Jinho Choi

System Engineer

Contact

Phone

010-4628-4501

E-mail

choijinho817@kaist.ac.kr

choijinho817@gmail.com

Github

https://github.com/jinho-choi123

Skills

Rust, Python, Pytorch, Web Development, Concurrent Programming

Languages

Korean (Native language)

English (Proficient)

Education

2019.03 – 2026.03(Expected) **School of Computing @KAIST**

Korea Advanced Institute of Science & Technology (KAIST)

Total GPA: 3.85/4.3, Credits Taken: 113/138

Research Interests

AI Accelerating Computer System

- Build efficient AI inference/training system
- Enjoy thinking a system design that can overcome the current limitation

Work/Internship Experience

(Mar 2025 – Present) **KAIST Casys Lab (Prof. Jaehyuk Huh) – Lab Intern for AI Accelerator**

- Research AI Accelerating System with efficient KV cache management

(Mar 2024 – Jul 2024) **KAIST Casys Lab (Prof. Youngjin Kwon) – Lab Intern for Secure Memory System**

- https://github.com/casys-kaist/BUDAlloc
- Research Use-After-Free memory bug detection/prevention system

(Mar 2024 – Jul 2024) **KAIST Casys Lab (Prof. Jaehyuk Huh) – Lab Intern for CXL System**

- Study efficient memory management for Tiered-Memory system

(Feb 2022 – May 2024) **SPARCS, KAIST – Full Stack Developer**

- https://zabo.sparcs.org/
- Contribute to KAIST Online-Board Service ZABO

Project Experience

(Mar 2025 – Present) **Compiler Design**

- Assignment project for KAIST CS.40200(Compiler Design)
- Design a C Compiler using Rust
- Implement IR generation feature, and apply several optimizations to the IR
- Apply Optimizations such as CFG simplification, register promotion, and global value numberings

(Mar 2024 – Jul 2024) **Safe API for Concurrent programming**

- Assignment project for KAIST CS.40301(Concurrent Programming)
- Design a Safe API for concurrent programming in Rust
- Implement lock coupled linked-list and lock-free hashtable
- Implement concurrent programming data-structure wrapper Arc(Atomic Reference Count)

(Mar 2024 – Jul 2024) **BUDAlloc, Use-After-Free detection/prevention system**

- Secure Memory System Project at KAIST Casys Lab(prof. Youngjin Kwon)
- Design a Use-After-Free memory bug prevention/detection system using One-Time-Allocator Concept
- Run SPEC CPU 2006, PARSEC 3.0 Benchmark and examine performance and memory footprint
- Suggest an optimization to minimize trie-search overhead, and implement it

(Mar 2024 – Jul 2024) **Application of NeuralODE to MNIST & CIFAR10 classification task**

- Assignment project for KAIST CS.30706(Machine Learning)
- Understand NeuralODE pytorch library(https://github.com/rtqichen/torchdiffeq)
- Train NeuralODE Solver for MNIST and CIFAR10 classification task

(Feb 2022 – May 2024) **ZABO, KAIST Online Board Service**

- Online board web service lauched at SPARCS, KAIST
- Implement Admin Panel and Advertisement Page
- Dockerize the environment and build CD pipeline
- Backend: https://github.com/sparcs-kaist/zabo-server-nodejs
- Frontend: https://github.com/sparcs-kaist/zabo-front-reactjs
- Advertisement Page: https://github.com/sparcs-kaist/zabo-boards

Computer System Study

KAIST CS.40200 – Compiler Design

- Building a C-Compiler with various optimizations

KAIST CS.40301 – Concurrent Programming

- Design a safe API, all kinds of locks and lock-free data structures, for concurrent programming

KAIST CS.30300 – Operating Systems and Lab

- Organize various design choice taken by Operating System.

KAIST CS.30101 – Computer Organization

- Deeply study about CPU architecture such as Caching, Pipelining, Logic Unit.

KAIST EE.30003 – Digital System

- Deep understanding of building Logic Unit

KAIST CS.20300 – System Programming

- Understand overall structure of Computer System

AI Study

KAIST CS.30706 – Machine Learning

- Understand ML in Bayesian perspective

Study Transformer Model

- Understand Transformer Model by reading the paper "Attention is All you need"
- Study variations of Transformer model such as Linear Transformer, ViT, and LoRA finetuning
- Personal blog posts: https://ball-blog.vercel.app/Machine-Learning-Transformer/

Understand Mamba Model

- Mathematically understand Mamba Model by reading papers
- Personal blog posts: https://ball-blog.vercel.app/Mamba/

AI Accelerator Study

- Read papers for accelerating Inference system such as "CachedAttention", "CacheBlend"
- Currently studying inference accelerating system that use Model-Context-Protocol(MCP)
- Personal blog posts: https://ball-blog.vercel.app/AI-Accelerator/

Awards

(Nov 2023) **Uni-DTHON 2nd Place Awards**

- Build Privacy-protection online business-card service
- Business card that selectively show personal information