

Shuning Zhang



My research focuses on **LLM and AI Agents' privacy, security and safety issues**. I address these challenges by measuring AI agent ecosystems' risks, conducting attacks and designing protection mechanisms. I publish 13 first (co-first) authored CCF-A full papers on CCS, CHI, CSCW, UbiComp conferences, and receive IJCAI25 DDL Challenge Best Paper Candidate, CCS25 HAIPS Distinguished Paper Award, and UbiComp25 Best Student Challenge Award. I interned at Bytedance, Tencent, and have collaborative projects with Ant Group, Tencent, etc.

Education

Tsinghua University

PhD. Student

Sep. 2023 – Present

Institute for Network Sciences and Cyberspace

- Scholarship: National Scholarship, Young Elite Scientists Sponsorship Program by CAST

Tsinghua University

Bachelor

Sep. 2019 – June. 2023

Computer Science and Technology

- Scholarship: National Scholarship (3/210), HAODUSHU Scholarship (1/210), ZHONGSHIMO Scholarship (3/210)

Internship Experience

Tencent, CDG

Research Intern

Jun. 2025 – Aug. 2025

Shenzhen, China

- Designed multi-agent retrieval and generation frameworks for Text2SQL and Text2DSL, improving accuracy by 14%.
- Developed an end-to-end evaluation benchmark to validate edge-case stability across multiple production scenarios.

Bytedance, Overseas education

Algorithm Intern

Mar. 2024 – Jun. 2024

Beijing, China

- Improved math solving models Intern-7B by 5% through process-oriented reward and rejection fine-tuning.
- Implemented Text2Video agents based on the Manim codebase to automatically generate video explanations for math problem-solving."

Visiting Experience

Research assistant

Proposed adaptive neural networks on physiological signals for emotion classification

Jun. 2022 - Aug. 2022

University of Washington

Research assistant

Designed REACTGENIE, a typescript framework for developers to facilitate multimodal apps' building (CHI24)

Jun. 2022 - Aug. 2022

Stanford University

Research assistant

Implemented template based NER model and benchmarked the model's transferability (ACL23)

Jun. 2021 - Aug. 2021

UCLA

Selected Publications

CoVer: Conflict-aware Claim Verification

(EMNLP26) **Shuning Zhang**, Dai Shi, Bohao Chu, Hui Wang, Yuwei Chuai, Yifan Wang, Jingruo Chen, Simin Li, Xin Yi, Hewu Li

Investigating Novice Researchers' Perceptions of Research Privacy Within LLM-Assisted Workflows

(CCS26) **Shuning Zhang**, Changxi Wen, Eve He, Ying Ma, Robert Xiao, Xin Yi, Hewu Li

MultiCom: Multi-Agent Simulation-Based Community Note Evaluation

(In Submission AAAI 27) Changxi Wen*, **Shuning Zhang***, Bohao Chu, Yuwei Chuai, Hui Wang, Xiaolong Li, Dai Shi, Xin Yi

Towards Benchmarking and Assessing Visual Naturalness of Physical World Adversarial Attacks

(CVPR23) Simin Li, **Shuning Zhang**, Gujun Chen, Dong Wang, Pu Feng, Jiakai Wang, Aishan Liu, Xin Yi, Xianglong Liu

PrivWeb: Unobtrusive and Content-aware Privacy Protection For Web Agents

(CHI26) **Shuning Zhang***, Yutong Jiang*, Rongjun Ma, Yuting Yang, Mingyao Xu, Zhixin Huang, Xin Yi, Hewu Li

Characterizing User-Reported Issues of GUI Agents for Web Browsing

(CHI26) **Shuning Zhang***, Jingruo Chen*, Zhiqi Gao, Jiajing Gao, Xin Yi, Hewu Li

Collab: Fostering Critical Identification of Deepfake Videos on Social Media via Synergistic Annotation

(CHI26) **Shuning Zhang**, Linzhi Wang, Shixuan Li, Yuanyuan Wu, Yuwei Chuai, Xin Yi, Hewu Li

VisGuardian: A Lightweight Group-based Visual Privacy Control Technique For Smart Glasses in Home Environments

