

# RONGWU XU

## PERSONAL

---

PRONOUNS: He/Him/His  
LOCATION: FIT building, Tsinghua University, Beijing, China  
E-MAIL: [0xrwxu@gmail.com](mailto:0xrwxu@gmail.com) (primary) or [xrw22@mails.tsinghua.edu.cn](mailto:xrw22@mails.tsinghua.edu.cn)  
HOMEPAGE: [rongwuxu.site](http://rongwuxu.site)

## RESEARCH

---




My primary interests lie in the fields of natural language processing (NLP) and trustworthy AI, as well as their intersections, *e.g.*, factuality, safety, and ethics in *up-to-date* NLP systems. Considering these intersections, my current exploration focuses on:

- **Investigating and Enhancing Trust in LLMs.** I identify risks present in large language models (LLMs), including misinformation, safety, ethical concerns. My ultimate goal is to contribute to more trustworthy NLP systems that align with human values.
- **Utilizing NLP in Public Discourse and Moderation.** I apply the advancements in NLP to analyze public opinions. This involves understanding, interpreting, and managing the vast array of information expressed in digital platforms.

**My vision is to harness the power of AI and NLP to create trustworthy and human-centric digital environments.**

## EDUCATION

---

Jun. 2025	MRES IN COMPUTER SCIENCE <b>Tsinghua University</b>  Beijing, China Advisor: Prof. <a href="#">Wei Xu</a>
Aug. 2022	Graduate Student at IIS (Headed by <a href="#">Andrew C. Yao</a> , the Turing award laureate'2000)
Jun. 2022	BENG IN COMPUTER SCIENCE <b>Tsinghua University</b>  Beijing, China
Sept. 2018	Bachelor Student at Department of Computer Science and Technology (DCST)
Jun. 2018	<b>Beijing No.4 High School</b>  Beijing, China
Sept. 2015	High School Student at the Class of Olympiad (Chemistry)

## PUBLICATIONS AND MANUSCRIPTS

---

11. Preemptive Answer “Attacks” on Chain-of-Thought Reasoning  
**Rongwu Xu\***, Zehan Qi\*, Wei Xu  
*Preprint*

10. Understandable and Singable Musical Lyrics Translation  
Jinhan Li, Zhuorui Ye, **Rongwu Xu**  
*Preprint*

9. Walking in Others’ Shoes: How Perspective-Taking Guides LLMs in Reducing Toxicity and Bias

**Rongwu Xu**, Zi'an Zhou, Tianwei Zhang, Zehan Qi, Su Yao, Ke Xu, Wei Xu, Han Qiu  
*Preprint*

8. Knowledge Conflicts for LLMs: A Survey

**Rongwu Xu\***, Zehan Qi\*, Cunxiang Wang, Hongru Wang, Yue Zhang, Wei Xu  
*arXiv Preprint*

7. The Earth is Flat because...: Investigating LLMs' Belief towards Misinformation via Persuasive Conversation

**Rongwu Xu**, Brian S. Lin, Shujian Yang, Tianqi Zhang, Weiyang Shi, Tianwei Zhang, Zhixuan Fang, Wei Xu, Han Qiu  
*arXiv Preprint*

6. Exploring Chinese Humor Generation: A Study on Two-part Allegorical Sayings

**Rongwu Xu**

In *International Joint Conference on Neural Networks (IJCNN)*, 2024

5. Tempo: Confidentiality Preservation in Cloud-based Neural Network Training

**Rongwu Xu** and Zhixuan Fang

In *International Joint Conference on Neural Networks (IJCNN)*, 2024

4. LSync: A Universal Timeline-synchronizing Solution for Live Streaming

Yifan Xu\*, Fan Dang\*, **Rongwu Xu**, Xinlei Chen, Yunhao Liu

In *IEEE/ACM Transactions on Networking (TON)*, 2024

3. MISO: Legacy-compatible Privacy-preserving Single Sign-on using Trusted Execution Environments

**Rongwu Xu**, Sen Yang, Fan Zhang, Zhixuan Fang

In *IEEE European Symposium on Security and Privacy (EuroS&P)*, 2023

2. LSync: A Universal Event-synchronizing Solution for Live Streaming

Yifan Xu, Fan Dang, **Rongwu Xu**, Xinlei Chen, Yunhao Liu

In *IEEE Conference on Computer Communications (INFOCOM)*, 2022

1. LifeRec: A Mobile App for Lifelog Recording and Ubiquitous Recommendation

Jiayu Li, Hantian Zhang\*, Zhiyu He\*, **Rongwu Xu\***, Pingfei Wu\*, Min Zhang, Yiqun Liu, Shaoping Ma

