Suxin Ji

1907 H Street, Union City, CA 94587 | + 1 (650) 507-9557 | suxinji@gmail.com | suxinji.github.io

Motivated and research-driven data scientist with industry and academic experience in systems programming, machine learning, and AI infrastructure. Completed an M.S.E. at the University of Pennsylvania with a 4.00 GPA and previously pursued Ph.D. studies at the University of Georgia (currently on leave). Co-author of multiple papers accepted at top-tier venues (e.g., SOSP, ICSE). Seeking admission to a competitive Ph.D. program to advance research in Large Language Models, Multi-Agent Systems, AI for Science, AI Safety.

EDUCATION

University of Pennsylvania — Philadelphia, PA
Master of Science in Engineering in Data Science — *Aug 2023 – Aug 2025*
- GPA: 4.0/4.0
- Core Courses: Big Data Analytics, Machine Learning and Statistics, Computer Systems Programming, Algorithms and Computation, Natural Language Processing, GPU Computing, Artificial Intelligence, Medical Image Analysis

University of Georgia — Athens, GA
Ph.D. Program in Computer Science (on leave) — *Aug 2024 – Sep 2025*
- GPA: 4.0/4.0
- Research Focus: Dynamic Binary Translation Optimizations

Southern Utah University — Cedar City, UT
Bachelor of Science (B.S.) Management Information Systems — *Jan 2018 - Aug 2021*
- GPA: 3.89/4.0
- Core Courses: Data Structure, Statistics, Discrete Structures, Algebra, Database, Algorithm, OOP in Java
- Awards: Honored with Summa Cum Laude and selected to the Dean's List

PUBLICATIONS

[1] Yage Hu *et al.* (including **Suxin Ji** and Wenwen Wang).
"WASIT: Deep and Continuous Differential Testing of WebAssembly System Interface Implementations."
In Proceedings of the 31st ACM Symposium on Operating Systems Principles (SOSP 2025), October 13–16, 2025, Seoul, Republic of Korea.

[2] Wen Zhang *et al.* (including **Suxin Ji** and Wenwen Wang).
"BFix: Automated Safe Memory-Leak Fixing for Binary Code."
In Proceedings of the 48th IEEE/ACM International Conference on Software Engineering (ICSE 2026), April 12–18, 2026, Rio de Janeiro, Brazil. (to appear)

WORKS IN PROGRESS (INDEPENDENT)

[3] "*TLRS-DPO: Token-Level Risk Scoring for Better Direct Preference Optimization.*"
Manuscript in preparation. *Target venue: Forty-Third International Conference on Machine Learning (ICML 2026).*
(Submission deadline: January 28, 2026.)

[4] "*AutoQSys: Automated QEMU Syscall Testing Framework.*"
Manuscript in preparation. Target venue: *ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA 2026).*
(Submission deadline: January 29, 2026.)

RESEARCH INTERESTS

Large Language Models, Reinforcement Learning, Multi-Agent Systems, AI for Science, AI Safety, Systems and Program Optimization, Runtime Systems and Program Analysis

RESEARCH EXPERIENCE

University of Georgia Athens, GA
Research Assistant *Aug 2024 – Dec 2024*
Research Advisor: Prof. Wenwen Wang

- Conducted research on system call emulation correctness in QEMU, focusing on Dynamic Binary Translation optimizations.
- Designed and implemented an independent syscall testing framework to compare native and QEMU-emulated behavior, with automated test generation, execution, and result analysis.
- Identified syscall emulation inconsistencies and limitations in existing testing tools, and developed mechanisms to uncover previously undetected QEMU bugs.

TEACHING EXPERIENCE

University of Georgia Athens, GA
Graduate Teaching Assistant *Jan 2025 – Jul 2025*

- CSCI 1302: Software Development
  Instructor: Prof. Bradley James Barnes
  Assisted in teaching Java-based object-oriented programming, modular software design, and data structures. Provided in-person office hours and project-based support.
- CSCI 3360E: Data Science
  Instructor: Prof. Sami Menik
  Supported instruction in Python programming, machine learning, dimensionality reduction, and data visualization. Answered student questions, clarified technical concepts, and graded assignments with timely feedback.

