Sean Nian

https://seannian.github.io/ snian2@illinois.edu 1(408)-370-8003

Education

University of Illinois Urbana-Champaign Champaign, IL
Doctor of Philosophy in Computer Science Aug. 2025 – Aug. 2030

San Jose State University San Jose, CA
Bachelor of Science in Computer Science with Honors Aug. 2021 – Dec. 2024

Publications

K. Hollingsworth, **S. Nian**, A. Gutierrez, and A. Padmanabhan. *An Analysis of Network Overhead in Distributed TinyML.* [Short Paper]. IEEE/ACM Symposium on Edge Computing (SEC), Rome, Italy, 2024, pp. 449-455. DOI: 10.1109/SEC62691.2024.00051.

S. Nian and G. Ishigaki. *Autoscaling in Knative for Serverless Computing Using Deep Reinforcement Learning.* [Short Paper]. IEEE Annual Congress on Artificial Intelligence of Things (IEEE AIoT), Osaka, Japan, 2025. (Under Review).

Presentations

An Analysis of Network Overhead in Distributed TinyML. [Sole Presenter]. INTERACT Workshop, IEEE/ACM Symposium on Edge Computing (SEC), Rome, Italy, 2024.

An Analysis of Network Overhead in Distributed TinyML. [Poster, Co-presented]. Harvey Mudd College CS Summer Research Multi-Institutional Seminar, 2024.

Other

S. Nian, A. Huang, and B. Reed. *Building a Mastodon Compatible Java Server for ActivityPub.* [Preprint]. arXiv:2412.09011 [cs.SI], 2024. DOI: 10.48550/arXiv.2412.09011.

S. Qi, X. I. Quan, T. Park, and B. Makani. *The Role of Artificial Intelligence in Enhancing Scholarly Research – AI Tools Evaluation.* [Acknowledged Research Assistance: **S. Nian**]. Quarterly Review of Business Disciplines, vol. 11, no. 3/4, Feb. 2025, pp. 90-114.

Research

User-Centric Machine Learning Systems with Prof. Fan Lai Aug. 2025 – Present
University of Illinois Urbana-Champaign

- Effectively serving machine-learning models in long-context settings, with an emphasis on resource-latency trade-offs

ML Systems for Serverless Computing with Prof. Genya Ishigaki Aug. 2024 – April 2025
San Jose State University

- Developed a Deep Reinforcement Learning (DRL) agent using PPO to perform simultaneous horizontal (pod count) and vertical (CPU/memory) scaling in Knative
- Achieved superior performance over native autoscalers, reducing memory consumption by up to 44% while increasing request throughput by 18.6% in high-concurrency scenarios
- Engineered a custom OpenAI Gym environment wrapping a live Kubernetes cluster to train the agent and programmatically apply scaling decisions by modifying Knative manifests in real-time

Optimizing TinyML Systems with Prof. Arthi Padmanabhan May 2024 – July 2024
Harvey Mudd College (NSF Funded REU)

- Engineered a distributed TinyML inference system using C++ on ESP32-S3 microcontrollers by partitioning a quantized TensorFlow MobileNet model and distributing intermediate computations over Bluetooth Low Energy (BLE)
- Optimized the BLE network communication protocol for enhanced fault-tolerance and energy efficiency, analyzing network overhead to reduce the distributed inference failure rate to 0%
- Authored a 6-page paper on the network overhead and challenges of distributed TinyML, accepted for publication and presentation at the INTERACT workshop at ACM/IEEE SEC 2024

Decentralized Social Media Networks with Prof. Ben Reed Jun. 2023 – May 2024
San Jose State University

- Explored decentralized social media networks by analyzing the ActivityPub Protocol's architecture
- Engineered moth, an open-source ActivityPub server in Java, by implementing RESTful API endpoints with JSON payloads using Spring WebFlux to ensure robust decentralized communication
- Reverse-engineered Mastodon's client-server API by analyzing network traffic with Wireshark,

enabling full compatibility between our moth server and the official Mastodon web frontend

Industry Experience

PACCAR Silicon Valley Innovation Center May 2025 – Present
R&D Software Engineer Intern

- Architected an edge-to-cloud data pipeline for an AI-powered dashcam system in Kotlin using Android Studio with low-level libIPC components, optimizing for resource-constrained hardware
- Engineered a vehicle data processing module by creating custom DBC files compliant with SAE J1939 standards; decoding 60+ CAN messages, mapping over 600 signals and 20,000 PGN/SPNs
- Implemented an edge management system to prevent on-device storage exhaustion, implementing a RAM and ROM monitoring system to automatically offload encrypted files to the cloud
- Enabled fleet-wide data visualization and analysis, creating a scalable RESTful API using Node.js and Express.js to ingest high-volume telemetry into the PACCAR data lake via MongoDB schemas
- Presented the end-to-end system at the PACCAR Innovation Center's Q3 review and briefed leadership at PACCAR Global Connected Services, demonstrating production viability and leading to plans for continued development

Awards

2025 NSF Graduate Research Fellow
2024 NSF Student Travel Grant IEEE/ACM SEC – $1,700
2021 SJSU Gail Fullerton Endowment – $1,000
2021 SJSU XILINX Scholarship – $2,500

References

Prof. Fan Lai
Assistant Professor of Computer Science at the University of Illinois Urbana-Champaign,
Email: fanlai@illinois.edu.

Prof. Genya Ishigaki
Assistant Professor of Computer Science at San Jose State University, Email: genya.ishigaki@sjsu.edu.

Prof. Arthi Padmanabhan
Assistant Professor of Computer Science at Harvey Mudd College, Email: arpadmanabhan@g.hmc.edu.

Prof. Ben Reed
Associate Professor of Computer Engineering at San Jose State University, Email: ben.reed@sjsu.edu.