

ZORA (ZHIRUO) WANG

✉ zhiruow@cs.cmu.edu | 🌐 zorazrw.github.io | 📖 google scholar | 🐙 github

EDUCATION

Carnegie Mellon University, Language Technologies Institute **Pittsburgh, PA**
PhD of Language Technologies Aug 2023 - Present
Advisor: Daniel Fried, Graham Neubig

Carnegie Mellon University, Language Technologies Institute **Pittsburgh, PA**
Master of Language Technologies (MLT), GPA: 4.02 / 4.3 Aug 2021 - Aug 2023
Advisors: Graham Neubig, Eric Nyberg

Beijing Normal University, School of Mathematical Sciences **Beijing, China**
B.S. in Mathematics & Applied Mathematics, major GPA: 3.98 / 4.0 Sep 2016 - Jun 2020
◦ Exchange, Chinese University of Hong Kong, full scholarship Spring 2019
◦ Summer School, The College of William & Mary, full scholarship Summer 2017

EXPERIENCE

Microsoft, Github Next Team **Remote**
Consultant, Host: Eddie Aftandilian Aug 2023 - Dec 2023
◦ Consult with code language model training and robust open-domain evaluation techniques

Amazon Alexa AI, Conversational Modeling Team **Sunnyvale, CA**
Research Intern, Manager: Dilek Hakkani-Tur May 2022 - Aug 2022
◦ Proposed a model self-talk schema for knowledge-guided dialogue generation, further bootstrapping large-scale data for adaptive training, under a special (book) domain and low-resource setting

Microsoft Research (Asia), Data Knowledge Intelligence Group **Beijing, China**
Research Intern → *Assistant Researcher*, Manager: Dongmei Zhang Aug 2020 - Aug 2021
◦ Designed and pre-trained a structure-aware table foundation model, *TUTA*, with tree-based masked attention and adaptive position encodings, and shaped it into Microsoft *Word* and *Excel*
◦ Built *HiTab* benchmark, supporting table QA and data analysis for complex-structured tables

SELECTED PREPRINTS

(* indicates equal contribution, ^m indicates mentorship)

[1] *Agent Workflow Memory*

Zora Zhiruo Wang, Jiayuan Mao, Daniel Fried, Graham Neubig
the International Conference on Learning Representations (ICLR), 2025 (under review)

[2] *CodeRAG-Bench: Can Retrieval Augment Code Generation?*

Zora Zhiruo Wang*, Akari Asai*, Xiyan Yu, Frank F. Xu, Yiqing Xie, Graham Neubig, Daniel Fried
the North American Chapter of the Association for Computational Linguistics (NAACL), 2025 (under review)

[3] *Learning to Filter Context for Retrieval-Augmented Generation*

Zora Zhiruo Wang, Jun Araki, Zhengbao Jiang, Md Rizwan Parvez, Graham Neubig

SELECTED PUBLICATIONS

[4] *What Are Tools Anyway? A Survey from the Language Model Perspective*

Zora Zhiruo Wang, Zhoujun Cheng, Hao Zhu, Daniel Fried, Graham Neubig
the Conference on Language Modeling (CoLM), 2024

- [5] *TroVE: Inducing A Verifiable and Efficient Toolbox for Programmatic Tasks*
Zora Zhiruo Wang, Graham Neubig, Daniel Fried
the International Conference on Machine Learning (ICML), 2024
- [6] *ECCO: Can We Improve Model-Generated Code Efficiency Without Sacrificing Functional Correctness?*
Siddhant Waghjale, Vishruth Veerendranath, **Zora Zhiruo Wang**^m, Daniel Fried
the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2024
- [7] *API-Assisted Code Generation for Question Answering on Varied Table Structured*
Yihan Can, Shuyi Chen, Ryan Liu, **Zora Zhiruo Wang**^m, Daniel Fried
the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023
- [8] *StarCoder: may the source be with you!*
Raymond Li, Loubna Ben Allal, ..., **Zora Zhiruo Wang**, ..., Leandro von Werra, Harm de Vries
the Transactions on Machine Learning Research (TMLR)
- [9] *Execution-Based Evaluation for Open-Domain Code Generation*
Zora Zhiruo Wang, Shuyan Zhou, Daniel Fried, Graham Neubig
the Conference on Empirical Methods in Natural Language Processing (EMNLP), Findings, 2023
- [10] *MCoNaLa: A Benchmark for Code Generation from Multiple Natural Languages*
Zora Zhiruo Wang, Grace Cuenca, Frank F. Xu, Shuyan Zhou, Graham Neubig
the European Chapter of the Association for Computational Linguistics (EACL), 2023
- [11] *DocPrompting: Generating Code By Retrieving the Docs*
Shuyan Zhou, Uri Alon, Frank F. Xu, **Zora Zhiruo Wang**, Zhengbao Jiang, Graham Neubig
the International Conference on Learning Representations (ICLR), 2023
- [12] *HiTab: A Hierarchical Table Dataset for Question Answering and Natural Language Generation*
Zhoujun Cheng*, Haoyu Dong*, **Zora Zhiruo Wang***, R. Jia, J. Guo, Y. Gao, S. Han, J. Lou, D. Zhang
the Association for Computational Linguistics (ACL), 2022

