

Heehyeon Kim

Ph.D student • School of Computing, KAIST
KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea
✉ heehyeon@kaist.ac.kr | 🏠 <https://h22hyeon.github.io/>

Education

Korea Advanced Institute of Science & Technology (KAIST)
Ph.D., School of Computing
- Advisor: Professor Joyce Jiyoung Whang

Daejeon, Republic of Korea
Sep. 2024 ~ Present

Korea Advanced Institute of Science & Technology (KAIST)
M.S., School of Computing
- Advisor: Professor Joyce Jiyoung Whang

Daejeon, Republic of Korea
Sep. 2022 ~ Aug. 2024

Chonnam National University
B.S. in IoT Artificial Intelligence Convergence

Gwangju, Republic of Korea
Feb. 2016 ~ Aug. 2022

Publications

* corresponding author; † equal contributions

Beneath the Facade: Probing Safety Vulnerabilities in LLMs via Auto-Generated Jailbreak Prompts

H. Kim, K. Lee, and J. J. Whang*

Findings of the Association for Computational Linguistics: EMNLP (Findings of EMNLP), 2025.

Unveiling the Threat of Fraud Gangs to Graph Neural Networks: Multi-Target Graph Injection Attacks against GNN-Based Fraud Detectors

J. Choi, H. Kim, and J. J. Whang*

AAAI Conference on Artificial Intelligence (AAAI), 2025.

SAIF: A Comprehensive Framework for Evaluating the Risks of Generative AI in the Public Sector

K. Lee, H. Kim, and J. J. Whang*

AI for Public Missions (AIPM) Workshop at AAAI Conference on Artificial Intelligence (AAAI), 2025.

SpOT-Mamba: Learning Long-Range Dependency on Spatio-Temporal Graphs with Selective State Spaces

J. Choi, H. Kim, M. An, and J. J. Whang*

Spatio-Temporal Reasoning and Learning (STRL) Workshop at International Joint Conference on Artificial Intelligence (IJCAI), 2024.

Dynamic Relation-Attentive Graph Neural Networks for Fraud Detection

H. Kim, J. Choi, and J. J. Whang*

Machine Learning on Graphs (MLog) Workshop at IEEE International Conference on Data Mining (ICDM), 2023.

HAZE-Net: High-Frequency Attentive Super-Resolved Gaze Estimation in Low-Resolution Face Images

JS. Yun, Y. Na, H. Kim, HI Kim, SB. Yoo*

Asian Conference on Computer Vision (ACCV), 2022.

LatentGaze: Cross-Domain Gaze Estimation through Gaze-Aware Analytic Latent Code Manipulation

I. Lee, JS. Yun, H. Kim, Y. Na, SB. Yoo*

Asian Conference on Computer Vision (ACCV), 2022.

Shared Knowledge Distillation for Robust Multi-Scale Super-Resolution Networks

Y. Na†, H. Kim†, and SB. Yoo*

IET Electronics Letters, 2022.

Professional Employment and Industry Expereience

| | |
|---|-----------------------|
| GNN-based Insurance Fraud Detection <i>Kyobo Life Insurance & DPLANEX</i> | Aug. 2022 ~ Nov.2025 |
| Building a Dataset Roadmap for Stability Assessment Across AI Risk Domains <i>Telecommunications Technology Association (TTA)</i> | Oct. 2024 ~ Dec. 2024 |
| Super-Resolved Gaze Estimation <i>Electronics and Telecommunications Research Institute (ETRI)</i> | Mar. 2021 ~ Jul. 2022 |

Teaching Experience

| | |
|---|----------------------|
| CS471 Graph Machine Learning and Mining <i>Korea Advanced Institute of Science & Technology (KAIST)</i> | Spring 2024 and 2025 |
| CS376 Machine Learning <i>Korea Advanced Institute of Science & Technology (KAIST)</i> | Fall 2023 and 2025 |