

RUOYAO WANG

Tel: 734-272-2321 • E-mail: ruoyao wang@arizona.edu • Website: ruoyao-wang.github.io

EDUCATION

University of Arizona , Tucson, AZ Ph.D., Information Science	2021-08 to 2025-05 (Expected) Advisor: Peter Jansen
University of Michigan , Ann Arbor, MI M.S., Computer Science	2018-09 to 2020-05 Advisor: Rada Mihalcea
Fudan University , Shanghai, China B.E., Microelectronics Science and Engineering	2014-09 to 2018-06

WORK EXPERIENCE

Microsoft Research, Montreal	Research Intern	2024.5-2024.8
Description: Use LLMs to code world models by interacting with simulation environments.		
University of Arizona	Teaching Assistant	2022.9 - 2024.8
Courses: Machine Learning (Graduate), Statistical Foundations (Undergrad)		
University of Michigan	Research Assistant	2020.7 - 2021.7
Description: I involved in constructing a video question answering dataset. I worked on data generation algorithm, baseline training, and paper writing. Related papers were published in ACL and LREC.		

TECHNICAL SKILLS

General Research Areas: Natural Language Processing, Machine Learning, Data Science

Specific Project Expertise: Virtual Environments, Text Games, Code Generation

Programming: Python, C++

Software: Pytorch, Huggingface Transformers, Numpy, Scipy, Sklearn

PUBLICATIONS

I published **6 first-author papers** on conferences and journals such as **ACL**, **EMNLP**, and **EACL**.

1. **Ruoyao Wang**, Graham Todd, Ziang Xiao, Xingdi Yuan, Marc-Alexandre Côté, Peter Clark, and Peter Jansen. Can language models serve as text-based world simulators? In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL)*, August 2024

Description: The state transition ability of large language models are evaluated in the context of text games. Experiments show that with only approximately 60% of transition prediction accuracy, large language models are not reliable world simulators yet.

2. **Ruoyao Wang**, Graham Todd, Eric Yuan, Ziang Xiao, Marc-Alexandre Côté, and Peter Jansen. Bytesized32: A corpus and challenge task for generating task-specific world models expressed as text games. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023

Description: The capacity of large language models to generate explicit, interpretable, and interactive world models of scientific and common-sense reasoning tasks is investigated. A corpus of 32 highly-templated Python text games, totalling 20k lines of code, is created. We had GPT-4 generating 96 text games using one of the 32 games as an example, and designed 4 automatic metrics evaluating the generated games.

3. **Ruoyao Wang**, Peter Jansen, Marc-Alexandre Côté, and Prithviraj Ammanabrolu. ScienceWorld: Is your agent smarter than a 5th grader? In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, December 2022

Description: ScienceWorld is designed as an interactive benchmark for scientific and common-sense reasoning. It includes 30 text games based on 10 common elementary school science topics. Instead of answering exam questions, agents show their understanding of the science concepts by conducting related experiments in

the game. We empirically show that agents need to be grounded in interactive environments to achieve such reasoning capabilities.

4. **Ruoyao Wang**, Peter Jansen, Marc-Alexandre Côté, and Prithviraj Ammanabrolu. Behavior cloned transformers are neurosymbolic reasoners. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, May 2023

Description: This work shows that interactive text games agents can be augmented with information from symbolic modules, much like humans use tools like calculators and GPS systems to assist with arithmetic and navigation.

5. **Ruoyao Wang** and Peter Jansen. Self-supervised behavior cloned transformers are path crawlers for text games. In *Findings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP Findings)*, 2023

Description: To overcome the difficulty of human game playthrough collection for behavior cloning transformers training in the context of text games, We propose a self-supervised data generation method that conduct path crawling and grouping to automatically collect generalizable text game agent training data.

6. **Ruoyao Wang**, Zhenghan Fang, Jiaqi Gu, Yi Guo, Shicong Zhou, Yuanyuan Wang, Cai Chang, and Jinhua Yu. High-resolution image reconstruction for portable ultrasound imaging devices. *EURASIP Journal on Advances in Signal Processing*, 2019(1):56, Nov 2019

Description: I built a super-resolution model to increase the resolution of portable ultrasound devices.

7. Santiago Castro, **Ruoyao Wang**, Pingxuan Huang, Ian Stewart, Oana Ignat, Nan Liu, Jonathan Stroud, and Rada Mihalcea. FIBER: Fill-in-the-blanks as a challenging video understanding evaluation framework. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*, May 2022

Description: FIBER is a video-based fill-in-the-blanks (Cloze) text task for video understanding evaluation that tries to remove the evaluation biases in multiple-choice video question answering tasks and overcome the difficulty in evaluating video captioning tasks.

8. Santiago Castro, Mahmoud Azab, Jonathan Stroud, Cristina Noujaim, **Ruoyao Wang**, Jia Deng, and Rada Mihalcea. LifeQA: A real-life dataset for video question answering. In *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC)*, May 2020

Description: LifeQA is a benchmark dataset for video question answering that focuses on day-to-day real-life situations.

9. Meaghan Emery-Wetherell and **Ruoyao Wang**. How to use academic and digital fingerprints to catch and eliminate contract cheating during online multiple-choice examinations: a case study. *Assessment & Evaluation in Higher Education*, 0(0):1–16, 2023

Description: This work introduces a case study on detecting and reducing cheating in online exams.

10. Jiaqi Gu, **Ruoyao Wang**, Jian Wang, Jinmei Lai, and Qinghua Duan. Remote embedded simulation system for sw/hw co-design based on dynamic partial reconfiguration. In *2017 IEEE 12th International Conference on ASIC (ASICON)*, pages 402–405, 2017

Description: A software/hardware co-design platform was built using FPGA dynamic partial reconfiguration.

ACADEMIC SERVICE

Beijing Academy of Artificial Intelligence, 2024

Invited Talk

Arizona State University Machine Learning Day, 2024

Invited Talk

EMNLP 2024

Reviewer

NAACL 2024

Reviewer

NLPCC 2024

Reviewer

WordPlay: When Language Meets Games @ ACL 2024

Reviewer