Shaoyuan Xie

Ph.D. Candidate, ICS, University of California, Irvine
Tel: (949)368-5560
Email: shaoyux@uci.edu
Homepage: https://daniel-xsy.github.io

Education

University of California, Irvine — Irvine, CA
Ph.D. in Computer Science, Advisor: Professor Qi Alfred Chen — *Expected 2027*

Huazhong University of Science and Technology (HUST) — Wuhan, CN
B.Eng. in Automation, GPA: 3.97 (Top 1st / 212 students) — *Jun 2023*

Industry Experience

Waymo — Mountain View, CA
Research Intern, AI Foundation | Mentor: Chen Song | Manager: Govind Thattai — *Jun 2026 – Sep 2026*

- Conduct research on a Gemma-based multimodal perception foundation model with latent visual reasoning, improving spatial perception and visual reasoning by 16% on internal data.
- Perform large-scale JAX training on 256 TPUs using FSDP, activation rematerialization, and tensor parallelism.

Mercedes-Benz Research & Development North America — San Jose, CA
Research Intern, AD Perception | Mentor: Thomas Monninger | Manager: Sihao Ding — *Jun 2025 – Sep 2025*

- Semantic distillation from LLM to BEV encoder for safer end-to-end driving, reducing closed-loop simulation collision rate by 16% on safety-critical scenarios.

Shanghai Artificial Intelligence Laboratory — Shanghai, CN
Research Intern, OpenMMLab | Mentor: Wenwei Zhang | Manager: Kai Chen — *Jun 2023 – Aug 2023*

- Post-training of InternLM language models on tool usage, improving math problem-solving performance on GSM8k dataset.

Selected Publications

10 top-tier publications, including **7 first/co-first-author papers** at TPAMI, ICLR, NeurIPS, ICCV, ECCV, CoRL, IROS, TMLR and NDSS

- **SafeDriveVLA: Navigation-Conditioned World Model Dreaming for Conflict-Aware End-to-End Autonomous Driving**
  Shaoyuan Xie*, Zihan Zhang*, Jingxuan Wang, Jiashu Qu, Xiaoqing Liang, Lingdong Kong, Junchi Lu, Henrik I Christensen, Qi Alfred Chen
  CoRL, 2026

- **BEVLM: Distilling Semantic Knowledge from LLMs into Bird's-Eye View Representations**
  Thomas Monninger*, Shaoyuan Xie* (**order determined alphabetically*), Qi Alfred Chen, Sihao Ding
  ECCV, 2026

- **FlyTrap: Physical Distance-Pulling Attack Towards Camera-based Autonomous Target Tracking Systems**
  Shaoyuan Xie, Mohamad Fakih, Junchi Lu, Fayzah Alshammari, Ningfei Wang, Takami Sato, Halima Bouzidi, Mohammad Al Faruque, Qi Alfred Chen
  NDSS, 2026

- **Are VLMs Ready for Autonomous Driving? An Empirical Study from Reliability, Data, and Metric Perspectives**
  Shaoyuan Xie, Lingdong Kong, Yuhao Dong, Chonghao Sima, Wenwei Zhang, Qi Alfred Chen, Ziwei Liu, Liang Pan
  ICCV, 2025

- **Can We Trust Embodied Agents? Exploring Backdoor Attacks against Embodied LLM-Based Decision-Making Systems**
  Ruochen Jiao*, Shaoyuan Xie* (**co-first author*), Justin Yue, Takami Sato, Lixu Wang, Yixuan Wang, Qi Alfred Chen, Qi Zhu
  ICLR, 2025

- **Benchmarking and Improving Bird's Eye View Perception Robustness in Autonomous Driving**
  Shaoyuan Xie, Lingdong Kong, Wenwei Zhang, Jiawei Ren, Liang Pan, Kai Chen, Ziwei Liu
  TPAMI, 2025

- **On the Adversarial Robustness of Camera-based 3D Object Detection**
  Shaoyuan Xie, Zichao Li, Zeyu Wang, Cihang Xie
  TMLR, 2024

- **The Heat is On: Understanding and Mitigating Vulnerabilities of Thermal Image Perception in Autonomous Systems**
  Hrushikesh Varma, Shaoyuan Xie, Michael Clifford, Qi Alfred Chen, Takeshi Sugawara, Sara Rampazzi
  NDSS, 2026

- **Revisiting Physical-World Adversarial Attack on Traffic Sign Recognition: A Commercial Systems Perspective**
  Ningfei Wang, Shaoyuan Xie, Takami Sato, Yunpeng Luo, Kaidi Xu, Qi Alfred Chen
  NDSS, 2025

- **RoboDepth: Robust Out-of-Distribution Depth Estimation under Corruptions**
  Lingdong Kong, Shaoyuan Xie, Hanjiang Hu, Lai Xing Ng, Benoit R. Cottereau, Wei Tsang Ooi
  NeurIPS, 2023

- **Adversarial Spatial Transform Marker: A Black-Box Pose Estimation Attack for Autonomous UAV Landing**
  Junchi Lu, Fayzah Alshammari, Shaoyuan Xie, Xiaoqing Liang, Qi Alfred Chen
  IROS, 2026

- **"Prompter Says": A Linguistic Approach to Understanding and Detecting Jailbreak Attacks Against Large-Language Models**
  Dylan Lee, Shaoyuan Xie, Shagoto Rahman, Kenneth Pat, David Lee, Qi Alfred Chen
  ACM CCS Workshop - LAMPS, 2024

Pre-Print (Under Review)

- **VPIXDistill: Securing Visual Inputs against Prompt Injection via Explainability-Guided Distillation**
  Junchi Lu, Aarifah Ullah, Chi Zhang, Shaoyuan Xie, Mahyar Ghazanfari, Peng Wei, Qi Alfred Chen

- **DistillTrace: Tracing LLM Distillation via Keyed Solution Preferences**
  Xiaoqing Liang, Xinqi Su, Jiashu Qu, Shaoyuan Xie, Ruihong Huang, Qi Alfred Chen

- **Beyond Surface Tokens: Watermarking Language Models through Private Solution Choices**
  Xiaoqing Liang, Xinqi Su, Jiashu Qu, Shaoyuan Xie, Ruihong Huang, Qi Alfred Chen

Research Impact

- Email acknowledgements from DJI on autonomous tracking vulnerabilities
- Responsible vulnerability disclosure acknowledged by a major automotive manufacturer for traffic-sign recognition vulnerabilities

Awards and Honors

- Best Demo, 3rd USENIX Symposium on Vehicle Security and Privacy (2025, *selected for best demo rate 18.2% = 2/11*)
- Honours Bachelor's Degree (2023, *top 1% in univ.*)
- Outstanding Undergraduate Student (2021, *top 1% in univ.*)
- National Scholarship, Ministry of Education of China (2021, *top 0.2% nation-wide*)
- National Scholarship, Ministry of Education of China (2020, *top 0.2% nation-wide*)

Academic Services

- *Competition Organizer*: IROS RoboSense (2025), ICRA RoboDrive (2024), ICRA RoboDepth (2023)
- *PC member*: ACM CCS Artifact Evaluation (2024, 2025, 2026), NDSS Artifact Evaluation (2025)
- *Reviewer*: CVPR (2023, 2024, 2025, 2026), ICCV (2025), ECCV (2026), BMVC (2026), ICLR (2026), CoRL (2026), IROS (2026), IJCV, RA-L, TDSC; TMM, TMLR
- *External Reviewer*: USENIX Security (2024, 2025), IEEE S&P (2024), ACM CCS (2024)