

Jiexiao Xu

jiexiaoxu9@gmail.com | (206)769-7823 | <https://jiexiaoxu.github.io>

Education

University of Washington, Seattle, WA 2021 - 2026 (Expected)

- Combined **BS/MS** in Computer Science
- **Coursework:** Computer Graphics*, Computer Systems*, Algorithms*, Differentiable PL*, Distributed Training System*, Deep Learning, Distributed Systems, Compiler Construction (*Graduate-level)

Published Research

Scalable and Accurate Application-level Crash-Consistency Testing via Representative Testing. OOPSLA 2025

Y. Gu*, I. Neal*, **J. Xu**, S. C. Lee, A. Said, M. Haydar, J. V. Geffen, R. Kadekodi, A. Quinn, B. Kasikci

- Co-designed a representative testing framework for crash-consistency bugs in MMIO and POSIX applications, balancing scalability and accuracy.
- Developed grouping heuristics for POSIX workloads that cluster consecutive syscalls and repeated patterns, reducing the crash-state search space by 80% and achieving up to 10× speedup over prior tools.
- Built a test suite compatible with prior work and conducted baseline experiments; reproduced and identified 17 bugs in 8 production-ready applications, including 7 previously unknown issues confirmed by maintainers.

Ongoing Research Experience

Research Assistant, Gilbert Bernstein's Group, University of Washington Sept. 2023 – Present

Project: *Optimizing Rendering Pipeline in 3D Gaussian Splatting*

Advisor: Gilbert Bernstein

- Analyzed 3D Gaussian Splatting scenes and identified depth-wise clustering patterns in fragment distributions that dominated sorting and rasterization cost.
- Prototyped an alternative rendering pipeline that clusters fragments (inspired by order-independent transparency) to reduce global sorting overhead.
- Derived and Implemented backward pass for sequential K-means clustering in the CUDA.

Research Assistant, System Lab, University of Washington

June 2025 – Present

Project: *End-to-end SLO-aware Scheduling in Microservices*

Advisors: Simon Peter, Ratul Mahajan

- Modeled scheduling policies using M/M/1 priority queues to reason about goodput and tail-latency tradeoffs under overload and to guide new policy designs.
- Designed evaluation for an SLO-aware RPC scheduler that prioritizes requests using their end-to-end SLO. Built experiment infrastructure handling multi-replica, fan-out topology, etc.
- Evaluated the system using Alibaba production traces, achieving ~13% goodput improvement at the knee of the load curve and up to 20× goodput improvement under overload compared to FIFO.

Internship

C++ Software Engineer, Shunwang Tech, Hangzhou, China

June 2023 – Sept. 2023

Project: *Real-time Invisible Watermark*

Mentor: Ziqun Ye

- Profiled existing watermarking algorithms based on DWT and DCT on live streaming platforms.
- Implemented a invisible watermark algorithm on Windows that met real-time (60 fps) insertion requirements.

Skills

- **Programming:** C/C++, Rust, Python, Java, CUDA, Vulkan, SystemVerilog
- **Systems:** PyTorch, gRPC, Tokio, Tonic, Nsight Systems