

Sehyun Lee

sehyun.lee@kaist.ac.kr | +82 10-4870-4248
sehyun-lee288.github.io | LinkedIn | Google Scholar

Research Profile

Ph.D. student at KAIST focused on model interpretability and multimodal language models, with first-work at ECCV, ICCV, NeurIPS, and Findings of EMNLP. Industry experience in multi-modal model evaluation and industrial time-series analysis.

Research areas: Explainable AI (XAI); mechanistic interpretability; concept circuits; polysemantic features; multimodal LLMs; self-supervised learning; time-series modeling and retrieval.

Education

Korea Advanced Institute of Science and Technology (KAIST) Mar 2020 – Present

Ph.D. Student in Artificial Intelligence

Kim Jaechul Graduate School of AI; Statistical Artificial Intelligence Lab

Advisor: Prof. Jaesik Choi

Ulsan National Institute of Science and Technology (UNIST) Mar 2018 – Feb 2020

M.S. in Computer Science and Engineering

Thesis: A General Compositional Operation in Random Process

Advisor: Prof. Jaesik Choi

Cumulative GPA: 3.91/4.3

Ulsan National Institute of Science and Technology (UNIST) Mar 2014 – Feb 2018

B.S. in Electrical and Computer Engineering

Cumulative GPA: 3.75/4.3

Selected Publications

Where Does the Sound Go? Tracing Acoustic Information Loss in Audio-Conditioned LLMs

Song-ha Jo, **Sehyun Lee**, Soyeon Kim, Jaesik Choi, and Sanghyuk Choi

Findings of EMNLP 2026 (accepted) | Co-first author | Paper

SPICE: Simple Polysemantic Feature Interpretation via Clustering-based Explanation

Sehyun Lee, Dahee Kwon, Damin Lee, and Jaesik Choi

ECCV 2026 | Paper

Granular Concept Circuits: Toward a Fine-Grained Circuit Discovery for Concept Representations

Dahee Kwon, **Sehyun Lee**, and Jaesik Choi

ICCV 2025 | Co-first author | Paper

Probing Network Decisions: Capturing Uncertainties and Unveiling Vulnerabilities without Label Information

Youngju Joung, **Sehyun Lee**, and Jaesik Choi

ICPRAI 2024 | Co-first author | Paper

Implicit Contrastive Representation Learning with Guided Stop-Gradient

Byeongchan Lee and **Sehyun Lee**

NeurIPS 2023 | Co-first author | Paper

UNIST SAIL System for TAC 2017 Cold Start Slot Filling

Seongwoo Lim, Soonjae Kwon, **Sehyun Lee**, Jaesik Choi

TAC 2017 system report | Paper

Technical Report Contributions

NAVER Cloud HyperCLOVA X Team — Technical reports, 2026

HyperCLOVA X 8B Omni | HyperCLOVA X 32B Think

Contribution: baseline model evaluation and benchmarking.

Industry Experience

NAVER Cloud (South Korea)

Oct 2025 - May 2026

Residency Program Participant, World Best LLM (WBL) Project

- Benchmarked vision-language models (VLMs) and large language models (LLMs) across standard evaluation suites; contributed baseline evaluations to the HyperCLOVA X 8B Omni and HyperCLOVA X 32B Think technical reports.
- Investigated audio-language model (ALM) architectures by analyzing the audio encoder, projector, and LLM separately; compared frozen audio encoders to assess downstream performance. Related work accepted to Findings of EMNLP 2026.

Gauss Labs (Seoul, South Korea; Palo Alto, USA)

Dec 2021 - Aug 2022

Applied Scientist Intern

- Developed an anomaly detection model for semiconductor process time series using Gaussian distributions to support root-cause analysis of manufacturing defects.

INEEJI (South Korea)

Aug 2020 - Feb 2021

Machine Learning Researcher

- Developed deep learning models to forecast financial time series and generate signals for downstream decision-making.

Projects

POSCO Time-Series VectorDB

Jun 2026 - Present

- Building a vector database that converts industrial sensor time-series into retrievable events and episodes, enabling semantic search over operational data, in collaboration with POSCO.

Automatic Debugging Techniques for Deep Learning with KDD

Dec 2022 - Present

- Developing explainable artificial intelligence (XAI) techniques to interpret and debug deep learning models for computer vision.

Smart Blast Furnace Project with POSCO

Mar 2018 - Dec 2019, Jun 2021 - Dec 2021

- Contributed to AI-based monitoring and control of internal and external blast furnace conditions in collaboration with POSCO.

Dacon KCB Financial Data Visualization Competition

May 2019 - Jul 2019

- The 8th KCB Financial Style Visualization Competition, Dacon, Korea.
- Achieved 4th place.

TAC 2017 Cold Start Slot Filling

2017

- Achieved 4th place in the 2017 TAC KBP competition.
- Developed a rule-based algorithm to detect the website class for the TAC Cold Start Slot Filling system.

Image Captioning System

Dec 2017

- Developed an image captioning system that extracts semantic concepts from images to generate natural-language descriptions.
- Received Naver UNIST Undergraduate Poster Award (2017).

Kaggle Two Sigma Financial Modeling Challenge

2017

- Placed 190th (Top 9.2%) out of 2071 teams.
- Predicted financial targets from anonymized features.

Talks & Presentations

AI EXPO KOREA 2026 – KAIST AI Technology Briefing

2026

Invited Talk, Technical Session I: AI Reliability, Model Interpretability, Medical AI, Data Synthesis

- Interpreting Neurons: From Understanding Concepts to Granular Concept Circuits.

KAIST XAI Tutorial Series

2024

Speaker, SAILab, KAIST AI

- Domain-Specific XAI Techniques for Time Series: attribution methods and temporal-pattern prototypes for interpreting time-series neural networks.

KAIST XAI Tutorial Series

2023

Speaker, SAILab, KAIST AI

- Perturbation-based Explanations: perturbation-based saliency methods (RISE and SISE) for interpreting model predictions.

Teaching & Mentoring

Research Mentor, KAIRI Internship Program

2025 - Present

KAIST Kim Jaechul Graduate School of AI

Teaching Assistant, AI702: Explainable Artificial Intelligence

Spring 2020

KAIST – Advisor: Prof. Jaesik Choi

Instructor, AI/DL Practical Training

Jul 2019

Samsung DS

Teaching Assistant, Advanced Machine Learning

Spring 2019

UNIST

Academic Service

Reviewer: NeurIPS 2026; ICML 2026 Workshop on Mechanistic Interpretability; BMVC 2026; ICCV 2025; CVPR 2025.