Nguyen
Binglin Zhou

Master's Students with Thesis

Chandan Akiti
Tianyang Zhao
Raj Ratn Pranesh
Sumit Kumar
Avi Iyer
Aashrith Madasu

Ph.D. Thesis Committee

Yanjun Gao
Maryam Zare
Athar Sefid
Saranya Venkatraman
Adaku Uchendu
Mukund Srinath
Muchao Ye
Zeinab Hakimi
Yuyang Deng
Zhaohui Li
Junyu Luo
Shijia Zhang
Jooyoung Lee
Vipul Gupta
Hangfan Zhang
Quan Li
Pouria Mahdavinia
Jiangtao Liu
Ting-Yao Hsu
Amutheezan Sivagnanam
Ashirbad Mishra
Abdullah Al Ishtiaq
Shihao Xia
Levent Toksoz
Yiran Wu
Berk Atil
Xiaochen Wang
Nafis Irtiza Tripto
Ruihao Pan
Tianrong Zhang
Bochuan Cao
Shu Zhao
Xiang Li
Xianren Zhang
Xinjie Li

Aofei Chang
Hamed Mahdavi
Mingyu Zhu
Ali Al-Lawati
Xin Gu
Zhenghao Li

Master's Thesis Committee
Saniya Naphade
Zeba Karishma
Rajal Nivargi
Sai Ajay Modukuri
Kavya Laalasa Karanam
Purushartha Singh
Abhijeet Kumar
Ashwath Raghav Swaminathan
Abdullah Al Ishtiaq
Parker Sell
Yuchen Wei

Undergraduate Students with Schreyer Honors Thesis
Jason Wang (CAFE AI Award for Best Undergraduate Honors Thesis)
Jacob Cunningham
Noah Gearhart
Michael Chan
Qiyu Chen
Yucheng Fang

Other Undergraduate Students
Lulu Liu
Theodore Steiner (UNC at Chapel Hill, REU)
Wenliang Zheng
Muyu Pan
Eugene Kwek

TEACHING

As Instructors

CSE 587 Deep Learning for NLP, Spring 2022, Spring 2023, Penn State University

CMPSC 448 Machine Learning, Fall 2020, Fall 2021, Fall 2022, Fall 2024, Penn State University

CMPSC 442 Artificial Intelligence, Spring 2021, Spring 2024, Spring 2025 Penn State University

As Teaching Assistants

Natural Language Processing, Spring 2018, Yale University

Artificial Intelligence, Fall 2017, Yale University

Natural Language Processing, Fall 2016, University of Michigan

Natural Language Processing, Coursera

Artificial Intelligence, Winter 2015, University of Michigan

Intro to Programming, Fall 2014, University of Michigan

East Asian Study, Summer 2013, Shanghai Jiao Tong University