

Yaman Yu

PHD CANDIDATE · SCHOOL OF INFORMATION SCIENCES

2078 Hazelwood Ct APT C, Urbana, IL 61801

✉ yamanyu2@illinois.edu | 🏠 yamanyume.github.io | 📺 [yaman-yu](https://www.youtube.com/channel/UCv33333333333333333333) | 🐦 [@YamanYu3](https://twitter.com/YamanYu3)

Research Interest

My research bridges human factors with system design and development to create human-centered safety, security, and privacy mechanisms for emerging technologies, including generative AI and decentralized systems.

Education

University of Illinois Urbana-Champaign

PHD IN INFORMATION SCIENCES

- Advisor: Professor Yang Wang

Champaign, IL

2021 – Present

Carnegie Mellon University

MASTER OF SCIENCE IN INFORMATION SECURITY POLICY AND MANAGEMENT

- Graduated with Highest Distinction/Advisor: Professor Alessandro Acquisti

Pittsburgh, PA

2019 - 2021

Tianjin University

BACHELOR OF SCIENCE IN INFORMATION SYSTEM AND MANAGEMENT

- Minors in Ancient Chinese Literature
- Honors thesis/undergrad research advisor: Professor Jin Tian

Tianjin, China

2014 - 2018

Awards, Fellowships, & Grants

2026	Responsible AI Amazon Research Awards: Safeguarding Youths in Multimodal Generative AI: Toward a Trainium-Powered Framework for Safety and Alignment, Co-PI	\$ 50,000
2025	Roblox & Prosocial Design Network: Teaching Young Game Developer Trust and Safety in Gaming, Lead PI	\$ 10,000
2025	MIT EECS Rising Star , (selection rate: 70/327 = 21%)	
2025	Google PhD Fellowship , Privacy, Safety, and Security	\$170,000
2025	LLM Agents Hackathon Winner , Berkeley RDI	\$800
2025	ACM CCS Travel Grant , ACM	\$1000
2024	CHI Best Paper honorable Mention Award , CHI 2024	
2023	Meta: People's Expectations and Experiences with Digital Privacy Research Awards , [contributed to ideation and grant writing], PI: Yang Wang	\$ 100,000
2023	IEEE Symposium on Security and Privacy Travel Grant , S&P	\$ 1,300
	CSRMP Scholar-Google's CS Research Mentorship Program , Google	
	Research Live Visionary Award , UIUC	\$200
2022	Graduate College Travel Grant 2022 , UIUC	\$ 1,000

Publications

PEER-REVIEWED PAPERS [* indicates equal contribution]

[C15] Principles of Safe AI Companions for Youth: Parent and Expert Perspectives.

Yaman Yu, Mohi, Aishi Debroy, Xin Cao, Karen Rudolph, Yang Wang.

In Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI26). Acceptance Rate: 25.3%

[C14] YouthSafe: A Youth-Centric Safety Benchmark and Safeguard Model for Large Language Models.

Yaman Yu, Yiren Liu, Jacky Zhang, Yun Huang, Yang Wang.

In Proceedings of ACM Conference on Computer and Communications Security (CCS25). Acceptance Rate: 16.9%

