Ruiling Xu

Tel: +86-19382273552 | Email: ruiling3@illinois.edu

EDUCATION

Zhejiang University (ZJU-UIUC Institute) | Haining, China — Aug 2022 – Jun 2026
- BEng. Electronic and Computer Engineering, GPA: 4.09/4.3

University of Illinois at Urbana-Champaign | Champaign, IL — Aug 2022 – Jun 2026
- BSc. Computer Engineering, GPA: 3.83/4
- **Computer skills:** Python, C/C++, CSS, JavaScript, TypeScript, HTML, MongoDB, RISC-V, LC3, Shell, Make, SQL
- **Tools:** Pytorch, HuggingFace, Git, Linux, Docker, Proteus

PUBLICATION

1. **Ruiling Xu**, Yifan Zhang, Qingyun Wang, Carl Edwards, Heng Ji*. oMeBench: Towards Robust Benchmarking of LLMs in Organic Mechanism Elucidation and Reasoning. ***ACL 2025*** (ARR Oct submission 3.33/5) https://arxiv.org/abs/2510.07731
2. Xingguo. Guo, Yaxin. Li, …, **Ruiling. Xu.**, ….& Bin. Hu*. Toward Engineering AGI: Benchmarking the Engineering Design Capabilities of LLMs. ***NeurIPS 2025*** (accepted) https://dblp.org/rec/journals/corr/abs-2509-16204.html
3. Qiyuan Wang, **Ruiling Xu**, Shibo He, Randall Berry, Meng Zhang*. Unlearning Incentivizes Learning under Privacy Risk. ***WWW'25*** *https://doi.org/10.1145/*
4. Zhiting Fan, Ruizhe Chen, **Ruiling Xu**, Zuozhu Liu*. BiasAlert: A Plug-and-play Tool for Social Bias Detection in LLMs. ***EMNLP 2024. 14778-14790.*** *https://aclanthology.org/2024.emnlp-main.820/*
5. Ruizhe Chen, Jianfei Yang, Huimin Xiong, **Ruiling Xu**, Yang Feng, Jian Wu, Zuozhu Liu*. Cross-center Model Adaptive Tooth Segmentation. ***Medical Image Analysis 101 (2025) 103443.*** *https://doi.org/10.1016/j.media.2024.103443*

RESEARCH

oMeBench: Towards Robust Benchmarking of LLMs in Organic Mechanism Elucidation and Reasoning
(BLENDER Lab at UIUC) — Champaign, IL
Research Intern, Supervisor: Prof. Heng Ji & Prof. Qingyun Wang — Apr 2025 – Oct 2025
- Built the first large-scale, expert-curated dataset of organic reaction mechanisms (oMe-Gold/Template/Silver), comprising 10k+ mechanistic steps with rich annotations (reaction types, intermediates, SMILES, difficulty).
- Proposed oMeS, a dynamic evaluation framework combining weighted Needleman–Wunsch alignment and Tanimoto similarity for partial credit, enabling fine-grained analysis of LLMs' actual mechanistic reasoning across 4 metrics.
- Analyzed performance of 10+ LLMs, revealing systematic failure patterns on domain-specific reaction reasoning.
- Conducted standard and COT fine-tuning on compact models, achieving consistent gains and increased performance by 50% over the leading closed-source model.

Unlearning Incentivizes Learning Under Privacy Risk (Nexus Lab at ZJU-UIUC Institute) — Haining, China
Research Assistant, Supervisor: Prof. Meng Zhang — Jul 2024 – Oct 2024
- Designed contracts to evaluate platform profitability impacts of enabling vs. disabling unlearning in federated learning.
- Modeled privacy-sensitive and risk-averse users using principal-agent theory and under federated learning scenarios, derived optimal contracts via backward induction and convex optimization methods (FOA, CVX solver), which were further validated by extensive numerical simulation.
- Validated the mechanism through survey data collected by WPP Media and simulations, showing that supporting unlearning increases both user participation and platform profitability in high-sensitivity settings.

