

Luchen Zhou

zhouluchen23@nudt.edu.com | No. 109 Deya Road, Changsha,

EDUCATION

National University of Defense Technology, College of Computer

Key Laboratory of Advanced Microprocessor Chips and Systems

Changsha, China

Ph.D. student in Large Language Model (LLM) Acceleration

Sep. 2023 – Present

Xiamen University, College of Computer

Xiamen, China

B.S. in Software Engineering (Comprehensive GPA: 3.72/4.00, Rank: 2/56)

Sep. 2019 – Jun. 2023

RESEARCH EXPERIENCE

Software Framework Optimization for Long-Sequence LLM Inference

Changsha, China

National University of Defense Technology

Jun. 2025 – Dec. 2025

- Developed the Multi-Granularity Bit-Level Hierarchical (MGBH) framework to optimize inference in long-sequence LLMs by integrating bit-level aggregation and dynamic memory management. Achieved a $1.45\times$ latency reduction over dense attention and $2.89\times$ over sparse-band, compared to SOTA methods, with minimal accuracy loss in BERT experiments.

Algorithmic Optimization for Long-Sequence LLM Inference

Changsha, China

National University of Defense Technology

May 2024 – May. 2025

- Investigated techniques for accelerating inference in LLMs with long sequences. Proposed methods like SVD compression, dynamic head routing, and sliding-window strategies to optimize memory usage and computational efficiency. The paper, titled "SEDPA: SVD-Enhanced Dual-Path Attention for Efficient Inference," is currently under review for ICASSP 2026.

Enhancing Web Attack Detection with Neural Networks

Xiamen, China

Xiamen University

Jun. 2022 – Sep. 2023

- Developed integrated machine learning and deep learning models for web attack detection, leveraging Char-SVM, Char-LSTM, CNN-SVM, and CNN-Bi-LSTM architectures. Outperformed previous models by surpassing the performance of earlier methods, including CNN-LSTM, and achieved an accuracy of 99.60% in binary classification with Char-SVM.

Web Application Injection Attack Detection Using One-Class SVM

Xiamen, China

Xiamen University

Oct. 2021 – Oct. 2022

- Developed an anomaly detection-based approach using one-class SVM to detect web application injection attacks. Achieved a detection rate of 94.04% and a false positive rate of 1.62%, demonstrating the effectiveness of one-class SVM.

PUBLICATIONS

- Luchen Zhou**, Hao Sun, Yiyue Hu, Zeyu Xue, Yongwen Wang. SEDPA: SVD-Enhanced Dual-Path Attention for Efficient Inference. **ICASSP 2026(CCF-B, under review)**
- Luchen Zhou**, Wei-Chuen Yau, Y.S. Gan, Sze-Teng Liong. E-WebGuard: Enhanced neural architectures for precision web attack detection. **Computers & Security Jan 2025(SCI II, CCF-B)**
- Luchen Zhou**, Tao Lu, Xiaobo Hu. Detecting Web Application Injection Attacks Using One-Class SVM. **IEEE CCET 2022 (EI)**
- Yiyue Hu, An Hu, Hui Guo, **Luchen Zhou**, Yongzhang Nie, Runzhang Mao, Gaoyang Zhao, Xia Zhao and Yongwen Wang. TRIDENT: An End-to-End Streaming Accelerator for TriSpGEMM. **DAC (CCF-A, under review)**
- Qi Wen Gan, Wei-Chuen Yau, Y.S. Gan, Iftekhar Salam, Shihui Guo, Chin-Chen Chang, Yubing Wu, **Luchen Zhou**. DSteganoM: Deep steganography for motion capture data Expert Systems with Applications **ESWA 2024 (SCI II, CCF-B)**

INTERNSHIP

Internship at Tencent - PCG Video Technology, Platform Product Group

Aug 2020 – Oct 2020, Shenzhen China

Optimized internal data platform using Hadoop and Spark, improving data sync by 31% and processing efficiency by 8%. Filed a patent for data privacy protection, titled "Data Management Method, Device, Computer Equipment, and Storage Medium".

HONORS AND AWARDS

Technical Skills: Linux, Python, PyTorch, C, C++, Matlab, Verilog

Language Proficiency: IELTS 6.5

Main awards: Second Prize in the 2nd China-Malaysia Youth Innovation and Entrepreneurship Competition, First-Class University Scholarship for three consecutive years (2020, 2021, and 2022).

Courses: Artificial Intelligence (4.0/4.0), Computer Architecture (4.0/4.0), Computer Networks (4.0/4.0), Data Structures (4.0/4.0), Java Script (4.0/4.0), Databases (4.0/4.0), Probability Theory (4.0/4.0), Advanced Computer Networks (4.0/4.0)

Interests: Traveling, Photography, Music, Swimming, Reading