

Rui Zhang

W329
Westgate Building
University Park, PA 16802

rmz5227@psu.edu
<https://ryanzhumich.github.io/>
+1-814-867-3870

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

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

PUBLICATIONS

[33] 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

[32] 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

[31] 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

[30] 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
- [29] 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
- [28] 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
- [27] 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
- [26] 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
- [25] 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
- [24] 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
- [23] 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
- [22] 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
- [21] 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
- [20] 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
- [19] 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

[18] 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.

*: Equal Contribution

In the 58th Annual Meeting of the Association for Computational Linguistics (ACL), 2020

[17] 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 LREC 2020 Workshop on Cross-Language Search and Summarization of Text and Speech, 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] SPaC: 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 Elbadrahy, 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

[3] 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

[2] 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

	<p>[1] 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. <i>In the First workshop on Interactive and Executable Semantic Parsing at EMNLP 2020, 2020</i></p>	
PATENT	<p>[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. <i>US20190341036</i>, 2019</p>	
GRANTS	Microsoft Block Grant Awards (\$4000)	2022
	Amazon Research Award (\$45,000 + \$75,000 AWS Credits)	2021
	Google Cloud Platform Research Credits (\$5000)	2020
PROFESSIONAL SERVICE	<p>Presenter, Tutorial on Contrastive Data and Learning for Natural Language Processing at NAACL 2022</p> <p>Organizing Committee member, SUKI: Workshop on Structured and Unstructured Knowledge Integration at NAACL 2022</p> <p>Organizing Committee member, MIA: Workshop on Multilingual Information Access at NAACL 2022</p> <p>Organizing Committee member, IntEx-SemPar: Workshop on Interactive and Executable Semantic Parsing at EMNLP 2020</p> <p>Area Chair (Summarization Track), EMNLP 2022</p> <p>Area Chair (Question Answering Track), NLPCC 2022</p> <p>Area Chair (Summarization Track), EMNLP 2021</p> <p>Area Chair (Summarization Track), NAACL 2021</p> <p>Area Chair (Summarization and Generation Track), NLPCC 2021</p> <p>Program Committee member, Workshop for Summarization of Creative Text at COLING 2022</p> <p>Program Committee member, The 3rd Workshop on New Frontiers in Summarization at EMNLP 2021</p> <p>Program Committee member, The First Workshop on Natural Language Processing for Programming (NLP4Prog) at ACL 2021</p> <p>Program Committee member, The 2nd Workshop on Data Science with Human-in-the-loop: Language Advances (DaSH-LA) at NAACL 2021</p> <p>Program Committee member, The Workshop on Cross-Language Search and Summarization over Text and Speech (CLSSTS) at LREC 2020</p> <p>Program Committee member, The 3rd Workshop on Natural Language Generation from the Seman-</p>	

tic 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 of TACL, 2022-2024

Reviewer, CCL 2022

Reviewer of 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

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 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

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

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

Student Volunteer, the Annual Conference of the North American Chapter of the Association for

STUDENTS

Ph.D. Students

Nan Zhang (Penn State)
Sarkar Das (Penn State)
Yusen Zhang (Penn State)
Haoran Zhang (Penn State)

Ph.D. Thesis Committee

Yanjun Gao (Penn State)
Maryam Zare (Penn State)
Athar Sefid (Penn State)
Saranya Venkatraman (Penn State)
Adaku Uchendu (Penn State)
Maryam Tabar (Penn State)
Mukund Srinath (Penn State)
Muchao Ye (Penn State)
Zeinab Hakimi (Penn State)
Yuyang Deng (Penn State)
Zhaohui Li (Penn State)
Junyu Luo (Penn State)
Shijia Zhang (Penn State)

Master's Students

Tianyang Zhao (Penn State)
Chandan Akiti (Penn State)

Master's Thesis Committee

Saniya Naphade (Penn State)
Zeba Karishma (Penn State)
Rajal Nivargi (Penn State)
Sai Ajay Modukuri (Penn State)
Kavya Laalasa Karanam (Penn State)
Purushartha Singh (Penn State)
Abhijeet Kumar (Penn State)

Undergraduate Students

Lulu Liu (Penn State)
Jason Wang (Penn State, Schreyer Honors Thesis Supervisor)
Jacob Cunningham (Penn State, Schreyer Honors Thesis Supervisor)
Noah Gearhart (Penn State, Schreyer Honors Thesis Supervisor)
Michael Chan (Penn State, Schreyer Honors Thesis Supervisor)
Qiyu Chen (Penn State, Schreyer Honors Thesis Supervisor)
Theodore Steiner (UNC at Chapel Hill, REU)

TEACHING

CSE 597 Deep Learning for NLP, Spring 2022, Spring 2023, Penn State University
CMPSC 448 Machine Learning, Fall 2020, Fall 2021, Fall 2022, Penn State University
CMPSC 442 Artificial Intelligence, Spring 2021, Penn State University

TEACHING
ASSISTANT

Natural Language Processing, Spring 2018, Yale University
Artificial Intelligence, Fall 2017, Yale University
Natural Language Processing, Fall 2016, University of Michigan

course evaluation: 4.4/5.0
course evaluation: 4.8/5.0

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

HONORS AND AWARDS

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

TALKS AND PRESENTATIONS

Invited Talk. Amazon AWS. August 4, 2022.

Invited Talk. MLNLP. July 31, 2022.

Invited Talk. Penn State NSF REU Site: Machine Learning in Cybersecurity. July 27, 2022.

Invited Talk. The University of Tokyo. July 7, 2022.

Invited Talk. NLP Reading Group, Google. April 19, 2021.

Invited Talk. Department of Computer Science, University of Pennsylvania. October 19, 2020.

Invited Talk. Department of Computer Science, Stony Brook University. April 10, 2020.

Invited Talk. Department of Computer Science, The Pennsylvania State University. March 30, 2020.

Invited Talk. Department of Computer Science, Virginia Tech. March 19, 2020.

Invited Talk. 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 Sys-
 tems”.
 Invited poster at the 5th IBM Research Cognitive Colloquium, the IBM T.J. Watson Research
 Center, New York, September 2016