- [C13] **Cluttered to Clear: Improving the Web Accessibility Design for Screen Reader Users in E-commerce With Generative AI.**
Yaman Yu, Bektur Ryskeldiev, Ayaka Tsutsui, Matthew Gillingham, Yang Wang.
In Proceedings of ACM SIGACCESS Conference on Computers and Accessibility (ASSETS25). *Acceptance Rate: 26%*
- [C12] **Youth-Centered GAI Risks (YAIR): A Taxonomy of Generative AI Risks from Empirical Data.**
Yaman Yu, Yiren Liu, Jacky Zhang, Yun Huang, Yang Wang.
In Proceedings of Symposium on Usable Privacy and Security (SOUPS25). *Acceptance Rate: 21.2%*
- [C11] **Using Affordance to Understand Usability of Web3 Social Media**
Yongqi Guan, Yaman Yu, Kanye Ye WANG.
In Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI25). *Acceptance Rate: 24.9%*
- [C10] **Exploring Parent-Child Perceptions on Safety in Generative AI: Concerns, Mitigation Strategies, and Design Implications**
Yaman Yu, Tanusree Sharma, Melinda Hu, Justin Wang, Yang Wang.
In Proceedings of the IEEE Symposium on Security and Privacy (S&P25). *Acceptance Rate: 14.8%*
- [C9] **Security Perceptions of Users in Stablecoins: Advantages and Risks within the Cryptocurrency Ecosystem**
Yongqi Guan, Yaman Yu, Tanusree Sharma, Z.Huang, K.Qin, Yang Wang, Kanye Ye Wang.
In Proceedings of the IEEE Symposium on Security and Privacy (S&P25). *Acceptance Rate: 14.8%*
- [C8] **Cross-Country Examination of People's Experience with Targeted Advertising on Social Media**
Smirity Kaushik, Tanusree Sharma, Yaman Yu, Yang Wang, Yixin Zou.
In Proceedings of the Privacy Enhancing Technologies Symposium (PETS25). *Acceptance Rate: 21.0%*
- [C7] **“Don’t Put All Your Eggs In One Basket”: How Cryptocurrency Users Choose and Secure Their Wallets**
Yaman Yu, Tanusree Sharma, Sauvik Das, Yang Wang.
In Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI24). *Acceptance Rate: 26.4%*
Best Paper Honorable Mention Award
- [C6] **Design and evaluation of inclusive email security indicators for people with visual impairments**
Yaman Yu, Saidivya Ashok, Smirity Kaushik, Yang Wang, Gang Wang.
In Proceedings of the IEEE Symposium on Security and Privacy (S&P23). *Acceptance Rate: 17.1%*
- [C5] **User Perceptions and Experiences of Targeted Ads on Social Media Platforms: Learning from Bangladesh and India**
Tanusree Sharma*, Smirity Kaushik*, Yaman Yu*, Ishtiaque Ahmed, Yang Wang.
In Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI23). *Acceptance Rate: 27.6%*
- [C4] **GuardLens: Supporting Safer Online Browsing for People with Visual Impairments**
Smirity Kaushik, Nata Barbosa, Yaman Yu, Tanusree Sharma, Zachary Kilhoffer, JooYoung Seo, Sauvik Das, Yang Wang.
In Proceedings of the Symposium on Usable Privacy and Security (SOUPS23). *Acceptance Rate: 22.0%*

LIGHTLY REVIEWED PUBLICATIONS

- [P3] **Designing Safer Generative AI Experiences for Youth.**
Yaman Yu.
In the Responsibility, Inclusion, Safety and Ethics AI Conference (RISE AI25).
- [P2] **Understanding Generative AI Risks for Youth: An Empirical Taxonomy for Safer Digital Futures.**
Yaman Yu, Yiren Liu, Jacky Zhang, Yun Huang, Zhiyi Qian, Yang Wang.
In the Association for Information Science and Technology (ASIS&T25).
- [C3] **“I Am a Blind Seller!”: Picture Taking Assistance for Visually Impaired Individuals in C2C Marketplaces**
Ayaka Tsutsui, Kengo Tanaka, Yaman Yu, Bektur Ryskeldiev.
In Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI LBWs25).
- [C2] **Safeguarding Children in Generative AI: Risk Frameworks and Parental Control Tools**
Yaman Yu.
The ACM international conference on Supporting Group Work (GROUP25 Doctoral Consortium).
- [C1] **Cross-Country Examination of People's Experience with Targeted Advertising on Social Media**
Smirity Kaushik, Tanusree Sharma, Yaman Yu, Yang Wang, Yixin Zou.
In Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI LBWs24).
- [P1] **Examining User Perceptions of Stablecoins: Understandings and Risks.**
Yongqi Guan, Yaman Yu, Tanusree Sharma, Kaihua Qin, Yang Wang, Ye Wang.

In Proceedings of the Symposium on Usable Privacy and Security (SOUPS23 Poster).

PAPERS UNDER REVIEW

Yaman Yu, Minglong Tang, Yuling Sun. 2026. Challenges and Generative AI Design Opportunities in Parent-Adult Child Communication.

