Research Statement

Sun Jun
School of Computing and Information Systems, Singapore Management University
Tel: (65) 6828-1312; Email: junsun@smu.edu.sg
2 (Day) 12 (Month) 2024 (Year)

Background

My research interests center on the application of formal methods to enhance the safety and security of a wide range of systems, with a recent emphasis on fundamental AI models and AI-enabled systems. I am particularly drawn to the potential of systematic and rigorous approaches, grounded in formal reasoning, to bring greater structure and reliability to complex technological landscapes. By fostering a more organized and predictable foundation for these systems, I aspire to contribute to creating a world that is not only more secure but also more conducive to human flourishing and enjoyment.

Research Areas

AI Safety and Security: My research group has been actively engaged in a series of studies focused on multiple critical aspects of AI safety and security, including:

1. Evaluating the safety and security of foundational AI models,

2. Developing systematic methodologies for improving the safety and security of AI systems, and

3. Establishing frameworks for certifying the safety and security of AI models and AI-enabled systems.

Our goal is to advance this line of inquiry in the coming years by developing innovative techniques and tools with tangible impact. These contributions may take the form of theoretical breakthroughs that reshape the community's understanding and approach to AI safety, or practical methodologies that gain widespread adoption in industry. We firmly believe that the importance of addressing AI safety and security cannot be overstated, as it is foundational to the responsible development and deployment of AI technologies.

New Approaches to Software Engineering: In addition to AI safety, my research group is investigating the transformative impact of AI on software engineering practices. Specifically, we are conducting a series of studies to evaluate whether AI technologies could potentially replace human programmers in the near future. This research is particularly significant given its potential to directly affect millions of software engineers worldwide. By understanding and addressing these changes, we aim to contribute to the evolution of software engineering practices in ways that ensure their relevance and utility in an increasingly AI-driven landscape.

Selected Publications and Outputs

- Shuang Liu, Chenglin Tian, Jun Sun, Ruifeng Wang, Wei Lu, Yongxin Zhao, Yinxing Xue, Junjie Wang, and Xiaoyong Du: Semantic Conformance Testing of Relational DBMS, VLDB 2025.
- Yufan Cai, Zhe Hou, David Sanan, Xiaokun Luan, Yun Lin, Jun Sun, Jin Song Dong: Automated Program Refinement: Guide and Verify Code Large Language Model with Refinement Calculus, POPL 2025.
- Hanmo Yu, Zan Wang, Xuyang Chen, Junjie Chen, Jun Sun, Shuang Liu, and Zishuo Dong:Mitigating Regression Faults Induced by Feature Evolution in Deep Learning Systems, TOSEM 2025.
- PeiYuan Tang, Xiaodong Zhang, Chunze Yang, Haoran Yuan, Jun Sun, Danfeng Shan, Zijiang James Yang: Unleashing the Power of Visual Foundation Models for Generalizable Semantic Segmentation, AAAI 2025.
- Zongxin Liu, Zhe Zhao, Fu Song, Jun Sun, Pengfei Yang, Xiaowei Huang, Lijun Zhang: Training Verification-Friendly Neural Networks via Neuron Behavior Consistency, AAAI 2025.
- Wei Zhao, Zhe Li, Yige Li, Ye Zhang, and Jun Sun: Defending Large Language Models Against Jailbreak Attacks via Layer-specific Editing, Findings of EMNLP 2024.
- Yihao Zhang, Zeming Wei, Jun Sun and Meng Sun: Towards General Conceptual Model Editing via Adversarial Representation Engineering, NeurIPS 2024.
- Jingnan Zheng, Han Wang, Tai D. Nguyen, An Zhang, Jun Sun, and Tat-Seng Chua: ALI-Agent: Assessing LLMs'Alignment with Human Values via Agent-based Evaluation, NeurIPS 2024.
- Ruihan Zhang, and Jun Sun: Certified Robust Accuracy of Neural Networks Are Bounded due to Bayes Errors, CAV 2024.
- Yang Sun, Chris Poskitt, Xiaodong Zhang, and Jun Sun: REDriver: Runtime Enforcement for Autonomous Vehicles, ICSE 2024.
- Ziqi Shuai, Zhenbang Chen, Kelin Ma, Kunlin Liu, Yufeng Zhang, Jun Sun, and Ji Wang: Partial Solution Based Constraint Solving Cache in Symbolic Execution, FSE 2024.
- Jinhao Dong, Jun Sun, Yun Lin, Yedi Zhang, Murong Ma, Jin Song Dong, and Dan Hao: Revisting the Conflict-Resolving Problem from a Semantic Perspective, ASE 2024.
- Pham Hong Long, and Jun Sun: Certified Continual Learning for Neural Network Regression, ISSTA 2024.
- Huijia Sun, Christopher M. Poskitt, Yang Sun, Jun Sun, and Yuqi Chen: ACAV: A Framework for Automatic Causality Analysis in Autonomous Vehicle Accident Recordings, ICSE 2024.
- Bing Sun, Jun Sun, Wayne Koh, and Jie Shi: Neural Network Semantic Backdoor Detection and Mitigation: A Causality-Based Approach, USENIX Security 2024.
- Yedi Zhang, Guangke Chen, Fu Song, Jun Sun and Jin Song Dong: Certified Quantization Strategy Synthesis for Neural Networks, Formal Methods 2024.
- Shunkai Zhu, Jingyi Wang, Jun Sun, Jie Yang, Xingwei Lin, Liyi Zhang, and Peng Cheng: Better Pay Attention Whilst Fuzzing, IEEE Transactions on Software Engineering, 2024.
- Zhe Zhao, Guangke Chen, Tong Liu, Taishan Li, Fu Song, Jingyi Wang, and Jun Sun: Attack as Detection: Using Adversarial Attack Methods to Detect

Abnormal Examples, ACM Transactions on Software Engineering Methodology, 2024.

- Yinxing Xue, Jiaming Ye, Wei Zhang, Jun Sun, Lei Ma, Haijun Wang, Jianjun Zhao: xFuzz: Machine Learning Guided Cross-Contract Fuzzing, IEEE Trans. Dependable Secur. Comput. 21(2): 515-529 (2024)
- Yuhan Zhi, Xiaofei Xie, Chao Shen, Jun Sun, Xiaoyu Zhang, and Xiaohong Guan: Seed Selection for Testing Deep Neural Networks, ACM Transactions on Software Engineering Methodology, 2024.
- Dongxia Wang, Tim Muller, Jun Sun: Provably Secure Decisions based on Potentially Malicious Information, IEEE Transactions on Dependable and Secure Computing, 2024.