Saemi Moon

Ph.D. Student, Machine Learning Lab @ POSTECH
saemi@postech.ac.kr
Personal page | Google Scholar | GitHub

About Me

Research Interests: Generative AI Safety; Machine Unlearning; Privacy-Preserving Learning
My research goal is to develop safe and robust generative models. I study how generative AI systems can be designed to be reliable, controllable, and suitable for real-world deployment. While my primary contributions focus on generative vision models, I am broadly interested in extending these principles to large language models and physical AI systems, where safety and robustness concerns are especially critical.

Education

Pohang University of Science and Technology (POSTECH) Pohang, Korea
Integrated Ph.D. in Computer Science & Engineering *Feb. 2022 – Present*
– Advisor: Prof. Dongwoo Kim, Machine Learning Lab

Pohang University of Science and Technology (POSTECH) Pohang, Korea
B.S. in Computer Science & Engineering *Feb. 2017 – Feb. 2022*

Publications

** denotes equal contribution.*

In-Place Feedback: A New Paradigm for Guiding LLMs in Multi-Turn Reasoning.
Youngbin Choi*, Minjong Lee*, **Saemi Moon**, Seunghyuk Cho, Chaehyeon Chung, MoonJeong Park, Dongwoo Kim.
Preprint. arXiv

Holistic Unlearning Benchmark: A Multi-Faceted Evaluation for Text-to-Image Diffusion Model Unlearning.
Saemi Moon*, Minjong Lee*, Sangdon Park, Dongwoo Kim.
ICCV 2025. paper | code | website

Feature Unlearning for Pre-trained GANs and VAEs.
Saemi Moon, Seunghyuk Cho, Dongwoo Kim.
AAAI 2024. paper
Outstanding Paper Award at KAIA 2023

Anonymization for Skeleton Action Recognition.
Saemi Moon*, Myeonghyeon Kim*, Zhenyue Qin, Yang Liu, Dongwoo Kim.
AAAI 2023. paper | code

Experience

Oral Presentation: Privacy-Preserving AI Mar. 2023
ExploreCSR Workshop, POSTECH
– ExploreCSR is a research-focused initiative for building computing research pathways and community, supported by Google Research.
– Presented AAAI 2023/2024 papers on skeleton action anonymization and feature unlearning.

Industry Project: Anomaly Detection for Sensor Temperature Forecasting Aug. 2022 - Apr. 2024
Korea Electric Power Corporation (KEPCO)
– Developed an STGCN-based forecasting and anomaly scoring pipeline for multivariate sensor temperature time series.

Teaching Assistant, Programming and Problem-Solving Fall 2022

Teaching Assistant, POSCO AI Experts Training Program Fall 2022

Teaching Assistant, Youth AI·BigData Academy Fall 2023, Spring 2024

Teaching Assistant, Machine Learning Spring 2024

Honors & Awards

1st Place - Graduate Venture Investment Competition (VCIC Korea) Dec. 2024
Venture Capital Investment Competition (VCIC)
– Advanced to and represented POSTECH at the Global Finals in North Carolina (Apr. 2025).

Hyundai Motor Chung Mong-Koo Scholarship Sep. 2024
Hyundai Motor Chung Mong-Koo Foundation
– Doctoral fellowship awarded for excellence in AI safety research; full tuition and merit-based research support.

Research Subsidy for Ph.D. Candidates Sep. 2024
National Research Foundation of Korea (NRF)
– Two-year subsidy (≈USD 35,000 total) supporting independent doctoral research.

Outstanding Paper Award Jul. 2023
Korea Artificial Intelligence Association (KAIA)
– Recognized for "Feature Unlearning for Pre-trained GANs and VAEs."

POSTECH Alchemist Fellowship Sep. 2022
– Awarded for entrepreneurial potential (≈USD 7,500 grant).

Cheon Sin-il Scholarship 2019

BOEING Scholarship 2018, 2019

POSTECH Global Leadership Program 2017

Professional Activities

- Graduate Student Council President - Dept. of CSE, POSTECH (2025)
- Reviewer - AAAI, ICCV, CVPR