Woogyeol Jin

Mail Linkedin GitHub

About

I am an incoming graduate student at KAIST, advised by Kimin Lee. My current research focuses on distillation and RL based post-training for reasoning tasks in LLM. Previously, I worked on the safety and alignment of LLM agents.

Education

Korea Advanced Institute of Science and Technology *Mar. 2020 – Feb. 2026*
B.S. in Computer Science / Electrical Engineering *(Expected)*
GPA: 3.81/4.3

Research Projects

LLM-Based Environment Shaping for Reinforcement Learning *Aug. 2025 – Current*

- Developing a code-generating agent that shapes Isaac Sim environments from natural language instructions to support RL training.

On-Policy Distillation for Mathematical Reasoning *Sep. 2025 – Current*

- Investigating the mode seeking in on-policy distillation and proposing teacher interleaving as a mitigation strategy.

Publications

C: Conference, **W**: Workshop, **P**: Preprint, *: Equal contribution

[P1] Learning Multi-View Spatial Reasoning from Cross-View Relations
Suchae Jeong, Jaehwi Song, Haeone Lee, Hanna Kim, Jian Kim, Dongjun Lee, Dong Kyu Shin, Changyeon Kim, Dongyoon Hahm, **Woogyeol Jin**, Juheon Choi, Kimin Lee
Under Review

[C2] Unintended Misalignment from Agentic Fine-Tuning: Risks and Mitigation
Dongyoon Hahm*, Taywon Min*, **Woogyeol Jin***, Kimin Lee
AAAI Conference on Artificial Intelligence (AAAI), 2026

[C1] Enhancing LLM Agent Safety via Causal Influence Prompting
Dongyoon Hahm, **Woogyeol Jin**, June Suk Choi, Sungsoo Ahn, Kimin Lee
Findings of the Association for Computational Linguistics (ACL Findings), 2025

Experiences

Undergraduate Research Intern at *RISE Lab* (Kimin Lee) *Sep. 2024 – Current*

- LLM Alignment and Safety

Undergraduate Research Intern at *CASYS Lab* (Youngjin Kwon) *Nov. 2023 – Mar. 2024*

- System for LLM Inference

Technical Skills

Languages: Python, C/C++, Rust, Scala
Developer Tools: Git, Cursor, VS Code
Libraries: NumPy, Matplotlib