BiasAlert: A Plug-and-play Tool For Social Bias Detection in LLMs (ZJU-UIUC Institute) — Haining, China
Research Assistant, Supervisor: Prof. Zuozhu Liu — Apr 2024 – Jul 2024
- Implemented a RAG-based plug-and-play tool to reliably detect social bias in LLM's open-text generations, integrating external knowledge retrieval with internal reasoning to improve adaptability and interpretability.
- Constructed a bias retrieval database with 3.9k+ data across 7 bias types, crafted an instruction-following dataset and implemented prompt engineering tricks to enhance internal reasoning abilities.
- Validated that BiasAlert achieves ~80% bias mitigation rate on multiple benchmarks and an average of 1.4 sec (dual RTX 3090s) to monitor a single bias, outperforming SOTA bias detection tools (Llama-Guard, LLMs-as-judge, etc).
- Presented research poster on the 2024 Conference on Empirical Methods in Natural Language Processing, Miami, 2024.

Cross-center Model Adaptive Tooth (CMAT) Segmentation (ZJU-UIUC Institute) — Haining, China
Research Assistant, Supervisor: Prof. Zuozhu Liu — Mar 2023 – Jul 2024
- Proposed a framework with 3 modules: tooth prototype alignment, progressive pseudo-label transfer, and tooth prior regularization information maximization for cross-center model adaptation without source data or extra annotated data.
- Constructed two cross-center tooth segmentation dataset, CrossTooth and AbnTooth, from five medical centers.
- Achieved superior segement performance, improving average mIoU by 7.5% on CrossTooth with 8.6% in multi-source, 4.8% in test-time adaption, and 1.4% on AbnTooth, showing strong generalization and privacy-preseving adaptability.

INTERSHIP

Hangzhou DeepSeek Artificial Intelligence Basic Technology Research Co., Ltd. — Beijing, China
AGI Engineer in Aligement Team — Nov 2025 – May 2026 (Estimate)
- Contributed to DeepSeek V3.2. Built high-complexity agent-user interactive evaluation environments with hundreds of tools, fragmented manuals, and asynchronous system dynamics, enabling systematic assessment of long-horizon reasoning and tool-use generalization; all state-of-the-art LLMs (e.g., Claude 4.5) achieve <20% accuracy.
- Developed a scalable agent-data generation framework, automatically synthesizing multi-step tool-use trajectories and diverse interaction patterns; enhanced DeepSeek's agentic capabilities via reinforcement-learning–based post-training, achieving notable improvements on internal reasoning and agent benchmarks.

COMPETITION

OpenAI GPT-OSS-20B Red-Teaming Challenge (Kaggle) Stanford, CA
Researcher, Collobration with Yuhui Zhang at Stanford Aug 2025 – Sep 2025

- Performed a red-teaming analysis of GPT-OSS-20B's agent framework, uncovering 5 reproducible alignment failures, including jailbreak exploits, unfaithful reasoning, and tool-driven data exfiltration (Python/web browsing/markdown rendering), which exposed risks of unauthorized code execution and private-data leakage.
- Developed a reproducible adversarial framework integrating multi-turn emotional framing, token-level gradient prompt tuning, and prompt-rewriting techniques to reproduce these failures and evaluate potential mitigations.

Yeastea Oscillate — AI-Driven Yeast Oscillation Platform for Sustainable Skincare Production Haining, China
iGEM (Gold Medal), Dry Lab Leader, Supervisor: Prof.Wenwen Huang Feb. 2024 – Oct. 2024

- Built **LuminoSeg**, a deep-learning platform integrating YOLOv8, CNN, and Cellpose Cyto3 for automated yeast cell segmentation and fluorescence pattern analysis, achieving 99.3% validation accuracy on 6.4k augmented images.
- Designed an end-to-end pipeline for long-term (70 hr+) oscillation tracking, integrating AI models with wet-lab workflows to automate microscopy analysis and cut manual annotation time by over 80%.
- Developed and deployed the team's iGEM project website with dynamic visualization and interactive model demos.