Lanjing Liu, **Yaman Yu**, Haojian Jin, Yaxing Yao. 2026. Designing AI Companions for Children: Design Framework from Parents' Perspectives.

Projects

Youth Safety Agent: Agentic LLM Framework for Context-Aware Threat Detection

Champaign, IL

RESEARCH INTERN - UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

Oct 2024 - present

- Advisor: Dr. Yang Wang
- **Youth-GAI Risk Taxonomy Development**: Built a risk classification framework for GAI Youth interactions by analyzing 10k+ adversarial examples (GAI chat logs, Reddit, AI incident DB).
- **LLM-Based Detection System**: Designed an agentic LLM system using Mistral-7B fine-tuning (PyTorch) for real-time detection of covert threats, outperforming GPT-4 Moderation API by 68% in accuracy.

GAI-Enhanced Accessibility Design

Tokyo, Japan

AI RESEARCH INTERN - MERCARI, R4D

Jun 2024 - present

- Advisor: Dr. Bektur RYSKELDIEV
- Optimizing the Blind and Low Vision Users' (BLVs) experience for C2C marketplace.
- Conducted 20+ semi-structured interviews to gain insights of BLVs usability and challenges.
- Developed a generative AI-powered agent to automatically enhance website accessibility design through fine-tuning and retrieval-augmented generation (RAG).
- User Testing with 10+ BLVs and proved the GAI improved version C2C website significantly improved on accessibility

Professional Experience

2025-now **UX Research Contractor**, R4D, Mercari, Supervisor: Dr. Bektur Ryskeldiev

2024 **Research Scientist Intern**, R4D, Mercari, Supervisor: Dr. Bektur Ryskeldiev

2021-now **Research Assistant**, Salt Lab, UIUC, Supervisor: Dr. Yang Wang

2019-2021 **Research Assistant**, Cylab, CMU, Supervisor: Dr. Alessandro Acquisti & Dr. Cristobal Cheyer

2017-2018 **Undergraduate Research Assistant**, Tianjin University, Supervisor: Dr. Jin Tian

Invited Talks & Seminar

2025. *Developing Inclusive Safeguards for At-Risk Users*. **University of Wisconsin-Madison**, Madison, WI, USA.

2025. *Safeguarding Youth in Generative AI*. **Character.ai**, Remote.

2025. *Human-Centered AI Safety: Developing Inclusive Safeguards for At-Risk Users Across Emerging Technologies*. **Johns Hopkins University**, Baltimore, MD, USA.

2025. *Tooling and Method to Empower Ethical Youth AI Developers and Designers*. **Roblox**, San Mateo, CA, USA.

2025. *Protecting and Empowering Next Generation in the Age of AI*. **Berkeley RDI Agentic AI Summit**, Berkeley, CA, USA.

2025. *Youth-Centric Safeguard Framework for Generative AI*. **Google**, Remote, CA, USA.

2024. *Safeguarding Youth in Generative AI*. **Berkeley Security Seminar**, Berkeley, CA, USA.

2024. *Safeguarding Children in Generative AI: Risk Frameworks and Parental Control Tools*. **UIUC School of Information Sciences Research Showcase**, Champaign, IL, USA.

2024. *GenAI for Accessibility: Website Re-generation for Screen Reader Users*. **University of Tokyo**, Tokyo, Japan.

2024. *Inclusive AI: Challenges and Solutions*. **Seminar of HCI**, Remote.

2024. *How Crypto Users Choose and Secure Wallets*. **Security & Privacy Seminar (SPRI)**, Champaign, IL, USA.

2023. *User Risk and Privacy Concerns in Generative AI*. Invited Talk at **UIUC IS Colloquium**, Champaign, IL, USA.

2022. *Managing Cyber Risks: Express Control Impact and Risk Analysis (ECI&RA)*. Invited talk: FloCon, Remote.

