Hyunseok Lee

https://hyunseoklee-ai.github.io
david990330@gmail.com / hs.lee@kaist.ac.kr
Google Scholar GitHub

Research Interests

My research interests lie in building intelligence that is self-aware and safe. To this end, I focus on developing Large Language Models (LLMs) that can reason, make decisions (i.e., exhibit agentic behaviors), and ensure safety. I am also broadly interested in the continual pretraining of LLMs and their multilingual capabilities.

Keywords: LLM, LLM Reasoning, LLM-based Agents, LLM Safety

Education

Ph.D. in Artificial Intelligence — Mar. 2024 - Present
Korea Advanced Institute of Science and Technology (KAIST)
Advisor: **Jinwoo Shin**

B.S. in Electrical Engineering and Computer Science (double) — Mar. 2018 - Feb. 2024
Korea Advanced Institute of Science and Technology (KAIST)

Work Experience

NAVER Cloud, Research Intern — Feb. 2025 - Present
with **Kang Min Yoo** — Seongnam, KR

- Topic: LLM reasoning, LLM Agents, Visual LM (VLM)

Publications

* denotes equal contribution

Preprints (available upon request)

[P1] ReVISE: Learning to Refine at Test-Time via Intrinsic Self-Verification
Hyunseok Lee*, Seunghyuk Oh*, Jaehyung Kim, Jinwoo Shin, Jihoon Tack

Conferences

[C1] ReMoDetect: Reward Models Recognize Aligned LLM's Generations
Hyunseok Lee*, Jihoon Tack*, Jinwoo Shin
NeurIPS 2024
Qualcomm Innovation Fellowship

Honors

Qualcomm Innovation Fellowship Korea 2024

Invited Talks

"Large Scale LLM Training and Cloud Computing Usage" — Dec. 2024
SKT Enterprise AIX CON Online (remote)

"ReMoDetect: Reward Models Recognize Aligned LLM's Generations" — Nov. 2024
Max Planck Institute for Security and Privacy (remote)

Industrial Project

Korean Multilingual LLM Training for Thesis Searching Service — Mar. 2024 - Dec. 2024

- LLM project with Nable Communications, the web service development company. The system will be deployed at the company's thesis searching service.
- Developed a multilingual Korean LLM continually trained from Llama-3.1-8B (Korean LLM for Thesis)
- Applicated core LLM techniques in the system: (i) multilingual continual pretraining by entangling first language, (ii) data synthesize for thesis data to pretrain and post-train, and (iii) RAG-specific training.

Academic Activities

Workshop Reviewer: Reasoning and Planning for LLMs@ICLR

Teaching Assistant, "CS101: Introduction to Programming", KAIST — Spring & Fall 2023

Tech. Skills

Programming: Python, C
Machine Learning: PyTorch, TensorFlow, huggingface transformers, deepspeed

Software

Open Source: PyTorch implementation and model

- `Korean LLM for Thesis Search`
- `https://github.com/hyunseoklee-ai/ReMoDetect` [C1]

Reference

Jinwoo Shin, Professor at KAIST
Contact: `jinwoos@kaist.ac.kr`

Kangmin Yu, Research Lead at Naver Cloud
Contact: `kangmin.yoo@navercorp.com`

last update: February 2025