SEONYOUNG KIM

seonyoungkim55@gmail.com
seonyoungkims.github.io

Interests

Efficient ML/AI, Model Compression, On-device AI, Hardware-Aware AI Design

Education

Korea Advanced Institute of Science and Technology (KAIST) Sep. 2019 - Feb. 2022
M.S. in Computer Science
- Thesis: Generating small anomaly detection models through distillation of long term dependency.

Hongik University Mar. 2015 - Aug. 2019
B.S.E. in Computer Engineering
- GPA: 4.24/4.50 (Ranked 3rd of 173 students)

Research Background

AI Researcher | Samsung Research Aug. 2022 - Present
- Data Intelligence Team (Upcoming, from Dec. 2025); SoC Architecture Team (Oct. 2022 - Nov. 2025)
- Focus: Model Compression, On-Device AI, LLMs, HW-SW co-design

Research Assistant | KAIST Sep. 2019 - Feb. 2022
- Database Lab, Advisor: Professor Myoungho Kim
- Focus: Knowledge Distillation, Time-series Anomaly Detection

Undergraduate Research Assistant | Hongik University Nov. 2018 - Aug. 2019
- Research Lab for Distributed INtelligence and Autonomy (DINA), Advisor: Professor Young Yoon
- Focus: AI, Data Analysis, Distributed system

Publications

In Progress

1. **RaBiT: Residual-Aware Binarization Training for Accurate and Efficient LLMs.**
   Youngcheon Yoo, **Seonyoung Kim**, Minseop Choi, Banseok Lee, Dongkyu Kim, Youngmin Kim. (2025).

2. **Bespoke LUT: Non-Linear Approximation for Integer-only Transformer Inference on NPUs.**
   Seonyoung Kim*, Jooeun Kim*, Meejeong Park, Sangjeong Lee, Hanjoo Cho, Hayoung Yoon, Heonjae Ha. (2025).

Thesis

1. **Generating small anomaly detection models through distillation of long term dependency.**
   Seonyoung Kim, Myoungho Kim. (2022). *Master's thesis*, School of Computing, KAIST.

Published

1. **Knowledge distillation for anomaly detection in multivariate time series data.**
   Seonyoung Kim, Myoungho Kim. (2021). *Oral Presentation* at *Korea Computer Congress*, Jeju, South Korea.

Projects

1. **Efficient AI Deployment on NPUs | Samsung Research** Aug. 2022 – Present
   - Developing hardware-aware model compression techniques for LLMs, focusing on ultra low-bit quantization and ongoing publications.

2. **Edge Computing–Based Anomaly Detection in Memory Semiconductor Processes | KAIST** Sep. 2020 – Sep. 2021
   - Developed an autoencoder–LSTM model for life prediction of memory semiconductors.
3. **Chemical Mechanical Planarization (CMP) Wafer Defect Detection Project | KAIST** Dec. 2019 – Jun. 2020
   - Developed wafer defect detection using vision models on CMP wafer surface images.
4. **Neouly Security Project | Hongik University** Dec. 2018 – Jun. 2019
   - Developed a malware detection system using DNN, achieving 97.8% accuracy.
5. **AI-based Restaurant Recommendation System | Hongik University** Jan. 2018 – Nov. 2018
   - Built a restaurant recommendation system using Bi-LSTM with Word2Vec.

Talks and Presentations

Internal Seminar (Journal Club) | Samsung Research Feb. 2023 - Present
- Presented 8 papers on model compression and LLMs as part of a rotating paper review seminar.

Graduate Seminar | KAIST Jan. 2019 – May. 2021
- Presented 6 talks on recent research papers, advanced methods, and my own thesis work.

Honors and Scholarship

Outstanding Teaching Assistant Award | KAIST Jun. 2020
- Recognized as an Outstanding Teaching Assistant based on top-tier student evaluations.

The Hongik scholarship | Hongik University Aug. 2015 - Sep. 2018
- Awarded $15,900 in total, covering approximately four semesters.

Korea Open Source Software Developers Hackathon | Korea IT Business Promotion Association Oct. 2016
- Contributed to the Linux perf open-source project and awarded 2nd place.

Teaching Assistant

Database System, Graduate Course | KAIST Mar. 2021 – Jun. 2021
System Programming, Undergraduate Course | KAIST Sep. 2020 – Dec. 2020
Data Structure, Undergraduate Course | KAIST Mar. 2020 - Jun. 2020

Activities

Vice President, Graduate Student Association | School of Computing, KAIST Mar. 2020 – Feb. 2021