Xinrui Shi

(+86) 182-2171-2378 | irene_05@sjtu.edu.cn | https://irene-shi.github.io/

Education

- **Shanghai Jiao Tong University (SJTU)** *Sep 2023 – Jun 2027 (Expected)*
  B.S. in Software Engineering, School of Computer Science Shanghai, China
  - GPA: **4.13/4.3**, Rank: **1/97**
  - Selected Courses: Linear Algebra (A+), Probability and Statistics (A+), Advanced Data Structure (A+), Introduction to Computer Systems (1) (2) (A+), Machine Learning (A+).
- **Stanford University** *Jun 2026 – Aug 2026*
  Undergraduate Visiting Research Intern Stanford, CA, USA

Honors and Awards

- **National Scholarship** (Top 0.4% Nationwide) *2024, 2025, 2026*
- Shanghai Jiao Tong University Undergraduate Category A Scholarship (Top 1% in SJTU) *2024, 2026*
- Shanghai Jiao Tong University Undergraduate Category B Scholarship *2025*
- CCF Elite Collegiate Award (1 of 101 Nationwide) *2026*

Research Interests

My research goal is to **build capable and trustworthy AI systems** that can operate autonomously in complex real-world environments. In particular, I'm interested in **agents** and **AI safety**, especially how interaction shapes agent behavior over time. My broader interests also include multimodal reasoning and machine learning systems.

Publications

[1] **Xinrui Shi**[*], Yanzhe Zhang[*], Diyi Yang. **Emergent Collusion in Long-Horizon LLM Agent Interaction.** *Submitted to ICLR 2027.* [PDF]

[2] **Xinrui Shi**[*], Kai Liu[*], Ziqing Zhang, Jianze Li, Anqi Li, Yulun Zhang. **DRScaffold: Boosting Dense-Scene Reasoning in Lightweight Vision Language Models.** *In Proceedings of the 40th Annual Conference on Neural Information Processing Systems (NeurIPS). 2026.* [PDF]

[3] Jiaxin Li*, Yuxiang Wu*, Zhenkai Zhang, **Xinrui Shi**, Haoyuan Wang, Yichen Zhao, Su Linxiang, Chenyang Yu, Mingyu Zhang, Yifan Ding, Boran Wen, Li Zhang, Ruiyang Liu, Yong-Lu Li. **HAT-4D: Lifting Monocular Video for 4D Multi-object Interactions via Human-Agent Collaboration.** *In Proceedings of the 19th European Conference on Computer Vision (ECCV). 2026.* [PDF]

[4] Tong Xin, **Xinrui Shi**, Mingkai Dong, Zeyu Mi. **Inference in the Shadows: Taming Memory Bandwidth Contention in Mobile LLM Inference with Sereno.** *In Proceedings of the 20th USENIX Symposium on Operating Systems Design and Implementation (OSDI). 2026.* [PDF]

Research Experiences

- **Multi-Agent Emergent Collusion** *May 2026 – Sep 2026*
  SALT lab, Stanford University; Advisor: Diyi Yang Stanford, CA, USA
  - Designed a long-horizon multi-agent framework to study emergent collusion, where agents solve tasks, exchange evidence, verify peers, and adapt from interaction history under conflicting environment.
  - Developed metrics and onset analyses for collusion across 10 frontier LLMs; observed collusion in 93.6% of trajectories with distinct formation patterns across model families.
  - Characterized the dynamics underlying emergent collusion, identifying peer influence, interaction history, and feedback signals as key factors shaping instruction violation and behavioral persistence.
  - **Invited talk, Microsoft AI Safety & Security Science Speaker Series, October 2026.**

- **Enhancing Reasoning in Lightweight Vision-Language Models** *Dec 2025 – May 2026*
  Advisor: Yulun Zhang (Shanghai Jiao Tong University) Shanghai, China
  - Designed a dense-scene reasoning benchmark and an automated data generation pipeline with structured annotations to evaluate reasoning capability and hallucination robustness.
  - Developed a structured supervision framework that decomposes dense-scene reasoning into entity grounding, relation modeling, step-wise reasoning, and answer generation.
  - Improved Qwen2.5-VL-3B dense-scene reasoning capability by 27.6%, surpassing Qwen2.5-VL-32B while maintaining gains on hallucination benchmarks without degrading general multimodal capabilities.
- **On-device LLM Bandwidth Control** *Aug 2025 – Dec 2025*
  IPADS lab, Shanghai Jiao Tong University; Advisor: Mingkai Dong, Zeyu Mi Shanghai, China
  - Diagnosed memory-bandwidth contention on Snapdragon 8 Elite and Dimensity 9400 platforms when running LLMs, identifying unobservable NPU traffic, missing partitioning support, and queue-depth imbalance.
  - Implemented speculative decoding with fine-grained bandwidth-control mechanisms, and designed a multi-source, length-weighted, self-purifying n-gram cache to recover decoding performance.
  - Evaluated the system on OnePlus 13 against PowerServe, MNN, and llama.cpp across 30 applications, reducing foreground jank rate by 58.5% and improving LLM decode throughput by 26.4% over PowerServe.
- **Monocular 4D Scene Reconstruction via Human-Agent Collaboration** *Dec 2024 – Nov 2025*
  RHOS lab, Shanghai Jiao Tong University; Advisor: Yonglu Li Shanghai, China
  - Led the construction of a high-fidelity 4D dataset and end-to-end data pipeline, including geometric calibration, temporal synchronization, down-sampling and pixel-level object and key-region annotation.
  - Developed a multi-level human-in-the-loop refinement protocol and used human-refined outputs to fine-tune 4D reconstruction models, creating a self-improving data engine.
  - Introduced a memory metric for information recovery in occluded regions and evaluated the system, showing 15.2% better temporal consistency and 22.4% better occluded-region recovery over SOTA baselines.

Project Experiences

- **Intelligent Tutoring System** *Jun 2025 – Jul 2025*
  - Built an end-to-end document processing and retrieval pipeline for multi-format teaching materials, enabling semantic search at both file and page levels.
  - Built an automated workflow to construct course knowledge graphs from teaching materials, with interactive visualization to explore concept relationships.
  - Developed a RAG-based tutoring system that integrates vector and knowledge graph retrieval for course question answering.

Skills

Languages	Chinese (Native), English (Fluent, TOEFL: 107, Speaking 26)
GRE	Verbal 162 / Quant 170 / Analytical Writing 4
Programming	C++, Python, Java
Tools	Git, LaTeX, PyTorch, Blender, MySQL, MongoDB, Neo4j