PROJECT

Exhitopia: Exhibition Booth Management Platform, UIUC Champaign, IL
Developer, Supervisor: Prof. Abdussalam Alawini Feb 2025 – May 2025

- Built a full-stack web application for managing anime exhibition reservations, supporting role-based permissions and real-world exhibition data integration.
- Developed exhibitor- and admin-facing interfaces for real-time inventory management, reservation control, and queue calling; implemented frontend and backend in Typescript.
- Designed and implemented complex database features: multi-condition triggers, stored procedures (e.g., atomic reservation handling), transactional integrity, and role-aware queue logic using MySQL.
- Deployed GCP with Docker and VPC, solving database access attacks through internal IP routing

Benchmarking the Engineering Design Capabilities of LLMs, UIUC Champaign, IL
Member, Supervisor: Prof. Bin Hu Feb 2025 – May 2025

- Designed and developed digital signal processing (DSP) benchmark tasks for ENGDESIGN, including task specification, evaluation pipeline, and automated scoring scheme.
- Contributed as a co-author to the paper Toward Engineering AGI: Benchmarking the Engineering Design Capabilities of LLMs, which was accepted by NeurIPS 2025.

Operating System Kernel Development — Illinix 39, UIUC Champaign, IL
Member, Supervisor: Prof. Kirill Levchenko & Prof. Dong Kai Wang Aug 2024 – Dec 2024

- Built a Unix-like OS kernel from scratch on RISC-V and QEMU, integrating virtio device drivers, ELF program loader, and preemptive multi-process scheduling.
- Implemented a block-device file system supporting open/read/write/seek, metadata operations, and block-level caching through the virtio protocol.
- Designed a virtual memory and process management subsystem based on Sv39 paging, enabling isolated address spaces, dynamic allocation, and secure kernel–user mode transitions.
- Deployed cloud infrastructure on GCP with Docker and VPC, mitigating database access attacks via internal IP routing and automating translation services.

SCHOLARSHIP & AWARD

1. National Scholarship 2023 (top 1%) Oct 2023
2. Zhejiang Government Scholarship (top 5%) Oct 2025
3. Second Class Scholarship, Zhejiang University (top 5%) Oct 2024 & 2025
4. First Class Scholarship, Zhejiang University (top 2%) Oct 2023
5. Gold Medal, International Genetically Engineered Machine (iGEM), 2024 Oct 2024
6. First Prize, The Chinese Mathematics Competitions for College Students Oct 2023

TEACHING & VOLUNTEER EXPERIENCE

Teaching Assistant, Math 213: Discrete Mathematics (Instructor: Prof. Meng Zhang) 2023 Fall & 2025 Fall

- Collaborated with the instructor to optimize course design based on my findings in the 2023 Fall, incorporating coding and algorithmic exercises to make the course better aligned with engineering students.
- Hold discussion sessions, collect and grade coursework, etc.

Teaching Assistant, CS 101: Intro Computing: Engineering & Science (Instructor: Prof. Ong Wee-Liat) 2025 Fall

- Developed Python-based interactive labs and an automated grading script to streamline grading processes for TAs.
- Led lab sessions and office hours, helping students debug code and understand problem concepts; analyzed recurring students' errors and provided actionable feedback to the instructor, assisting him in improving teaching materials.

Volunteer Experience (420+ Hours) Sep 2022 – Present

- **President of Volunteer Practice Center:** led 10+ community service initiatives and expanded outreach from campus to city-wide programs; trained and mentored 100+ new volunteers, improving their engagement.
- **Core volunteer in the 19th Asian Games**: 1) provided information support, including winner prediction, guest liaison and schedule planning; 2) participated in producing music video, including arrangement and filming.
- **Volunteer Teacher**: designed teaching materials and taught English, music and science courses; performed fieldwork on education in China's central and western regions, authored a report that was awarded the outstanding paper.