

Xinye Zhao

xzhao24@nd.edu | xinye-zhao.github.io | github.com/Xinye-Zhao | linkedin.com/in/xinye-zhao

EDUCATION

University of Notre Dame

PhD in Computer Science and Engineering, GPA: 3.92

South Bend, IN

Expected May 2029

Georgia Institute of Technology

Master of Science in Electrical & Computer Engineering, GPA: 4.0

Atlanta, GA

August 2020 – July 2024

Dalian Maritime University

Bachelor of Engineering in Electrical Engineering, GPA: 3.95

Dalian, China

September 2016 – July 2020

SELECTED PUBLICATIONS

- **Xinye Zhao**, Joanna C. S. Santos, “Towards Budget-Aware Early Candidate Selection for Vulnerability Analysis,” accepted to **EXPRESS 2026**, 2026.
- **Xinye Zhao**, Joanna C. S. Santos, “TriVul: Improving Precision of Static Vulnerability Detection through Multi-Agent Reasoning,” accepted to **SAS 2026**, 2026.
- ML Siddiq, **Xinye Zhao**, VC Lopes, B Casey, J Santos, “Security in the Age of AI Teammates: An Empirical Study of Agentic Pull Requests on GitHub,” *arXiv preprint arXiv:2601.00477*, 2026.
- **Xinye Zhao**, Spyridon Mastorakis, “SemShareKV: Efficient KVCache Sharing for Semantically Similar Prompts via Token-Level LSH Matching,” **IJCNLP 2025**, 2025.
- Ziqi Zhang, **Xinye Zhao**, Mehak Bindra, Peng Qiu, Xiuwei Zhang, “scDisInFact: Disentangled Learning for the Integration and Prediction of Multi-Batch, Multi-Condition Single-Cell RNA-Sequencing Data,” **Nature Communications**, 2024.
- **Xinye Zhao**, Alexander Du, Peng Qiu, “scMODD: A model-driven algorithm for doublet identification in single-cell RNA-sequencing data,” **Frontiers in Systems Biology**, 2023.

RESEARCH EXPERIENCE

University of Notre Dame

Graduate Research Assistant, Computer Science and Engineering

South Bend, IN

September 2025 – Present

- Conduct research in AI for software engineering security, focusing on agentic and reasoning-based methods for vulnerability detection and analysis workflows.
- Developed *TriVul*, a multi-agent reasoning framework that improves the precision of static vulnerability detection for Java and was accepted to SAS 2026.
- Proposed a budget-aware early candidate selection approach for vulnerability analysis workflows, accepted to EXPRESS 2026.
- Built an AI-based orchestrator for the ARPA-H UPGRADE Hackathon that integrates SAST and DAST tools for medical-device software analysis.

University of Notre Dame

Graduate Research Assistant, Computer Science and Engineering

South Bend, IN

August 2024 – September 2025

- Designed *SemShareKV*, a KV-cache sharing framework for semantically similar prompts that minimizes recomputation during LLM inference.
- Introduced token-level fuzzy matching with positional information preservation to enable robust cache reuse across similar prompts.
- Achieved a 6.25x speedup and a 42% reduction in memory usage during LLM inference; published at IJCNLP 2025.

Georgia Institute of Technology

Graduate Research Assistant, Electrical and Computer Engineering

Atlanta, GA

September 2023 – August 2024

- Studied network behavior of 100+ IoT smart-speaker sessions, focusing on eavesdropping and targeted advertising.
- Built an automated testing platform with one tablet and three smart speakers to analyze voice-triggered traffic behaviors.
- Leveraged FFT-based temporal features to analyze traffic traces from pcap files.

Georgia Institute of Technology

Graduate Research Assistant, Computer Science and Engineering

Atlanta, GA

June 2022 – June 2024

- Contributed to *scDisInFact*, a VAE-based framework for batch effect removal, condition-associated key gene detection, and perturbation prediction.

- Utilized multiple encoders to disentangle gene expression data into shared and unshared biological factors.
- Integrated ELBO, MMD, classification, and group lasso losses to improve model performance.

Graduate Research Assistant, Biomedical Engineering

June 2022 – June 2023

- Developed *scMODD*, a model-based framework for doublet detection in scRNA-seq data by constructing singlet models from cell clustering.
- Modeled gene expression distributions with Negative Binomial and Zero-Inflated Negative Binomial formulations.
- Performed binary classification of singlets and artificial doublets using a multilayer perceptron.

ByteDance

Shenzhen, China

Frontend Engineer Intern, Lark Document

December 2020 – June 2021

- Built the *Popover Toolbar* for the Table component in Lark Document, improving usability and editing interactivity.
- Implemented 10+ end-to-end tests for *CodeBlock* using Mocha, Puppeteer, and Chai to improve reliability and test coverage.
- Maintained and analyzed a 200+ entry dataset to support data-informed product decisions.

TEACHING & SERVICE

- **Teaching Assistant**, Systems Programming (CSE 20289), University of Notre Dame Fall 2025
- **Teaching Assistant**, Computer Networks (CSE 30264), University of Notre Dame Spring 2025
- **Student Volunteer**, SPLASH/ISSTA 2026 2026
- **Program Committee Member**, ICDM 2025 Demo Track 2025
- **Reviewer**, HPDC 2025 2025

HONORS

- **Striving for Excellence in College and University Teaching**, Certificate of Completion 2025
- **Top Graduate (Rank 1/120)**, Dalian Maritime University 2020
- **National Scholarship (B.S.)**, awarded three consecutive years 2017, 2018, 2019

TECHNICAL SKILLS

Programming Languages: Python, Java, JavaScript, C++, SQL

ML, Security, and Systems: PyTorch, LLM inference systems, multi-agent reasoning workflows, static vulnerability analysis, Wireshark, Docker, LLVM

Data and Engineering Tools: Pandas, React, Styled-Components, Mocha, Puppeteer, Chai, Git

Research Areas: AI for Software Engineering, Software Security, LLM Systems, IoT Traffic Analysis, Computational Biology