(CHI26) **Shuning Zhang***, Qucheng Zang*, Yongquan Hu, Jiachen Du, Xueyang Wang, Yan Kong, Xinyi Fu, Suranga Nanayakkara, Xin Yi, Hewu Li

A Scoping Review and Guidelines on Privacy Policy's Visualization from an HCI Perspective

(CHI26) **Shuning Zhang**, Eve He, Sixing Tao, Yuting Yang, Ying Ma, Ailei Wang, Xin Yi, Hewu Li

“Privacy across the boundary”: Examining Perceived Privacy Risk Across Data Transmission and Sharing Ranges of Smart Home Personal Assistants

(CHI26) **Shuning Zhang**, Shixuan Li, Haobin Xing, Jiarui Liu, Yan Kong, Xin Yi, Kanye Ye Wang, Hewu Li

The Pervasive Blind Spot: Benchmarking VLM Inference Risks on Everyday Personal Videos

(UbiComp26) **Shuning Zhang**, Zhaoxin Li, Changxi Wen, Ying Ma, Simin Li, Gengrui Zhang, Ziyi Zhang, Yibo Meng, Hantao Zhao, Xin Yi, Hewu Li

Exploring Collaboration Patterns and Strategies in Human-AI Co-creation Through the Lens of Agency: A Scoping Review of Top-tier HCI Literature

(CSCW25) **Shuning Zhang**, Hui Wang, Xin Yi

Actual Achieved Gain and Optimal Perceived Gain: Modeling Human Take-over Decisions Towards Automated Vehicles’ Suggestions

(CHI25) **Shuning Zhang**, Xin Yi, Shixuan Li, Chuye Hong, Gujun Chen, Jiarui Liu, Xueyang Wang, Yongquan ‘Owen’ Hu, Yuntao Wang, Hewu Li

PrivCAPTCHA: Interactive CAPTCHA to Facilitate Effective Comprehension of APP Privacy Policy

(CHI25) **Shuning Zhang**, Xin Yi, Shixuan Li, Haobin Xing, Hewu Li

The Manipulative Power of Voice Characteristics: Investigating Deceptive Patterns in Mandarin Chinese Female Synthetic Speech

(IMWUT25) **Shuning Zhang***, Han Chen*, Yabo Wang, Yiqun Xu, Jiaqi Bai, Yuanyuan Wu, Shixuan Li, Xin Yi, Hewu Li

Towards Aligning Personalized Conversational Recommendation Agents with Users’ Privacy Preferences

(🏆 CCS25 HAIPS workshop distinguished paper) **Shuning Zhang**, Ying Ma, Jingruo Chen, Simin Li, Xin Yi, Hewu Li

SituGuard: LLM-based Fine-grained Smart Glass Privacy Control in Home Environments

(🏆 UbiComp25 best student challenge award) **Shuning Zhang**, Ying Ma, Qucheng Zang, Yongquan Hu

Synlab: A Spatio-Temporal Aggregation Framework for Collaborative Visual Misinformation Detection in Videos

(🏆 IJCAI25 DDL Workshop best paper candidate) **Shuning Zhang**, Linzhi Wang, Ruokai Zhao, Yiqun Xu, Yuanyuan Wu, Yabo Wang, Simin Li, Xin Yi, Hewu Li

Evaluating the Privacy Valuation of Personal Data on Smartphones

(IMWUT24) Lihua Fan*, **Shuning Zhang***, Yan Kong, Xin Yi, Yang Wang, Xuhai ‘Orson’ Xu, Chun Yu, Hewu Li, Yuanchun Shi

Squeeze’In: Private Authentication on Smartphones based on Squeezing Gestures

(CHI23) Xin Yi, **Shuning Zhang**, Ziqi Pan, Louisa Shi, Fengyan Han, Yan Kong, Hewu Li, Yuanchun Shi

Technical Skills, Language Skills, and Interests

Programming Languages: Python, Javascript, Java, C Script

Language: Chinese (Native), English (GRE 332+5, Toefl 109)

Interests: Marathon, Chinese Folk Dance