

## EDUCATION

## University of Illinois Urbana-Champaign

Urbana, IL

*Ph.D. in Computer Science*2024 - 2029 (*expected*)

- GPA: 3.94/4.00.
- Advisor: Prof. Jiaxuan You
- Research area: Large Language Model, Language Agent, Multimodal Machine Learning, Graph-based Machine Learning

## Carnegie Mellon University

Pittsburgh, PA

*M.S. in Intelligent Information Systems*

2022 - 2024

- GPA: 4.12/4.33.
- A+ Coursework: Natural Language Processing, Advanced Topics in Multimodal Machine Learning, etc

## Zhejiang University

Hangzhou, China

*B.Eng. in Computer Science*

2018-2022

- GPA: 3.96/4.00. Rank 7/134.
- Outstanding Graduates. Provincial Government Scholarship 2019.
- A+ Coursework: Discrete Mathematics and Application, Object-Oriented Programming, Computer Networks, etc

SELECTED  
PUBLICATION

1. **Haofei Yu**<sup>†</sup>, Zhengyang Qi<sup>†</sup>, Yining Zhao<sup>†</sup>, Kolby Nottingham, Keyang Xuan, Bodhisattwa Prasad Majumder, Hao Zhu, Paul Pu Liang, Jiaxuan You. Sotopia-RL: Reward Design for Social Intelligence. *Preprint 2025*. [\[Paper\]](#) [\[Code\]](#) [\[Model\]](#) [\[Data\]](#)
2. Xiaojun Shan<sup>†</sup>, Qi Cao<sup>†</sup>, Xing Han<sup>†</sup>, **Haofei Yu**<sup>†</sup>, Paul Pu Liang. MINT: Multimodal Instruction Tuning with Multimodal Interaction Grouping. *Preprint 2025*. [\[Paper\]](#)
3. **Haofei Yu**<sup>†</sup>, Zhaochen Hong<sup>†</sup>, Zirui Cheng<sup>†</sup>, Kunlun Zhu<sup>†</sup>, Keyang Xuan, Jinwei Yao, Tao Feng, Jiaxuan You. ResearchTown: Simulator of Human Research Community. *ICML 2025*. [\[Paper\]](#) [\[Code\]](#) [\[Data\]](#)
4. **Haofei Yu**<sup>†</sup>, Zhengyang Qi<sup>†</sup>, Lawrence Jang<sup>†</sup>, Ruslan Salakhutdinov, Louis-Philippe Morency, Paul Pu Liang. MMOE: Enhancing Multimodal Models with Mixtures of Multimodal Interaction Experts. *EMNLP 2024*. [\[Paper\]](#) [\[Code\]](#)
5. **Haofei Yu**<sup>†</sup>, Keyang Xuan<sup>†</sup>, Fenghai Li<sup>†</sup>, Kunlun Zhu, Zijie Lei, Jiaxun Zhang, Ziheng Qi, Kyle Richardson, Jiaxuan You. TinyScientist: An Interactive, Extensible, and Controllable Framework for Building Research Agents. *EMNLP 2025 Demo Track*. [\[Paper\]](#)
6. **Haofei Yu**<sup>†</sup>, Cunxiang Wang<sup>†</sup>, Yue Zhang, Wei Bi. TRAMS: Training-free Memory Selection for Long-range Language Modeling. *Findings of EMNLP 2023*. [\[Paper\]](#) [\[Code\]](#)
7. Ruiyi Wang<sup>†</sup>, **Haofei Yu**<sup>†</sup>, Wenxin Zhang<sup>†</sup>, Zhengyang Qi<sup>†</sup>, Maarten Sap, Graham Neubig, Yonatan Bisk, Hao Zhu. SOTOPIA- $\pi$ : Interactive Learning of Socially Intelligent Language Agents. *ACL 2024*. [\[Paper\]](#) [\[Code\]](#) [\[Data\]](#) [\[Model\]](#) [\[Media\]](#)
8. Zhaochen Hong<sup>†</sup>, **Haofei Yu**<sup>†</sup>, Jiaxuan You. ConsistencyChecker: Tree-based Evaluation of LLM Generalization Capabilities *ACL 2025*. [\[Paper\]](#) [\[Code\]](#)
9. Cunxiang Wang<sup>†</sup>, **Haofei Yu**<sup>†</sup>, Yue Zhang. RFiD: Towards Rational Fusion-in-Decoder for Open-Domain Question Answering. *Findings of ACL 2023*. [\[Paper\]](#) [\[Code\]](#)
10. Yijie Hao, **Haofei Yu**, Jiaxuan You. Beyond Facts: Evaluating Intent Hallucination in Large Language Models *ACL 2025*. [\[Paper\]](#)
11. Chiyu Song<sup>†</sup>, Hongliang He<sup>†</sup>, **Haofei Yu**, Pengfei Fang, Leyang Cui, Zhenzhong Lan. Uni-Encoder: A Fast and Accurate Response Selection Paradigm for Generation-Based Dialogue Systems. *Findings of ACL 2023*. [\[Paper\]](#) [\[Code\]](#)

