Jaehyeok Lee/이재혁/李在赫

Ph.D. student
Sungkyunkwan University, Suwon, South Korea
hjl8708@skku.edu
https://jaehyeoklee-119.github.io/

Research Interests

I am broadly interested in building AI systems that can reason and act robustly and efficiently in real-world scenarios. I also explore how to equip these systems with human values and emotions.

Work Experiences

Microsoft Research Asia, Beijing, China — Jul 2025 — Dec 2025
Research Intern
Mentor: Dr. Xiaoyuan Yi

Education

Sungkyunkwan University, Suwon, South Korea — Mar 2025 — *Present*
Doctor of Philosophy, Artificial Intelligence
Advisor: Prof. JinYeong Bak

Sungkyunkwan University, Suwon, South Korea — Mar 2023 — Feb 2025
Master of Science, Artificial Intelligence — Cumulative GPA: 4.36/4.50
Advisor: Prof. JinYeong Bak

Thesis: Improving LLMs' Reasoning with Consistency-Driven Rationale Evaluation for Self-Training
Thesis committee: Prof. JinYeong Bak, Prof. Yun-Gyung Cheong, and Prof. Jongwuk Lee.

Sungkyunkwan University, Suwon, South Korea — Mar 2017 — Feb 2023
Bachelor of Science, Double Major: Computer Science and Engineering, Chemical Engineering — Cumulative GPA: 3.75/4.50

Publications

[1] Sooyung Choi*, **Jaehyeok Lee***, Xiaoyuan Yi, Jing Yao, Xing Xie, and JinYeong Bak. Unintended harms of value-aligned LLMs: Psychological and empirical insights, 2025. (ACL 2025).

[2] **Jaehyeok Lee**, Keisuke Sakaguchi, and JinYeong Bak. Self-training meets consistency: Improving LLMs' reasoning with consistency-driven rationale evaluation, 2025. (NAACL 2025).

Talks

"Unintended harms of value-aligned LLMs: Psychological and empirical insight"
Microsoft Research Asia Online Seminar, 2025.8.20

"Psychological and Empirical Insights on the Harms of Aligning LLMs with Human Values"
Microsoft Research Asia Intern Tech Talk, 2025.7.24

Teaching Experience

"Open Source Software Practice"
Teaching Assistant, Sungkyunkwan University, Spring 2025

Academic Services

W-NUT 2025 — Student Volunteer, Reviewer

Extracurricular Activities

"Optimistic, Pessimistic and Realistic of Large Language Models" KOFST, 2023
Event Assistant

Programming Skills

- Languages and Frameworks: C, C++, Python, Java, JavaScript, PyTorch, Android
- GitHub repositories: https://github.com/JaehyeokLee-119

REFERENCES

Prof. JinYeong Bak, SKKU, jy.bak@skku.edu