

Liu Yu

Chengdu, China | liu.yu@std.uestc.edu.cn | [Google Scholar](#) | [Homepage](#)

Education

University of Electronic Science and Technology of China (UESTC) Chengdu, China
Ph.D. in Software Engineering, School of Information and Software Engineering Sept. 2023 – Present
Advisor: [Prof. Ping Kuang](#). Research on trustworthy generative AI, bias mitigation, and hallucination reduction in LLMs and LVLMs.

University of Auckland Auckland, New Zealand
Visiting PhD Student in Computer Science / Joint Program Sept. 2025 – Sept. 2026
Advisor: [Prof. Gillian Dobbie](#). Research on trustworthy generative AI and hallucination mitigation in LLMs and LVLMs.

University of Electronic Science and Technology of China (UESTC) Chengdu, China
M.Eng. in Software Engineering, School of Information and Software Engineering June 2020 – June 2023
Advisor: [Prof. Fan Zhou](#). Research on information propagation and cascade prediction in social networks.

Research Interests

- Responsible and trustworthy AI, large language models, large vision-language models, hallucination mitigation, social bias mitigation, causal intervention, and intelligent information propagation.
- Published or accepted 20 papers in venues and journals including ICML, AAAI, ACM MM, ACL, SIGIR, ICASSP, ICME, and international journals.

Selected Publications

1. **Dismantling Pathological Shortcuts: A Causal Framework for Faithful LVLm Decoding.**
ICML 2026 *First Author*
2. **Causally-Grounded Dual-Path Attention Intervention for Object Hallucination Mitigation in LVLms.**
[AAAI 2026](#) *First Author*
3. **Bimodal Debiasing for Text-to-Image Diffusion: Adaptive Guidance in Textual and Visual Spaces.**
[ACM MM 2025](#) *First Author*
4. **A Mixup-based Unified Framework to Overcome Gender Bias Resurgence.**
[SIGIR 2023](#) *First Author*
5. **Causal-Debias: Unifying Debiasing in Pretrained Language Models and Fine-tuning via Causal Invariant Learning.**
[ACL 2023](#) *Corresponding Author*
6. **Transformer-Enhanced Hawkes Process with Decoupling Training for Information Cascade Prediction.**
[Knowledge-Based Systems 2022](#) *First Author*
7. **Amplifying Commonsense Knowledge via Bi-directional Relation Integrated Graph-based Contrastive Pre-training from Foundation Models.**
[Information Processing & Management 2025](#) *First Author*
8. **Bridging the Fairness Gap: Enhancing Pre-trained Models with LLM-Generated Sentences.**
[ICASSP 2025](#) *First Author*
9. **Knowledge Graphs Acquisition via Forward-Reverse Relation Enhanced Contrastive Pretraining from Large-scale Models.**
[ICME 2025](#) *First Author*
10. **Decoupling Representation and Regressor for Long-Tailed Information Cascade Prediction.**
[SIGIR 2021](#) *Student First Author*
11. **Decoupling Overlapped Feature Spaces: When Continual Learning Meets Fine-grained Classification.**
ICME 2025 *Corresponding Author*
12. **Delight-UPS: Uncalibrated Photometric Stereo with Diffusion Model-Based Relighting.**
ICME 2025 *Corresponding Author*

Honors and Awards

- Second Prize, Sichuan Science and Technology Progress Award, 2024.
- Young Elite Scientists Sponsorship Program for PhD Students, China Association for Science and Technology, 2025.
- National Scholarship, 2021, 2022, 2025; Outstanding Graduate, Sichuan Province, 2020, 2023.
- President’s Scholarship, China Petroleum Scholarship, Shenzhen Stock Exchange Scholarship, and First-Class Academic Scholarship.
- ACM MM 2025 Student Travel Award; AAAI 2026 Student Travel Award.
- Academic Rising Star, Academic Young Talent, Outstanding Graduate Student, and Outstanding Teaching Assistant.

Academic Exchange and Service

- One-year joint PhD program at the University of Auckland, New Zealand, conducting collaborative research on trustworthy generative AI.
- Attended ICML 2026 in Seoul, Korea; AAAI 2026 in Singapore; ACM MM 2025 in Dublin, Ireland; and ICME 2025 in Nantes, France for international academic exchange.
- Reviewer for ICML 2026, NeurIPS 2026, AAAI 2026, ACL 2026, CVPR 2025, ICME 2025, IEEE TKDE, and Information Processing & Management.
- On-site volunteer for ACM MM 2025 and AAAI 2026; student volunteer for SIGIR 2021 Online Conference.

Research Projects

Mechanism and Regulation of the Fact-Fairness Dual Alignment Dilemma in Multimodal Large Models	2026
Core Member. National Natural Science Foundation of China; responsible for mechanism analysis, method design, and experimental validation.	
Trustworthy Generation of Vision-Language Large Models via Causal Reasoning and Multi-Agent Collaboration	2026
Core Member. Sichuan Natural Science Foundation; working on causal reasoning and multi-agent collaboration for trustworthy generation.	
Media Convergence Service Technology R&D and Integrated Application	2022
Participant. National Key R&D Program; responsible for cross-media information integration and analysis over audio and image data.	
Intelligent Media Technology Platform R&D and Industrial Application	2025
Core Member. Sichuan Key R&D Program; responsible for efficient media production, trustworthy communication, and algorithm implementation.	
Comprehensive Public Opinion Situation Awareness and Application	2024
Core Member. Sichuan Key R&D Program; responsible for hotspot prediction, propagation modeling, and experimental analysis.	
Unbiased Decision Optimization for Public-Security Network Governance	2024
Core Member. Chengdu Key R&D Support Program; responsible for unbiased method design, model implementation, and scenario validation.	

Patents

- Long-tailed cascade popularity prediction model, training method, and prediction method. **Granted.**
- Information cascade prediction system and method based on Transformer-enhanced Hawkes process. Filed.
- Decoupling method, representation method, and representation system for cross-modal data. **Granted.**
- Potential hotspot content identification method based on a mainstream-value vertical-domain knowledge base. **Granted.**
- Network content propagation method and system based on cognitive understanding and intelligent governance. **Granted.**
- Deepfake detection and localization method, system, storage medium, and device. Filed.
- Debiasing method, system, storage medium, and terminal for LLM-based pretrained language models. Filed.
- Commonsense knowledge generation method, system, storage medium, and terminal based on bidirectional relation-integrated graphs. Filed.