

Ziyu Yao

Department of Computer Science
George Mason University
Nguyen Engineering Building 4415
Fairfax, VA 22030 USA

ziyuyao.org ziyuyao@gmu.edu [ZiyuYao](https://twitter.com/ZiyuYao) [GScholar](#)

(Last update: August 12, 2022)

Research Interests

I am interested in broad natural language processing and artificial intelligence problems. In particular, my current research explores building natural language interfaces (e.g., question answering systems) that can interact with humans while assisting them in knowledge discovery and task completion. Some of my work is about interdisciplinary research topics, such as AI/NLP for Software Engineering and Bioinformatics.

Education

- 2015–2021 **M.S./Ph.D.**, The Ohio State University (OSU), Computer Science & Engineering.
(Thesis: On Advancing Natural Language Interfaces: Data Collection, Model Development, and User Interaction. Advisor: Prof. Huan Sun)
- 2011–2015 **B.E.**, Beijing University of Posts and Telecommunications (BUPT), Communication Engineering.

Professional Experience

- 2021–present **Assistant Professor**, George Mason University (GMU).
Affiliation: C4I & Cyber Center, Center for Advancing Human-Machine Partnership, and Institute for Digital InnovAtion (IDIA) at GMU
- 2021 Summer **Research Intern**, Microsoft Semantic Machines (Mentor: Anthony Platanios).
- 2020 Summer **Research Intern**, LTI, Carnegie Mellon University (Mentor: Graham Neubig).
- 2018 Summer **Research Intern**, Microsoft Research, Redmond (Mentor: Yizhe Zhang, Bill Dolan).
- 2017 Summer **Research Intern**, Fujitsu Lab of America (Mentor: Mehdi Bahrami, Wei-Peng Chen).
- 2014–2015 **Research Intern**, Tsinghua University, Beijing (Mentor: Minlie Huang).

Awards and Honors

- 2021 **Best Paper Award**, The IEEE International Conference on Bioinformatics and Biomedicine (BIBM).
- 2021 **Graduate Student Research Award**, OSU, CSE.
- 2020 **Presidential Fellowship**, OSU.
- 2020 **Rising Star in EECS**, UC Berkeley.
- 2020 **Graduate Students Poster Award (Honorable Mention)**, CSE, OSU.
- 2019 **NSF Student Travel Award**, The Web Conference (WWW).
- 2019 **Graduate Students Poster Award (Honorable Mention)**, CSE, OSU.
- 2018 **Travel Grant for Graduate Cohort for Women, San Francisco**, CRA-W.
- 2015 **Beijing Outstanding Graduate Award**, Beijing, China.
- 2012, 2014 **First-class Scholarship (top 3%)**, BUPT, China.
- 2013 **Second-class Scholarship (top 5%)**, BUPT, China.

Publications

- Tianbao Xie*, Chen Henry Wu*, Peng Shi, Ruiqi Zhong, Torsten Scholak, Michihiro Yasunaga, Chien-Sheng Wu, Ming Zhong, Pengcheng Yin, Sida 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, “UnifiedSKG: Unifying and Multi-Tasking Structured Knowledge Grounding with Text-to-Text Language Models” *equal contribution. [[Preprint](#), [Code](#), [Project Website](#)]
- Wenhui Chen, Xinyun Chen, Zhiyu Chen, **Ziyu Yao**, Michihiro Yasunaga, Tao Yu, Rui Zhang, “Proceedings of the Workshop on Structured and Unstructured Knowledge Integration (SUKI)”, **co-located at NAACL 2022**. [[Workshop Proceedings](#), [Website](#)]
- Peizhao Li, Xuchao Zhang, **Ziyu Yao**, Wei Cheng, Haifeng Chen, Hongfu Liu, “Code Editing from Few Exemplars by Adaptive Multi-Extent Composition”, The Deep Learning for Code Workshop at The Tenth International Conference on Learning Representations (**ICLR 2022 DL4Code**). [[Paper](#)]
- Xiang Yue, **Ziyu Yao**, Huan Sun, “Synthetic Question Value Estimation for Domain Adaptation of Question Answering”, The 60th Annual Meeting of the Association for Computational Linguistics (**ACL 2022, acceptance rate: 20.75%**). [[Paper](#), [Code](#)]
- **Ziyu Yao**, “On Advancing Natural Language Interfaces: Data Collection, Model Development, and User Interaction”, The Ohio State University, **PhD Dissertation, 2021**. [[Dissertation](#)]
- **Ziyu Yao**, Frank Xu, Pengcheng Yin, Huan Sun, Graham Neubig, “Learning Structural Edits via Incremental Tree Transformations”, The Ninth International Conference on Learning Representations (**ICLR 2021**). [[Paper](#), [Code](#)]
- **Ziyu Yao**, Yiqi Tang, Scott Wen-tau Yih, Huan Sun, Yu Su, “An Imitation Game for Learning Semantic Parsers from User Interaction”, 2020 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2020, acceptance rate: 22.4%**). [[Paper](#), [Code](#)]
- Xiang Yue, Xinliang Frederick Zhang, **Ziyu Yao**, Simon Lin, Huan Sun, “CliniQG4QA: Generating Diverse Questions for Domain Adaptation of Clinical Question Answering”, **IEEE BIBM 2021 (Best Paper Award)**. [[Paper](#), [Code](#)]
- Royi Lachmy, **Ziyu Yao**, Greg Durrett, Milos Gligoric, Junyi Jessy Li, Ray Mooney, Graham Neubig, Yu Su, Huan Sun, Reut Tsarfaty, “Proceedings of the 1st Workshop on Natural Language Processing for Programming (NLP4Prog 2021)”, **co-located at ACL-IJCNLP 2021**. [[Workshop Proceedings](#), [Website](#)]
- **Ziyu Yao**, Yu Su, Huan Sun, Scott Wen-tau Yih, “Model-based Interactive Semantic Parsing: A Unified Formulation and A Text-to-SQL Case Study”, 2019 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2019, acceptance rate: 23.8%**). [[Paper](#), [Code](#)]
- Boyuan Pan, Hao Li, **Ziyu Yao**, Deng Cai, Huan Sun, “Reinforced Dynamic Reasoning for Conversational Question Generation”, Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (**ACL 2019, acceptance rate: 25.7%**). [[Paper](#), [Code](#)]
- **Ziyu Yao**, Jayavardhan Reddy Peddamail, Huan Sun, “CoaCor: Code Annotation for Code Retrieval with Reinforcement Learning”, The Web Conference 2019 (**WWW 2019, acceptance rate: 18%**). [[Paper](#), [Code](#)]
- **Ziyu Yao**, Xiujun Li, Jianfeng Gao, Brian Sadler, Huan Sun, “Interactive Semantic Parsing for If-Then Recipes via Hierarchical Reinforcement Learning”, The AAAI Conference on Artificial Intelligence 2019 (**AAAI 2019, acceptance rate: 16.2%; SPOTLIGHT**). [[Paper](#), [Code](#)]