12. Pengrui Han†, Peiyang Song†, **Haofei Yu**, Jiaxuan You. In-Context Learning May Not Elicit Trustworthy Reasoning: A-Not-B Errors in Pretrained Language Models. *Findings of EMNLP 2024*. [\[Paper\]](#) [\[Code\]](#)
13. Kunlun Zhu, Jiaxun Zhang, Ziheng Qi, Nuoxing Shang, Zijia Liu, Peixuan Han, Yue Su, **Haofei Yu**, Jiaxuan You. SafeScientist: Toward Risk-Aware Scientific Discoveries by LLM Agents. *EMNLP 2025*. [\[Paper\]](#) [\[Code\]](#)
14. Xuhui Zhou†, Hao Zhu†, Leena Mathur, Ruohong Zhang, **Haofei Yu**, Zhengyang Qi, Louis-Philippe Morency, Yonatan Bisk, Daniel Fried, Graham Neubig, Maarten Sap. Sotopia: Interactive evaluation for social intelligence in language agents. *ICLR 2024 (top 5% submission, Spotlight)*. [\[Paper\]](#) [\[Code\]](#)
15. Leonie Weissweiler†, Valentin Hofmann†, Anjali Kantharuban†, Anna Cai, Ritam Dutt, Amey Hengle, Anubha Kabra, Atharva Kulkarni, Abhishek Vijayakumar, **Haofei Yu**, Hinrich Schütze, Kemal Oflazer, David R Mortensen. Counting the Bugs in ChatGPT’s Wugs: A Multilingual Investigation into the Morphological Capabilities of a Large Language Model. *EMNLP 2023*. [\[Paper\]](#) [\[Code\]](#)
16. Paul Pu Liang†, Akshay Goindani†, Talha Chafekar, Leena Mathur, **Haofei Yu**, Ruslan Salakhutdinov, Louis-Philippe Morency. HEMM: Holistic Evaluation of Multimodal Foundation Models. *NeurIPS 2024 Dataset and Benchmark Track*. [\[Paper\]](#) [\[Code\]](#)

## OPEN-SOURCE SOFTWARE

### Research Town: Simulator for Human Research Community

Project Leader

🔗 [research-town](#) ★ 177 📄 23

[\[Python package\]](#)

### TinyScientist: A Minimal Framework for Building Research Agents

Project Leader

🔗 [tiny-scientist](#) ★ 106 📄 13

[\[Python package\]](#)

### Sotopia: An Open-ended Social Learning Environment

Core contributor

🔗 [sotopia](#) ★ 246 📄 37

[\[Python package\]](#)

### TinyChat: A Lightweight Platform for Social Simulations

Core contributor

🔗 [tiny-chat](#) ★ 11 📄 2

[\[Python package\]](#)

## INDUSTRIAL EXPERIENCE

### Allen Institute for AI | Seattle, WA

2025.05 - 2026.08

- Defined the automatic SOTA discovery task as a link discovery task when framing the HuggingFace platform as the artifact graph.
- Proposed a two-stage framework (predict-and-verify) for scalable auto discovery on the HuggingFace platform.

### MIT-IBM Watson AI Lab | Boston, MA

2024.05 - 2024.08

- Conducted a comprehensive evaluation on open-sourced and close-sourced LLMs on SWE-bench.
- Proposed a fine-grained analysis on the skill difference between open-sourced LLMs and close-sourced ones.

### Apple AIML | Seattle, WA

2023.05 - 2023.08

- Delivered an LLM-driven hierarchical prompting system including one document retriever, multiple summarizers, and one QA model to disambiguate and accurately respond to challenging real-world Siri user queries.
- Enhanced satisfaction on internal user data and chosen to present to Siri Senior Director *Robby Walker* (top 10 in SII).

### Tencent AI Lab | Shenzhen, China

2022.02 - 2022.08

- Designed a training-free memory selection metric in Transformer-XL, gaining 0.19 perplexity drop on WikiText-103.
- Proposed a diffusion-based approach for NER, achieving results comparable to SpanBERT on CoNLL03 and OntoNotes.

ACADEMIC EXPERIENCE	<b>University of Illinois Urbana-Champaign</b>   Urbana, IL	2024.08 - Now
	<ul style="list-style-type: none"> <li>• Proposed a multi-agent environment for research community simulation.</li> <li>• Built a minimal and extensible platform for building research agents.</li> </ul>	
	<b>Carnegie Mellon University</b>   Pittsburgh, PA	2022.08 - 2024.05
	<ul style="list-style-type: none"> <li>• Proposed contrastive ranking loss for multi-task training on code generation, enhancing pass@100 by 6.65% on ODEX.</li> <li>• Implemented LLM training on the SOTOPIA benchmark using ReST, reaching a 37.8% boost in social goal achievement.</li> <li>• Introduced multimodal MoE to handle 3 self-defined multimodal interactions, achieving a state-of-the-art MUsTARD performance.</li> <li>• Analyzed LLM's morphological ability in 4 languages and compared few-shot prompting with sub-word-level models.</li> </ul>	
	<b>Westlake University</b>   Hangzhou, China	2021.02 - 2022.02
	<ul style="list-style-type: none"> <li>• Proposed Rational Fusion-in-Decoder, augmented with passage rationale training, achieving a 2.8% increase in accuracy on NaturalQuestions, a 0.9% rise on TriviaQA, and a 15.4% enhancement in understanding across multiple documents</li> <li>• Proposed a SoTA response selection model with a 2.9% R10@1 improvement and 4x faster inference on Ubuntu-v2.</li> </ul>	
TEACHING EXPERIENCE	<b>TA for UIUC CS-512</b>   Data Mining Principles <a href="#">[Material]</a>	2025 (Fall)
	<b>TA for CMU 11-877</b>   Advanced Topics in Multimodal ML <a href="#">[Material]</a>	2024 (Spring)
	<b>TA for CMU 11-777</b>   Multimodal Machine Learning <a href="#">[Material]</a>	2023 (Fall)
AWARDS	<b>Outstanding Reviewer</b>   EMNLP 2024	2024.11