PROFESSIONAL EXPERIENCE

Data Witching Fremont, CA
Solutions Engineer (Broadcom Project) *Oct 2025 – Present*

- Support AI projects with Broadcom by migrating data workflows from Alteryx to Incorta Data Studio.
- Collaborate directly with clients to analyze business requirements and design AI-driven business solutions leveraging Incorta.
- Extract, transform, and model data using Incorta schemas, formulas, materialized views, and analyzer tables.
- Write test plans and scripts, analyze discrepancies, and conduct thorough solution testing.
- Train business users and coordinate cross-team communication for project success.

Incorta San Mateo, CA
Machine Learning Engineer *Nov 2021 – Nov 2022*

- Developed and deployed machine learning solutions for business optimization using Python, R, and PySpark.
- Designed and documented the Incorta DataPrep library to enable robust data manipulation.
- Conducted software integration testing and provided feedback to enhance product functionality.
- Contributed extensively to the Incorta Community by recording tutorials and answering customer queries.

Incorta San Mateo, CA
Technical Intern *Jan 2021 – Apr 2021*

- Tested Incorta's external notebook data API using Jupyter Notebook.
- Developed time series forecasting models using ARIMA and Facebook Prophet.
- Created product feature demos with data visualizations.
- Documented ML pipeline instructions via blog posts with 100+ daily views.

Incorta San Mateo, CA
Intern *Jun 2019 – Aug 2019*

- Designed and developed schemas and dashboards for monitoring Incorta data loading jobs.
- Built and refactored Python log parsers to unify data extraction tools across versions.

PROFESSIONAL SERVICE

Artifact Evaluation Committee Member, *IEEE/ACM International Symposium on Code Generation and Optimization (CGO 2025, CGO 2026)*

PROJECT EXPERIENCE

Kiwi - Database-Driven Web App Apr 2024
Database & Information Systems, University of Pennsylvania
- Designed and implemented a full-stack web app with SQL-backed data management and custom query functionalities.
- Developed both client and server modules for user interaction, data retrieval, and persistent storage.

Redfin Housing Market Prediction Dec 2023
Big Data Analytics, University of Pennsylvania
- Preprocessed large-scale housing market data using KNN imputation, outlier removal, and PCA.
- Built and evaluated predictive models including linear regression (Spark & scikit-learn), random forest, gradient boosting, and neural networks.

Kaggle Competition - Commonlit Readability Prize Jun 2021
Goal: Rating the complexity of reading passages for grade K 12 classroom use, optimizing reading selections, and benefiting reading skills
- Led in a team of four in Kaggle competition in PyTorch on Google Colab platform.
- Implemented a BERT model for NLP regression tasks in pretraining the passages and predicting the readability.
- Achieved top 1% scores in the leaderboard and won the silver medal.

Capstone Project - School Learning Management System Jan 2021
Goal: Improving the student and teacher performance with unstructured big data analysis
- Analyzed the requirements and functionality of the Canvas LMS using Canvas Docker image deployed in Azure Cloud.
- Developed ETL process for data manipulation and warehousing through PostgreSQL database.
- Built dashboards and reports for various dimensions for teachers, students, and administrators using Incorta.

CERTIFICATIONS

- Mathematics for Machine Learning: Linear Algebra (Coursera)
- Mathematics for Machine Learning: Multivariate Calculus (Coursera)
- Machine Learning (Coursera)
- Python Data Structures (Coursera)
- Applied Data Science (Coursera)
- Tableau Desktop (Tableau)
- Alibaba Cloud Certified Associate (Alibaba Cloud)

TECHNICAL SKILLS

- **Languages:** Python, Java, C/C++, SQL, R
- **Frameworks & Tools:** Spark, PyTorch, CUDA, NumPy, Pandas, Scikit-learn, Docker, Git, Linux, Bash, MySQL, PostgreSQL, Tableau, Excel, Incorta
- **Cloud Platforms:** Azure, GCP, Alibaba Cloud