- **Ziyu Yao**, Yizhe Zhang, Xiujun Li, Jianfeng Gao, Michel Galley, Chris Brockett, Huan Sun, Bill Dolan, “IEC: Towards Interest-Eliciting Neural Conversational Agents”, **Manuscript 2019**, [Paper].
- Jayavardhan Reddy Peddamail, **Ziyu Yao**, Zhen Wang, Huan Sun, “A Comprehensive Study of StaQC for Deep Code Summarization”, **SIGKDD 2018 Deep Learning Day**. [Paper]
- **Ziyu Yao**, Daniel S. Weld, Wei-Peng Chen, Huan Sun, “StaQC: A Systematically Mined Question-Code Dataset from Stack Overflow”, The Web Conference 2018 (**WWW 2018**, acceptance rate: **14.8%**). [Paper, Code]
- Li Zhao, Minlie Huang, **Ziyu Yao**, Rongwei Su, Yingying Jiang, Xiaoyan Zhu. “Semi-Supervised Multinomial Naive Bayes for Text Classification by Leveraging Word-Level Statistical Constraint”, The AAI Conference on Artificial Intelligence 2016 (**AAAI 2016**, acceptance rate: **25.8%**). [Paper]
- **Ziyu Yao**, Shouzhong Tu, Minlie Huang, Xiaoyan Zhu. “A Semi-Supervised Method for Filtering Chinese Spam Tweets”. **Journal of Chinese Information Processing 2016**, 30(5): 176-186. [Paper in Chinese]

Talks

- 2022 **Building Interactive and Adaptive Question Answering Systems**, Invited talk at CLSP (Center for Language and Speech Processing) Seminar, Johns Hopkins University. Host: Kevin Duh, Ben Van Durme.
- 2022 **Building Interactive and Adaptive Question Answering Systems**, Invited talk at Center for Advancing Human-Machine Partnership, George Mason University.
- 2021 **Building Interactive Natural Language Interfaces**, Invited talk at Brandeis University (virtually). Host: Hongfu Liu.
- 2021 **Building Interactive Natural Language Interfaces**, OSU AI Club (virtually).
- 2020 **An Imitation Game for Learning Semantic Parsers from User Interaction**, EMNLP (virtually).
- 2020 **On Building an Interactive Natural Language Interface: An intelligent agent that can communicate, collaborate and learn**, Invited talk at Microsoft Research AI Breakthroughs Workshop (virtually). Invited by: Ahmed Hassan Awadallah, Alex Polozov.
- 2020 **Building Natural Language Interfaces to Programs**, Invited talk at VMware, Beijing (virtually). Host: Jin He, Kevin Song.
- 2020 **Towards Building Interactive and Collaborative Natural Language Interfaces**, Invited talk at NeuLab, Carnegie Mellon University (virtually). Host: Graham Neubig.
- 2020 **Learning Semantic Parsers from User Interaction**, CSE AI Seminar, OSU.
- 2019 **Towards Building Interactive and Collaborative Natural Language Interfaces**, Invited talk at Chair of Management Information System, ETH Zurich, Host: Stefan Feuerriegel.
- 2019 **CoaCor: Code Annotation for Code Retrieval with Reinforcement Learning**, WWW, San Francisco, USA.
- 2019 **Interactive Semantic Parsing for If-Then Recipes via Hierarchical Reinforcement Learning**, AAI, Hawaii, USA.
- 2018 **Interactive Semantic Parsing for If-Then Recipes via Hierarchical Reinforcement Learning**, CSE AI Seminar, OSU.
- 2018 **StaQC: A Systematically Mined Question-Code Dataset from Stack Overflow**, WWW, Lyon, France.

