Changsheng Sun

Email: cssun@u.nus.edu · **Mobile:** +65-8039-5393 · Singapore
Homepage: sunchangsheng.com · **LinkedIn:** changshengsun

Changsheng Sun is a Ph.D. candidate at the PLSE Lab, School of Computing, National University of Singapore, advised by Prof. Dong Jin Song. His research focuses on trustworthy AI, spanning GNN explainability, risk-aware robust optimization (CVaR-based), and LLM safety. He has published at venues including NeurIPS, AAAI, ICSE, ASE, and IEEE S&P Workshops (Google Scholar: ~590 citations, h-index: 7), and serves as a reviewer for NeurIPS, ICML, and AAAI etc.. Previously, he worked on distributed reinforcement learning systems at JD Intelligent Cities Research.

Education

National University of Singapore **Jan 2022 – Early 2026**
Doctor of Philosophy, Computer Science
Thesis: *Trustworthy Geometric Learning: From Structural Biases to Risk-Aware Robustness*
Advisor: Prof. Dong Jin Song

National University of Singapore **Aug 2020 – Dec 2021**
Master of Computing, Artificial Intelligence

Xidian University, Xi'an, China **Sep 2015 – Aug 2019**
Bachelor of Engineering, Computer Science

Skills

Programming	Python, Bash Scripting, PyTorch
Systems	Linux/Unix, Git, Docker, Multi-GPU Training, ClickHouse, Spark, Ray
Research Interests	Trustworthy AI, Graph Neural Networks, LLM Safety, Multi-Agent Systems
Language	English, Mandarin

Research Experience

NUS School of Computing **Jul 2021 – Dec 2021**
Research Assistant Singapore

- Built and maintained a reproducible evaluation pipeline for trustworthiness assessment of deep learning systems, emphasizing uncertainty-aware reliability analysis.
- Contributed to InputReflector, published at ASE 2022.
- Hosts: Prof. Dong Jin Song and Prof. Xiao Yan.

NUS-Singtel Cyber Security R&D Lab **Jan 2020 – Jun 2020**
Research Intern Singapore

- Developed a white-box testing approach for deep neural networks by leveraging intermediate-layer signals and density estimation for reliability assessment.
- Contributed to work published at ICSE 2021.
- Host: Prof. David S. Rosenblum.

JD Intelligent Cities Research, JD.com (MSRA Urban Computing Group) **Jan 2018 – Jun 2018**
Research Intern & Algorithm Engineer Beijing, China

- Prototyped a distributed DQN research system on spatiotemporal data using the Ray execution framework; focused on training scalability and system performance profiling.
- Hosts: Prof. Yu Zheng and Dr. Junbo Zhang.

Publications

Google Scholar: ~590 citations, h-index: 7.

[1] *Ignoring Directionality Leads to Compromised Graph Neural Network Explanations*. Changsheng Sun, Xinke Li, Jin Song Dong. IEEE S&P Workshops (SPW), **2025**.

[2] *PointCVaR: Risk-optimized Outlier Removal for Robust 3D Point Cloud Classification*. Xinke Li, Junchi Lu, Henghui Ding, Changsheng Sun, Joey Tianyi Zhou, Chee Yeow Meng. AAAI, **2024**.

[3] *Repairing Failure-inducing Inputs with Input Reflection*. Yan Xiao, Yun Lin, Ivan Beschastnikh, Changsheng Sun, David S. Rosenblum, Jin Song Dong. ASE, **2022**.

[4] *PIANO: Influence Maximization Meets Deep Reinforcement Learning*. Hui Li, Mengting Xu, Sourav S. Bhowmick, Shafiq Joty Rayhan, Changsheng Sun, Jiangtao Cui. IEEE Trans. Computational Social Systems, **2022**.

[5] *Self-Checking Deep Neural Networks in Deployment*. Yan Xiao, Ivan Beschastnikh, David S. Rosenblum, Changsheng Sun, Sebastian Elbaum, Yun Lin, Jin Song Dong. ICSE, **2021**.

[6] *Digraph Inception Convolutional Networks*. Zekun Tong, Yuxuan Liang, Changsheng Sun, Xinke Li, David S. Rosenblum, Andrew Lim. NeurIPS, **2020**.

Preprints:

[7] *From Attack Surfaces to Actual Operations: A Survey of Modern LLM Jailbreaks*. Ruikang Zhou, Changsheng Sun, Mark Huasong Meng. Under Review, **2025**.

[8] *Risk-Aware Robust Graph Network Explanation*. Changsheng Sun, Xinke Li, Jin Song Dong. Under Review, **2025**.

[9] *Generalizing Neural Networks by Reflecting Deviating Data in Production*. Yan Xiao, Yun Lin, Ivan Beschastnikh, Changsheng Sun, David S. Rosenblum, Jin Song Dong. arXiv:2110.02718, **2021**.

[10] *Directed Graph Convolutional Network*. Zekun Tong, Yuxuan Liang, Changsheng Sun, David S. Rosenblum, Andrew Lim. arXiv:2004.13970, **2020**.

[11] *Disco: Influence Maximization Meets Network Embedding and Deep Learning*. Hui Li, Mengting Xu, Sourav S. Bhowmick, Changsheng Sun, Zhongyuan Jiang, Jiangtao Cui. arXiv:1906.07378, **2019**.

Services & Awards

Program Committee / Reviewer:
NeurIPS, ICML, AAAI, WWW, FSE, ASE, ISACE (2022–2026)

Teaching Assistant:
CS5232 Formal Specification & Design Techniques (2024);
CS4218 Software Testing (2023);
CS4211 Formal Methods for Software Engineering (2022)

Awards:
Graduate Student Travel Grants (IEEE S&P 2025, AAAI 2024, ASE 2022);
NUS Research Scholarship (2022–2026)