AOXUAN (SILVIA) ZHANG

Tel: 010-2184-3072 | LinkedIn: https://www.linkedin.com/in/aoxuan-silvia-zhang | Email: silvia@deepauto.ai

EDUCATION

Korea University, Seoul, South Korea
Major in Computer Science and Engineering; Double major in Mathematics;
March 2021 - August 2025

Hong Kong University of Science and Technology, **Hong Kong, HK**
Exchange Program in School of Science;
August 2023 - June 2024
Major in Mathematics

EXPERIENCE

DeepAuto.ai AI Engineer
Agentic AI Systems, Jun 2025 – Present

- Contributed to the development of a three-stage agentic AI workflow: compile, implement, and execute, supporting structured workflow automation.
- Assisted in building and testing modules that transform high-level workflow plans into atomic functions and execute them with LLM-powered agents.

KAIST MLAI Research Internship
Weight Generation for Large Language Models, January 2025 - June 2025

- Conducted a comprehensive literature survey on generative models for weight generation and alternative approaches to weight-space learning in neural networks.
- Ran and analyzed experimental results from existing codebases, assisted in debugging and reproducing key experiments to validate methodologies.
- Co-authored a paper (under review at ICLR 2026) proposing a VAE-based framework for merging heterogeneous large language models in a shared latent space.
- Investigated weight distribution properties (kurtosis, compressibility) and contributed to the design and implementation of latent-space fusion experiments.

Hyperparameter Optimization, January 2024 - May 2024

- Developed and implemented baseline optimization algorithms, including Bayesian Optimization with Hyperband (BOHB), Differential Evolution with Hyperband (DEHB), and Functional Surrogate-Based Optimization (FSBO) to efficiently optimize complex black-box functions.
- Evaluated the performance of these algorithms on benchmark problems and real-world applications, demonstrating their effectiveness in sample-efficient hyperparameter tuning and optimization.
- Co-authored a paper (NeurIPS 2025): *Cost-Sensitive Freeze-Thaw Bayesian Optimization for Efficient Hyperparameter Tuning*, introducing a novel cost-aware strategy to improve resource allocation in freeze-thaw Bayesian optimization.

DeepAuto.ai AI Researcher
LLM Agent on hyperparameter optimization, October 2024 - December 2024

- Conducted a literature review on state-of-the-art hyperparameter optimization techniques for large language models (LLMs).
- Analyzed existing codebases and replicated experiments to understand optimization workflows.
- Implemented and tested existing optimization methods to assess their impact on model performance.

AWS AI & ML Scholarship'22

Udacity, AI Programming with Python, July 2022 – December 2022

- Participated in the AWS DeepRacer Student League and received the AWS AI & ML Scholarship
- Having a collaborative virtual course that teaches programming tools and techniques fundamental to machine learning, with support from Udacity teachers in weekly group sessions
    - Project 1: Use a Pre-trained Image Classifier to Identify Dog Breeds
    - Project 2: Create an Image Classifier

Technical Consulting Virtual Internship Program

SAP, Forage, May 2022

- Virtual experience program participant in SAP, through Forage
- Practical task modules in assembling the data, data analysis, and presenting the results

UNIVERSITY PROJECTS

CA: 32-bit ARM Single Cycle Processor Design and Implementation Analysis

- Analyzed the design and implementation of a 32-bit ARM single-cycle processor using an FPGA.
- Developed expertise in ARM architecture, register files, control units, instruction extenders, and ALUs.
- Conducted simulation tests to validate functionality and performance.
- Demonstrated expertise in low-level programming, processor design, and simulation.

OS: Context Switching Overhead Analysis with Round-Robin Scheduling

- Analyzed context-switching overhead using the round-robin scheduling algorithm.
- Designed and implemented a benchmarking framework to measure overhead.
- Collected precise timing measurements and analyzed the impact of different time slices.
- Demonstrated expertise in low-level programming, scheduling algorithms, and performance analysis.

CSD: Mini-OS Design using Zedboard Zynq

- Designed a mini-OS for the Zedboard Zynq platform, providing a low-level interface between hardware and high-level programs.
- Implemented system initialization tasks, including stack setup for different modes, Global Interrupt Controller (GIC) configuration, private timer setup, UART configuration, and Task Control Block (TCB) setup for user programs.
- Developed task scheduling functionality using the round-robin mechanism.
- Configured the timer interval for task scheduling and implemented context switching in the Interrupt Service Routine (ISR) upon private timer interrupts.

SKILLS & INTERESTS

- Research Interests: Machine Learning, Deep Learning, Large Language Models (LLMs), Generative Models, Weight Generation, Model Optimization, Agentic AI Systems
- Framework and Library: Pytorch, Keras, Tensorflow, Scikit-learn, Numpy, Matplotlib
- Technical Proficiencies: Python, C, Matlab, C++, Verilog, R, Assembly
- Languages: English (professional proficiency), Korean (limited working proficiency), Mandarin (Native)