Grants

“**AI Engineering Seedlings**”, US Department of Defense, contract from UMD Applied Research Laboratory for Intelligence and Security (ARLIS), \$100,000, 07/2021–07/2022. Role: Co-PI.

Teaching Experience

At GMU:

- 2022 Fall **CS678: Advanced Natural Language Processing.**
- 2022 Spring **CS499: Natural Language Processing.**
- 2021 Fall **CS695: Special Topics on Natural Language Processing.**

At OSU:

- 2020 Fall **CSE Panel: Graduate School Experience and Expectation**, CSE, OSU. Panelist.
- 2020 Fall **Foundations of Speech and Language Processing**, CSE, OSU. Guest lecturer.
- 2018 Fall **Course Panel: Data Management in the Cloud**, CSE, OSU. Guest panelist.
- 2017 Spring **Introduction to Computer Programming in C++**, CSE, OSU. Student instructor.

Mentoring Experience

Mentored PhD Students at GMU.

Saurabh Srivastava (2022-)
Daking Rai (2022-)
Hao Yan (2022-)

(Co-)Mentored/Collaborated Students at OSU.

Jayavardhan Reddy Peddamail (Master Student; now at Google)
Varun Ranganathan Dhanvanth (Master Student; now at Facebook)
Heming Sun (Undergraduate Student)
Tianshu Zhang (PhD Student)
Leslie Zhou (Master Student)

School/Department Services

Department Committees at GMU.

Faculty Search Committee (Fall'21 - Spring'22)

Comprehensive Exam Committees at GMU.

Fahim Faisal (Fall'22)
Jonathan Mbuya (Fall'22)
Md Mahfuz Ibn Alam (Spring'22)
Abhisekh Rana (Spring'22)

Professional Activities

Editor Services.

ACM SIGAI: [Newsletter](#) co-editor

2022 –

Organizing Committees.

SUKI at NAACL2022: Workshop on Structured and Unstructured Knowledge Integration 2022
NLP4Prog at ACL2021: The First Workshop on NLP for Programming 2021

Area Chairs and Senior Program Committees.

<u>AAAI 2023</u> : Fast Track	2023
<u>EMNLP 2022</u> : Efficient NLP Track	2022

Reviewers (Journal).

<u>TKDD</u> : ACM Transactions on Knowledge Discovery from Data	2022 –
<u>TKDE</u> : IEEE Transactions on Knowledge and Data Engineering	2022 –

Program Committees/Reviewers (Conference).

<u>ARR</u> : ACL Rolling Review	2021 –
<u>MSR</u> : Mining Software Repositories Conference	2021
<u>NAACL</u> : Annual Conference of the North American Chapter of the Association for Computational Linguistics	2021
<u>EMNLP</u> : Conference on Empirical Methods in Natural Language Processing	2018, 2020, 2021
<u>AACL</u> : Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics	2020
<u>AAAI</u> : The AAAI Conference on Artificial Intelligence	2020
<u>NLPCC</u> : Conference on Natural Language Processing and Chinese Computing	2020
<u>CoNLL</u> : Conference on Computational Natural Language Learning	2019

Secondary/External Reviewer:

<u>ACL</u> : Annual Meeting of the Association for Computational Linguistics	2018, 2020
<u>KDD</u> : SIGKDD Conference on Knowledge Discovery and Data Mining	2017, 2019, 2020
<u>NAACL</u> : Annual Conference of the North American Chapter of the Association for Computational Linguistics	2019
<u>ICDM</u> : International Conference on Data Mining	2016, 2018
<u>WWW</u> : International World Wide Web Conference	2017
<u>CIKM</u> : International Conference on Information and Knowledge Management	2016

Program Committees/Reviewers (Workshop).

<u>HCI+NLP at NAACL 2022</u> : Bridging Human-Computer Interaction and Natural Language Processing	2022
<u>DL4Code at ICLR 2022</u> : Deep Learning for Code	2022
<u>IntEx-SemPar at EMNLP 2020</u> : Interactive and Executable Semantic Parsing Workshop	2020
<u>NLI at ACL 2020</u> : Natural Language Interface Workshop	2020
<u>SRW at ACL 2019</u> : Student Research Workshop	2019