Teaching Experience

Spring 25	Methods of Data Science , Co-Instructor	UIUC
	Facilitating Learning Excellence Award, 2025	
Fall 24	Data, Statistical Models, and Info , Teaching Assistant	UIUC
Spring 24	User Research & Evaluation , Teaching Assistant	UIUC
Fall 20	Privacy In The Digital Age , Teaching Assistant	CMU
Spring 17	Network Security , Teaching Assistant	TJU

Mentoring

2025-	Yifan Qiao , Worked on older adults fraud project, UIUC undergraduate student
2025-	Yolanda Xu , Worked on older adults fraud project, UIUC Ph.D. student
2025-	Mohi FNU , Worked on youth AI project, UIUC master student
2025-	Jinwu Hu , Worked on youth AI project, Tsinghua undergraduate student
2025-	Aishi Debroy , Worked on youth AI project, Swarthmore undergraduate student
2025-	Xin Cao , Worked on youth AI project, UIUC undergraduate student
2024-2025	Yuqing Wen , Worked on youth AI project, NUS master student
2024-2025	Xiangyi Liu , Worked on youth AI project, UIUC undergraduate student
2024-2025	Jacky Zhang , Worked on youth AI project, UIUC undergraduate student
2024-2025	Ayaka Tsutsui , Worked on GenAI accessibility project, University of Tsukuba Ph.D. student
2024-2025	Yongqi Guan , Worked on Web3 Security, University of Macau Ph.D. student
2023-2024	Melinda Hu , Worked on youth AI project, high school student
2023-2024	Justin Wang , Worked on youth AI project, high school student
2023-2024	Mengke Wu , Worked on older adult security and privacy project, UIUC Ph.D. student
2022-2023	Yunhao Mei , Worked on targeted ads project, UIUC Master Student
2022-2023	Kailan Mao , Worked on targeted ads project, CMU Master Student

Professional Service

PROGRAM COMMITTEE

2027	ACM ASIA CCS Conference on Computer and Communications Security , Program Committee
2026	1st Workshop on Developmentally Safe Generative AI Environments for Youth , Organizer
2026	CHI conference on Human Factors in Computing Systems , Associate Chair
2025	CHI conference on Human Factors in Computing Systems , Associate Chair
2024	9th Workshop on Inclusive Privacy and Security (WIPS) , Co-Organizer
2023	8th Workshop on Inclusive Privacy and Security (WIPS) , Committee Member
2023	IEEE Symposium on Security and Privacy (S&P) , Session Student Volunteer

SPECIAL RECOGNITIONS

- 2026 Special Recognition for Outstanding Reviews, CHI 2026
- 2026 Special Recognition for Outstanding Reviews, DIS 2026
- 2026 Special Recognition for Outstanding Reviews, IDC 2026
- 2025 Special Recognition for Outstanding Reviews, CHI 2025
- 2025 Special Recognition for Outstanding Reviews, CSCW 2025
- 2024 Special Recognition for Outstanding Reviews, CHI 2024

EXTERNAL REVIEWER

- [CHI] ACM Conference on Human Factors in Computing Systems: 2023, 2024, 2025, 2026
- [CSCW] ACM Conference On Computer-Supported Cooperative Work And Social Computing: 2024, 2025, 2026
- [DIS] Designing Interactive Systems Conference: 2025, 2026
- [UIST] ACM Symposium on User Interface Software and Technology: 2026
- [IDC] Interaction Design and Children Conference: 2026
- [MobileHCI] ACM International Conference on Mobile Human-Computer Interaction: 2023
- [WIPS] SOUPS Workshop on Inclusive Privacy and Security: 2023, 2024

Media Press

- 2026 [What Teens Are Doing With Those Role-Playing Chatbots](#), New York Times
- 2025 [Character.AI to Bar Children Under 18 From Using Its Chatbots](#), New York Times
- 2025 [Protecting and Empowering Next Generation in the Age of AI](#), Business Insider
- 2024 [Exploring Risks in Generative AI](#), Illinois News Bureau
- 2024 [Teens Are Developing Relationships With AI](#), Futurism
- 2024 [Research Highlights in AI Safety](#), EurekAlert!

Technical Skills

Programming: Pytorch, Python, Java, R, JavaScript, SQL, HTML, Prompt Engineering, LLM Fine-tuning; **UX Skills:** Qualitative Research, Quantitative Research, Experiment Design, UX Design, Prototype Development, Usability Testing