Hyunseung Hwang

Email: aguno@kais.ac.kr
Mobile number: 010-4003-7805
Website: https://aguno.github.io

Professional Summary

Machine learning researcher specializing in AI interpretability, model robustness, and alignment. I stress-test the explanation methods practitioners rely on to audit high-stakes models, exposing their hidden vulnerabilities and engineering more transparent, robust alternatives. My work spans from developing explainability-first clustering frameworks to demonstrating how widely used tools like SHAP can be highly sensitive to data-engineering choices. I aim to build trustworthy AI systems where model risk is strictly managed, ensuring that automated decisions in critical domains—from AI alignment to financial services—are safe, accountable, and highly confident.

Education

Ph.D. School of Electrical Engineering, KAIST, Sep 2021 - Aug 2026
M.S., School of Electrical Engineering, KAIST, Sep 2019 -Aug 2021
B.S., School of Comuting, KAIST, Sep 2013 -Aug 2019
Langley High School, Fairfax County, VA Sep 2010- Aug 2013

WORK EXPERIENCE

New York University (NYU), Center for Responsible AI | New York, NY *Incoming Postdoctoral Researcher* (Sep 2026 – Present)

- Will conduct advanced research on the interpretability and safety of machine learning models under the supervision of Prof. Julia Stoyanovich.

Global Frontier Lab | Daejeon, South Korea *Research Collaborator* (Fall 2025)

- Collaborated with the NYU Center for Responsible AI to investigate the sensitivity and robustness of AI explainability techniques.
- Developed methodologies to stress-test explanation models, ensuring safe, confident, and defensible outputs for end-users and stakeholders.

Naver Corporation | Seongnam, South Korea *Research Intern* (Summer 2019)

- Implemented large-scale community detection algorithms (Girvan-Newman) using PySpark.
- Clustered massive unsupervised datasets to optimize and enhance the video recommendation engine.

Kiswe Mobile | *Software Engineering Intern* (Summer 2018)

- Engineered and shipped core Android features for a live video-streaming service.
- Upgraded the application architecture to meet and exceed Android Oreo (8.0+) performance and security requirements.

Published Papers

SHAP-based Explanations are Sensitive to Feature Representation

- Accepted at ACM FAccT 2025

XClusters: explainability-first clustering

- Accepted at AAAI Conference on Artififial Intelligence 2025

Open-World COVID-19 Data Visualization

- Accepted at VLDB Workshop on Data Management and Healthcare 2021