COMMUNITY SERVICES

Workshop & Tutorial Organization

- SIGIR Tutorial Organizer & Presenter: “Large Language Models for Tabular Data” Jul 2024
- Head Organizer of the 1st CMU Agent Workshop May 2024
- Head Organizer of CMU Student Research Symposium Aug 2022

DEI Services

- Member & Student Representative of CMU LTI DEI Committee 2024
- Chair of Faculty Hiring Student Committee; LTI, CMU 2024

Paper Review

- Program committee for NeurIPS, ICLR; ACL, EMNLP, NAACL, EACL; AAAI

TEACHING EXPERIENCE

Teaching Assistant

- Neural Code Generation, Spring 2024: lead student discussion, advise students research projects
- Advanced NLP, Spring & Fall 2023: head TA; prepare assignments, advise students research projects
- Data Science Seminar, Fall 2021: organize research paper discussion, advise students research proposal

Guest Lectures

- “Evaluation & Benchmarks for Code Generation” at Neural Code Generation and Advanced NLP
- “Tool Use and Language Agents” at Advanced NLP

Student Mentoring

Xinyue Liu	CMU Master (MCDS)	Fall 2024 – Present
Vishwa Shah	CMU Master (MIIS)	Fall 2024 – Present
Yibo Peng	CMU Master (ECE)	Spring 2024 – Present
Vishruth Veerendranath ^[6]	CMU Master (MIIS)	Fall 2023 – Present
Eduardo Treviño	CMU Master (HCI)	Summer – Fall 2024
Hugo Contant	CMU Undergrad	Summer – Fall 2024
Siddhant Waghjale ^[6]	CMU Master (MIIS)	Fall 2023 – Spring 2024
Afreen Shaikh	CMU Master (MIIS)	Fall 2023 – Spring 2024
Jennifer Hsia	CMU PhD (MLD)	Fall 2023
Ryan Liu ^[7]	CMU Master (MSML) → Princeton PhD	Spring 2023
Shuyi Chen ^[7]	CMU Master (MISM) → CMU PhD (ML&OR)	Spring 2023
Yihan Cao ^[7]	CMU Master (MISM) → LinkedIn MLE	Spring 2023
Ethan Chen	CMU Master (MSAII) → SJTU PhD	Spring 2023

AWARDS & HONORS

CMU Presidential Fellowship	2023
China National Scholarship recipient	2019

INVITED TALKS

“Agent Workflow Memory”	
CAMEL-AI	Sep 2024
“What Are Tools Anyway?”	
FLAME (Foundation and LAnguage Model) seminar	May 2024
“Tool using, learning, and making with LLMs”	
LLM as Agent Seminar	Feb 2024
“Execution-based Open-Domain Coding”	
JetBrains Research Team	Apr 2023

MEDIA COVERAGE

MarkTechPost	Sep 2024
Agent Workflow Memory (AWM): An AI Method for Improving the Adaptability and Efficiency of Web Navigation Agents	
MarkTechPost	Aug 2024
A Reproducible AI Benchmark for Evaluating Program Efficiency via Two Paradigms	
MarkTechPost	Jul 2023
Google Research Introduces SPAE: An AutoEncoder For Multimodal Generation With Frozen Large Language Models (LLMs)	
SyncedReview	May 2023
‘May the Source Be With You!’ –BigCode’s Open-Access StarCoder Outperforms All Existing Open Code LLMs	