

# Changsheng Sun

Email: [cssun@u.nus.edu](mailto:cssun@u.nus.edu) · Mobile: +65-8039-5393 · Singapore

Homepage: [sunchangsheng.com](http://sunchangsheng.com) · LinkedIn: [changshengsun](https://www.linkedin.com/in/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, and LLM safety. He has published at venues including NeurIPS, AAAI, ICSE, ASE, and IEEE S&P Workshops ([Google Scholar](#): ~600 citations, h-index: 7), and serves as a reviewer for NeurIPS, ICML, and AAAI etc.. Previously, he was an intern working on distributed reinforcement learning systems at JD Intelligent Cities Research.

## Education

---

<b>National University of Singapore</b> Doctor of Philosophy, Computer Science Thesis: <i>Trustworthy Geometric Learning: From Structural Biases to Risk-Aware Robustness</i> Advisor: <a href="#">Prof. Dong Jin Song</a>	<b>Jan 2022 – Early 2026</b>
<b>National University of Singapore</b> Master of Computing, Artificial Intelligence	<b>Aug 2020 – Dec 2021</b>
<b>Xidian University, Xi'an, China</b> Bachelor of Engineering, Computer Science	<b>Sep 2015 – Aug 2019</b>

## Skills

---

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

## Experience

---

<b>NUS School of Computing</b> <i>Research Assistant</i>	<b>Jul 2021 – Dec 2021</b> Singapore
<ul style="list-style-type: none"><li>Built and maintained a reproducible evaluation pipeline for trustworthiness assessment of deep learning systems, emphasizing uncertainty-aware reliability analysis.</li><li>Contributed to InputReflector, published at ASE 2022.</li><li>Hosts: Prof. Dong Jin Song and Prof. Xiao Yan.</li></ul>	
<b>NUS-Singtel Cyber Security R&amp;D Lab</b> <i>Research Intern</i>	<b>Jan 2020 – Jun 2020</b> Singapore
<ul style="list-style-type: none"><li>Developed a white-box testing approach for deep neural networks by leveraging intermediate-layer signals and density estimation for reliability assessment.</li><li>Contributed to work published at ICSE 2021.</li><li>Host: <a href="#">Prof. David S. Rosenblum</a>.</li></ul>	
<b>JD Intelligent Cities Research, JD.com (MSRA Urban Computing Group)</b> <i>Research Intern &amp; Algorithm Engineer</i>	<b>Jan 2018 – Jun 2018</b> Beijing, China
<ul style="list-style-type: none"><li>Prototyped a distributed DQN research system on spatiotemporal data using the Ray execution framework; focused on training scalability and system performance profiling.</li><li>Hosts: <a href="#">Prof. Yu Zheng</a> and Dr. Junbo Zhang.</li></ul>	

## Publications

---

Google Scholar: ~600 citations, h-index: 7.

- [1] *From Attack Surfaces to Actual Operations: A Survey of Modern LLM Jailbreaks*. Ruikang Zhou, Changsheng Sun, Mark Huasong Meng. ACL Findings, **2026**.
- [2] *Ignoring Directionality Leads to Compromised Graph Neural Network Explanations*. Changsheng Sun, Xinke Li, Jin Song Dong. IEEE S&P Workshops (SPW), **2025**.
- [3] *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**.
- [4] *Repairing Failure-inducing Inputs with Input Reflection*. Yan Xiao, Yun Lin, Ivan Beschastnikh, Changsheng Sun, David S. Rosenblum, Jin Song Dong. ASE, **2022**.
- [5] *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**.
- [6] *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**.
- [7] *Digraph Inception Convolutional Networks*. Zekun Tong, Yuxuan Liang, Changsheng Sun, Xinke Li, David S. Rosenblum, Andrew Lim. NeurIPS, **2020**.

### Preprints:

- [7] *Risk-Aware Robust Graph Network Explanation*. Changsheng Sun, Xinke Li, Jin Song Dong. Under Review, **2025**.
- [8] *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**.
- [9] *Directed Graph Convolutional Network*. Zekun Tong, Yuxuan Liang, Changsheng Sun, David S. Rosenblum, Andrew Lim. arXiv:2004.13970, **2020**.
- [10] *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)

SoC Tuition Fee Support Award (2026)