Hamin Koo

hamin2065@yonsei.ac.kr
KR +82-10-3346-4896

RESEARCH INTEREST

My research interests center on building fair and safe Large Language Models (LLMs), with a focus on robustness, bias mitigation, and reliable evaluation across diverse languages and usage scenarios. To achieve this goal, my prior work has explored culturally informed multilingual research and optimization-based jailbreak methods for improving LLM safety.

EDUCATION

Yonsei University (M.S.)   South Korea | May. 2025 ~ Current
- Advised by Prof. Jaehyung Kim at ML^3 Lab
- Research Focus: Safety and Fairness of LLM
- Major: Artificial Intelligence

Hankuk University of Foreign Studies (B.S.)   South Korea | May. 2020 ~ Feb. 2024
- Graduated with Honors (College of Engineering)
- Major: Computer Engineering
- Total GPA: 4.39/4.5

University of California at Berkeley (Visiting Student)   United States | Jan. 2023 ~ May.2023
- Visiting Student
- Relevant coursework: Introduction to Artificial Intelligence, Introduction to Machine Learning

PUBLICATIONS

- **Hamin Koo**, Minseon Kim, and Jaehyung Kim, "Align to Misalign: Automatic LLM Jailbreak with Meta-Optimized LLM Judges", ICLR (2026)

- **Hamin Koo** and Jaehyung Kim, "EMCee: Improving Multilingual Capability of LLMs via Bridging Knowledge and Reasoning with Extracted Synthetic Multilingual Context", Under Review (2025)

- **Hamin Koo**, Minseon Kim, and Sungju Hwang, "Optimizing Query Generation for Enhanced Document Retrieval in RAG", Arxiv preprint (2024)

- **Hamin Koo**, Yuri Choi, Jongryong Choi, and Byunghwan Jeon, "Training Multi-Resnet Model for Classification of Manufacturing Products from Multi-view Images", 35th Workshop on Image Processing and Image Understanding, IPIU, 2023, Jeju, Korea (2023)

RESEARCH EXPERIENCES

Research Intern at *ML³ Lab, Yonsei University*   **Aug. 2024 ~ Feb. 2025**
- Advisor: Prof. Jaehyung Kim
- LLM, Safety, Fairness

Research Intern at *MLAI Lab, KAIST AI*   **Jul. 2023 ~ Jul. 2024**
- Advisor: Prof. Sungju Hwang
- LLM, Retrieval Augmented Generation (RAG)
- Arxiv preprint, Patent

Research Intern at *Machine Intelligence Lab, HUFS*   **Apr. 2022 ~ Jun. 2023**
- Advisor: Prof. Byunghwan Jeon
- Classification of Manufacturing products using Multi-Resnet Model
- Poster Presentation at IPIU 2023

PATENT

- **Hamin Koo**, Taewook Nam, and Sungju Hwang, "METHOD OF RECOMMENDING PROCESS RECIPE AND LEARNING METHOD OF PROCESS PREDICTOR" 10-2024-0042648
  - South Korea, Filed on Mar 28, 2024 (Pending)
  - United States (Pending)

AWARDS / SCHOLARSHIPS

- Academic Excellence Scholarship, Hankuk University of Foreign Studies, Korea (2020 & 2021)
- National Excellence Scholarship, Korea Student Aid Foundation (2022 & 2023)

TEACHING & ADVISING

- HUFS Residential College English Tutor   Mar. 2021 ~ Jun. 2021
- Open Source SW & Experiments TA   Mar. 2022 ~ Jun. 2022
- HUFS Coding Zone Tutor   Mar. 2022 ~ Jun. 2022

OTHER EXPERIENCES

- Internship in 'PINEONE SOFT'   Jul.2021 ~ Aug. 2021