LONGGE YUAN

847 West Georgia Street Tallahassee FL ·

Phone Number: 612-442-7344

ly23a@fsu.edu

EDUCATION

09/2023 - 12/2025 **Florida State University**
Florida, Tallahassee *MASTER OF COMPUTER SCIENCE*

08/2018 – 05/2023 **Winona State University**
Minnesota, Winona *BACHELOR OF SCIENCE IN COMPUTER SCIENCE*

RELEVANT COURSEWORK

Scientific Data Compression Prediction
Course Project, Spring 2025

- *Built a predictive model to estimate lossy compression ratio using features extracted from scientific field data (e.g., variance, Hurst exponent, frequency energy ratio).*
- *Applied Random Forest and XGBoost models and evaluated using R^2 and MAE across datasets like Hurricane and XGC.*

Optimizing LLM Inference on Laptops via TinyChat Engine
Courese Project, Fall 2024

- *Implemented advanced system-level optimization techniques on a quantized LLaMA-2 model using a self-developed C++ inference engine based on TinyChat.*
- *Achieved significant latency reduction from 12.4s to 1.4s/token (88.7% improvement) on a MacBook M3 Pro (ARM architecture) without using external libraries.*
- *Proposed future acceleration strategies and demonstrated applicability to other LLMs.*

DISP-LLM: Dimension-Independent Structural Pruning for LLMs
Research Assistant to Prof. Shangqian Gao, Florida State University

- *Reproduced the proposed DISP-LLM pruning method for large language models based on the NeurIPS 2024 submission, including index-based residual bypassing and hypernetwork-guided layer width selection.*
- *Verified pruning effectiveness across OPT, LLaMA, LLaMA-2, and Phi models with different pruning ratios and zero-shot tasks.*

INTERNSHIP

Shanghai Cenoreach Technology Co., Ltd. – *Engineering Intern*
May 2023 – Sep 2023 | Tech Stack: Qt, C++, Java, Embedded System

- *Participated in developing a discharge monitoring system the high-voltage DC valve hall of converter station.*
- *Developed equipment for real-time monitoring, remote shooting, setting equipment parameters, and switching camera angles.*
- *Developed collaborative inspection control functions for multiple UV monitoring points in the valve hall to formulate, manage, and query sensor inspection plans.*
- *Used the Ebus interface to connect Digital Low-Light UV Camera and Ubuntu, read video streams from cameras, and display them.*

Shanghai GrandVision Technology Co., Ltd. – *Engineering Intern*
May 2024 – August 2024 | Tech Stack: Qt, C++, python

- *Developed on infrared and visible light video fusion.*
- *Developed video fusion GUI interfaces.*
- *According to the provided ebus python code, transcribe it into the corresponding C++ code and provide an interface.*

PUBLICATIONS

*Contributor to **ToMoE: Converting Dense Large Language Models to Mixture-of-Experts through Dynamic Structural Pruning***
Under review at ICML 2025

LANGUAGE&SKILL

- **Programming Languages:** Java, python, PHP, C++, C, Javascript
- **Databases**: MySQL, Oracle
- **Web Technologies**: HTML, CSS, NodeJS
- **Operating System**: Linux, Ubuntu, Embedded
- **Research Direction**: LLM

EXTRACURRICULAR ACTIVITIES

04/2023 As an ambassador represented the University by delivering a speech to high school.

05/2023 Actively participated in admissions meetings on behalf of the university. Translated for the Director of Admissions and introduced the University to prospective students and parents.