In *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR)*, 2022

(\* equal contribution)

## HONORS AND AWARDS

---

2023	Overall Excellence Scholarship@Tsinghua University
2020	Technological Innovation Scholarship@Tsinghua University
2020	1 <sup>st</sup> in "Youth in Action" Social Practice@Tsinghua University
2019	Tsinghua-Panasonic Scholarship@Tsinghua University
2018	Outstanding Volunteers in Beijing
2017	2 <sup>nd</sup> Prize (Preliminary) in Chinese Chemistry Olympiad (CChO)
2017	1 <sup>st</sup> Prize in Chinese Chemistry Olympiad (Beijing Regional Qualifiers)

## TALKS AND PRESENTATIONS

---

Apr. 2024	Investigating large language models' beliefs and behaviors under misinformation	Propaganda film@IIIS, Tsinghua
May. 2023	Privacy-preserving authentication	Oral report@EuroS&P conference

## EXPERIENCES

---

### Mentoring

It is my pleasure to collaborate with the following brilliant students:

Apr. 2024 -	Yishuo Cai	Undergrad, SE@Central South Univ.
Mar. 2024 -	Yishu Yin	Undergrad, CS@Tsinghua Univ.
Mar. 2024 -	Priscilla Chen	Undergrad, EECS@UC Berkeley
Mar. 2024 -	Xinghan Li	Undergrad, IIIS@Tsinghua Univ.
Mar. 2024 -	Xuan Qi	Undergrad, IIIS@Tsinghua Univ.
Dec. 2023 -	Zi'an Zhou	Undergrad, Zhili College@Tsinghua Univ.
Sep. 2023 - Feb. 2024	Shujian Yang	Master, SPEIT@Shanghai Jiao Tong Univ.
Jun. 2023 -	Tianqi Zhang	Undergrad, CS@Tsinghua Univ.
Jun. 2023 -	Brian S. Lin	Undergrad, CS@Tsinghua Univ.
Oct. 2021 - Jul. 2022	Xingyu Dang	Undergrad, IIIS@Tsinghua Univ.

### Teaching

Spring 2024	<b>Teaching Assistant &amp; Organizer</b> , Tsinghua University <i>Introduction of Large Language Model Applications</i> Co-designed labs, organized the curriculum and assisted labs in class.
Fall 2023	<b>Teaching Assistant</b> , Tsinghua University <i>Operating System and Distributed System</i> Held discussion and office hours, graded exams, assignments and projects.
Spring 2022	<b>Teaching Assistant</b> , Tsinghua University <i>Distributed System and Blockchain</i> Held discussion and office hours, graded exams, assignments and projects.

### Exchanging

Apr. 2021 - Oct. 2022	<b>Research Assistant (Remote)</b> , Duke University <i>Granted summer overseas internship (undergrad) by Tsinghua</i> Conducted research in privacy-preserving authentication. Host: Prof. <a href="#">Fan Zhang</a>
-----------------------	--

### Internship

Apr. 2024 -	<b>TongYi Vision Intelligence Lab, Alibaba Inc.</b> , Beijing, China TBA TBA
Dec. 2022 - Jan. 2023	<b>Shanghai Qi Zhi Institute</b> , Shanghai, China Conducted research in decentralized finance (DeFi). Worked on MEV arbitrage forecasting algorithms using graph neural networks (GNNs). Mentor: Prof. <a href="#">Zhixuan Fang</a>

## SKILLS AND EXPERTISE

---

- **Research:** Experienced in deep learning/data analysis, also ability with computer systems/applied cryptography/software engineering.
- **Programming Language:** Proficient in C++/Go/Python/Java/JavaScript/L<sup>A</sup>T<sub>E</sub>X, also ability with Verilog/System Verilog/ASM/Rust/P4/HTML/Matlab.
- **Technological:** Proficient in Pytorch/NumPy/Matplotlib/Git/Linux OS/Markdown/, also ability with Docker/Wireshark/Microsoft Office.
- **Other Expertise:** Good communication skills, love collaborating with people, experi-

enced in mentoring students, works well in a team.

## SERVICES AND MEMBERSHIPS

---

### Academic

- Mar. 2024 - Current **Member**, International Neural Network Society (INNS)
- Mar. 2024 - Current **Member**, Institute of Electrical and Electronics Engineers (IEEE)
- Mar. 2024 - Current **Member**, ACL SIGSEC, Association for Computational Linguistics

### Societal

- Jun. 2024 - Current **2024 Graduate Freshman Assistant**, IIS, Tsinghua University
- Apr. 2024 - Current **Social Practice Stewardship**, IIS, Tsinghua University
- Nov. 2023 - Current **Member**, IIS Students Congress, Tsinghua University