Dongyoon Hahm

Email Address: hahmdong@kaist.ac.kr
Homepage: https://hahmdy.github.io

ABOUT

I am a undergraduate student in KAIST, advised by Prof. Kimin Lee. I am currently interested in safety of AI systems.

EDUCATION

Mar. 2021 ~ Present **KAIST** Daejeon, Republic of Korea
School of Computer Science
Candidate for Bachelor of Computer Science and Bio and Brain Engineering
Cumulative GPA: 3.84/4.3

PUBLICATIONS

1. Sayash Kapoor, Benedikt Stroebl, Peter Kirgis, Nitya Nadgir, Zachary S Siegel, Boyi Wei, Tianci Xue, Ziru Chen, Felix Chen, Saiteja Utpala, Franck Ndzomga, Dheeraj Oruganty, Sophie Luskin, Kangheng Liu, Botao Yu, Amit Arora, **Dongyoon Hahm**, Harsh Trivedi, Huan Sun, Juyong Lee, Tengjun Jin, Yifan Mai, Yifei Zhou, Yuxuan Zhu, Rishi Bommasani, Daniel Kang, Dawn Song, Peter Henderson, Yu Su, Percy Liang, and Arvind Narayanan, "Holistic Agent Leaderboard: The Missing Infrastructure for AI Agent Evaluation"
under review

2. **Dongyoon Hahm***, Taywon Min*, Woogyeol Jin*, Kimin Lee, "Unintended Misalignment from Agentic Fine-Tuning: Risks and Mitigation"
AAAI Conference on Artificial Intelligence (AAAI), AI Alignment Track 2026

3. **Dongyoon Hahm**, Woogyeol Jin, June Suk Choi, Sungsoo Ahn, Kimin Lee, "Enhancing LLM Agent Safety via Causal Influence Prompting",
Annual Meeting of the Association for Computational Linguistics (ACL), 2025 (findings)
4. Juyong Lee*, **Dongyoon Hahm***, June Suk Choi*, W. Bradley Knox, Kimin Lee, "MobileSafetyBench: Evaluating Safety of Autonomous Agents in Mobile Device Control"
AAAI Conference on Artificial Intelligence (AAAI), AI Alignment Track 2026
5. Juyong Lee, Taywon Min, Minyong An, **Dongyoon Hahm**, Haeone Lee, Changyeon Kim, Kimin Lee, "B-MoCA: Benchmarking Mobile Device Control Agents across Diverse Configurations"
Conference on Lifelong Learning Agents (CoLLAs), 2025
6. Woongbi Cho, **Dongyoon Hahm**, Jae Ha Yim, Jun Hee Lee, Yun Ju Lee, Dong-Gyun Kim, Yong Seok Kim, Jeong Jae Wie, " Programmable Building Blocks via Internal Stress Engineering for 3D Collective Assembly"
Advanced Materials Technologies, 2020

EXPERIENCES

- **Research Intern** at *Graduate School of AI,* KAIST Jan. 2024 ~ Present
  - Enhancing & monitoring the safety of LLM agents.
- **Machine Learning Engineer** at ACTNOVA Jul. 2023 ~ Dec. 2023
  - Develop computer vision model for animal behavior experiments.
- **Research Intern** at *School of Electrical Engineering,* KAIST Dec. 2022 ~ Jul. 2023
  - Develop quantum clustering, classification algorithm

PATENTS

1. **Dongyoon Hahm,** " Method and server for providing platform for trading at least one interior modeling file used in augmented reality", Application No. 10-2865776-0000
2. **Dongyoon Hahm**, Taegun Eom, " Apparatus for collecting fine dust by using driving force of vehicle", Registration No. 10-2136771-0000
3. Cho Hyuna, **Dongyoon Hahm**, "Notification providing system and notification providing method", Registration No. 10-1805575-0000

AWARDS AND HONORS

2025	Next-generation Engineer, Institute for Promotion of Engineering and Science of Korea
2024	Dean's List, KAIST College of Engineering
2020	KAIST President's Award
2018	Korea Petrochemical Industry Association President's Award

SERVICES

- Program committee member
  AAAI AIA 2026

EXTRACURRICULAR ACTIVITIES

- Google Developer Student Club — Sep. 2022 ~ Jul. 2023
- Entangled Quantum Society — Sep. 2022 ~ Jul. 2023
- KAIST Official Student Ambassador KAINURI — Mar. 2021 ~ Jan. 2023