

## RESEARCH INTEREST

---

Accelerating LLM Inference, Parallel Computing, SW/HW Co-Design

## EDUCATION

---

- **Seoul National University** Seoul, South Korea  
*B.S. in Electrical and Computer Engineering; GPA: 3.65/4.3* *Mar. 2021 - Present*
  - Courses: Scalable High-Performance Computing, Embedded System Design, Digital System Design, Computer Organization, Operating System
- **Busan-Il Science High School** Busan, South Korea  
*Mar. 2019 - Feb. 2021*

## PROJECTS

---

- **Accelerate Pipeline-Parallel LLM Inference**  
Conducted as research for my graduation thesis.
- **Accelerate Diffusion Inference on Embedded System**  
Accelerate Diffusion Inference on Nvidia Jetson Orin Nano(CPU, GPU) and XILINX ZYNQ-7000(FPGA).  
Conducted as a term project for the Embedded System Design course.
- **Accelerate CNN based Text Classifier on GPU Cluster**  
Accelerate CNN based Text Classifier on GPU Cluster using CUDA.  
Conducted as a term project for the Scalable High-Performance Computing course.
- **Implementing CNN Accelerator**  
Implementing systolic-array CNN accelerator on FPGA using verilog.  
Conducted as a term project for the Digital System Design course.

## SKILLS

---

- **Programming** - C, C++, Python, CUDA, Verilog
- **Frameworks** - Pytorch, vLLM

## RESEARCH EXPERIENCES

---

- **MLSYS, Seoul National University** Seoul, South Korea  
*Undergraduate Research Intern (Advisor: Jiwon Seo)* *Mar. 2025 - Present*
- **AI Sys, Seoul National University** Seoul, South Korea  
*Undergraduate Research Intern (Advisor: Jinho Lee)* *Jun. 2024 - Aug. 2024*
- **AI Sys, Seoul National University** Seoul, South Korea  
*Undergraduate Research Intern (Advisor: Jinho Lee)* *Dec. 2023 - Feb. 2024*

## SCHOLARSHIP

---

- **Sooyoungro Church Scholarship**  
*Received a merit-based full-tuition scholarship, regardless of religious affiliation.* *Mar. 2021 - Feb. 2025*