Mijin Koo

: starmj09@snu.ac.kr
: https://mijinkoo.github.io/

Research Interests

Embodied AI safety; runtime safeguards and monitoring for robot policies; physical safety reasoning for real-world deployment; adversarial robustness and generative model safety.

Education

Carnegie Mellon University — Pittsburgh, PA, USA
Visiting Scholar, School of Computer Science (S3D) — Sep. 2026–Feb. 2027

Seoul National University — Seoul, Korea
Ph.D. in Intelligence and Information (Advisor: Prof. Nojun Kwak) — Sep. 2024–Present

Ewha Womans University — Seoul, Korea
B.S. in Computer Science and Engineering — Mar. 2020–Aug. 2024

Publications

[1] K. Kim, C. Kim, J. Shin, T. Kwon, J. Kim, **M. Koo**, and H. Park, "Phasor: Phase-anchored universal action representations for humanoid embodiments", *arXiv preprint arXiv:2606.01851*, 2026. [link].

[2] S. Panpatil, A. Singh, **M. Koo**, C. Kim, H. Park, and D. Choi, "Egosafetybench: A diagnostic egocentric video benchmark for evaluating embodied vlms as runtime safety guards", *arXiv preprint arXiv:2607.00218*, 2026. [link].

[3] J. Lee, **M. Koo**, and N. Kwak, "Csf: Black-box fingerprinting via compositional semantics for text-to-image models", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026, pp. 16 572–16 582. [link].

[4] H. Lee*, **M. Koo***, Y. Song, and N. Kwak, "Targeted data protection for diffusion model by matching training trajectory", in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 40, 2026, pp. 5854–5862. [link].

[5] Y. Song, J. Lee, **M. Koo**, J. Lee, and N. Kwak, "Point-to-point: Sparse motion guidance for controllable video editing", *arXiv preprint arXiv:2511.18277*, 2025. [link].

[6] M. Park, C. Park, S. Lim, **M. Koo**, H. Lee, W. W. Ro, and S. Kim, "Adversarial purification via super-resolution and diffusion", in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025, pp. 4605–4615. [link].

* Equal contribution in [4].

Research Experience

AIM Intelligence — Seoul, Korea
AI Researcher — Jan. 2026–Jul. 2026

- **Project: Runtime Safeguard for Physical AI Agents**
  - Developed a vision-language model (VLM)-based runtime safeguard for physical AI agents, focusing on real-time detection and intervention of unsafe behaviors.

- Deployed the safeguard model on a Unitree G1 humanoid robot and validated real-time safety intervention in collaboration with an on-device engineer.
- Contributed to embodied AI safety research, including EgoSafetyBench and PHASOR.

Korea Institute of Science and Technology (KIST) Seoul, Korea
Undergraduate Researcher, AI & Robotics Institute Jun. 2023–Dec. 2023

- **Project: Adversarial Purification with Diffusion and Super-Resolution Models**
  - Developed adversarial defense methods using diffusion and super-resolution models for image purification.
  - Conducted comparative analysis of purification performance across multiple adversarial attack scenarios.

Honors & Awards

IITP Government Fellowship – CMU AI Advanced Program 2026
Ministry of Science and ICT, Korea

Selected Projects

A Single Word Bypass – How Name Tokens Break Data Protection [Paper]
Course Project, Seoul National University Mar. 2025–Jun. 2025

- Identified a token-dependent failure mode in data protection methods for diffusion model personalization, showing that common name tokens can bypass existing safeguards.
- Conducted mechanistic analyses of parameter updates, cross-attention patterns, and text embeddings to characterize the source of protection failure.

CareSpoon: AI Nutrition Monitoring Service for Seniors
Capstone Design, Ewha Womans University Feb. 2023–Jun. 2023

- Developed a PyTorch-based object detection pipeline for food recognition and deployed the inference system using FastAPI.
- Constructed and curated a Korean food image dataset from AI HUB for model training and evaluation.

Teaching

Samsung Electronics
Teaching Assistant, AI Expert Program – Adversarial Attack & Defense 2024–2026

Leadership Experience

SK LOOKIE
President, Social Venture Club Feb. 2023–Feb. 2024

AI Club Euron
Manager / Academic Lead – Computer Vision study and technical seminars Mar. 2021–Aug. 2022