

Seungwoo Kang

Undergraduate Student

Seoul, Republic of Korea — swkang6348@naver.com — +82-10-4828-6348

GitHub — LinkedIn — Website

Education

Hongik University

B.S. in Computer Engineering

Seoul, S. Korea

Mar 2020 – Present

- Total GPA of 3.84 / 4.5
- Major GPA of 4.01 / 4.5

Research and Internship Experience

UTL Lab, Korea University

Undergraduate Researcher

Seoul, S. Korea

Aug 2026 – Present

- Investigating offline-to-online reinforcement learning to improve the generalization of VLA models.

SCAI Lab, Chung-Ang University

Undergraduate Researcher

Seoul, S. Korea

Jan 2026 – May 2026

- Researched knowledge distillation techniques for speech self-supervised learning models to improve inference efficiency.
- Investigated training-free steering methods for improving instruction following in large language models.
- Assisted with validation and quality assurance for KoALa-Bench, a benchmark for Korean speech understanding and faithfulness evaluation in large audio language models.

Project Experience

Strategy-Aware Activation Steering for Diverse LLM Decoding

Mar 2026 – Present

Independent Research Project

- Designing SAAS, a test-time activation steering framework for strategy-aware diverse decoding.
- Developing strategy detection and steering methods based on task-specific activation clusters.
- Planning evaluations on math, coding, and creative generation tasks.

Jailbreak-Robust ZeroTuning for LLMs [GitHub] [Youtube]

May 2026

Team Leader

- Advanced Machine Learning Course Project
- Reproduced and evaluated ZeroTuning-based attention sink modulation
- Investigated the relationship between attention modulation and jailbreak robustness.

Face-Age Estimation [GitHub]

Oct 2025 – Dec 2025

Team Leader

- Basic Machine Learning Course Project
- Built a face age estimation model using ResNet/VGG backbones on the UTKFace dataset.
- Enhanced model generalization performance through data augmentation strategies

Honors and Scholarships

- Creativity Scholarship, Hongik University 2026
- Cooperation Scholarship, Hongik University 2025

Research Interests

- Large Language Models
- Robot Learning
- Robotic Manipulation

Others

English Proficiency

- TOEIC: 835 / 990

Jan 2026