

# JIANCHANG SU

✉ jianchang.su@uconn.edu · 🌐 <https://jianchang.su> · 📍 Storrs, CT

## Education

---

### University of Connecticut, USA

Ph.D. in Computer Science and Engineering

Aug. 2023 – Present

Advisor: Prof. Wei Zhang

### Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

M.S. in Artificial Intelligence

Apr. 2021 – Aug. 2023

### Shanghai Normal University, China

B.E. in Electronic Information Engineering

Sep. 2016 – Jun. 2020

## Research Interest

---

My research focuses on building efficient, scalable, and trustworthy infrastructure for ML/LLM systems, spanning distributed training and serving, systems-level performance optimization, and secure computing in cloud environments.

## Selected Projects

---

- **Trusted Agent Function Call** (current) – Designing a runtime trust framework for multi-agent LLM function calling across multi-cloud environments.
- **TrustWeave** – Developed a runtime attestation system for multi-cloud LLM serving using Intel TDX and Linux IMA, reducing measurement overhead by 99.95%.
- **In-Kernel Broadcast & Aggregation** – Implemented eBPF-based collective communication primitives in the Linux kernel, achieving RDMA-competitive throughput on commodity hardware.
- **PISeL** – Designed a layer-grouping pipeline for serverless DNN inference cold start, achieving 37–63% latency speedup and 59% memory reduction.
- **FastMatch** – Built a selective data caching system that eliminates CPU–GPU preprocessing bottleneck, accelerating distributed training pipelines.

## Experience

---

University of Connecticut – Research Assistant

May 2024 – Present

University of Connecticut – Teaching Assistant

Aug. 2023 – Present

CSE 3250 Introduction to Cloud Computing (3 semesters), CSE 4095 Cloud Computing, CSE 3150 C++ Essentials

## Selected Publications

---

- **J. Su\***, W. Zhang\*, Y. Zhang, K. Chu, H. Guo, Y. Lu, W. Zhang. *TrustWeave: Integrity Measurement and Attestation For Multi-Cloud LLMs*. **EuroSys '26**. (\*Co-First)  
Runtime attestation framework extending Linux IMA with Intel TDX for secure multi-cloud LLM agent deployments.
- **J. Su**, Y. Zhang, W. Zhang. *In-Kernel Aggregation and Broadcast Acceleration for Distributed Communication*. **NINEs '26**.  
Moves collective communication into the Linux kernel via eBPF, achieving RDMA-competitive throughput on commodity cloud.
- D. Xiang, W. Xu, K. Chu, T. Ding, Z. Shen, Y. Zeng, **J. Su**, W. Zhang. *Promptsculptor: Multi-agent based text-to-image prompt optimization*. **EMNLP '25**.  
Multi-agent system that iteratively optimizes text-to-image prompts for higher-quality image generation.
- **J. Su**, M. Rahimi Jafari, Y. Zhang, W. Zhang. *FastMatch: Enhancing Data Pipeline Efficiency for Accelerated Distributed Training*. **ICCD '24**.  
Selective data caching system that reduces CPU preprocessing bottleneck to better utilize GPU compute.
- M. Rahimi Jafari\*, **J. Su\***, Y. Zhang, O. Wang, W. Zhang. *PISeL: Pipelining DNN Inference for Serverless Computing*. **CIKM '24**. (\*Co-First)  
Layer-grouping pipelining for serverless DNN cold start; 37–63% speedup and up to 59% memory reduction.

## Technical Skills

---

**Languages:** C/C++, Python, Golang, JavaScript/TypeScript, Bash  
**ML/Systems:** PyTorch, TensorFlow, vLLM, CUDA, eBPF, Linux kernel  
**Infrastructure:** Docker, Kubernetes, AWS, GCP, Git, CMake, CI/CD

## Service & Awards

---

**Reviewer:** ICML 2026, NeurIPS 2026

**Awards:** Predoctoral Fellowship Tier 3 (2025), Conference Participation Award (2025)