Haeone Lee

haeone.lee@kaist.ac.kr | Site | GitHub | Seoul, Korea

Research interest

I am interested in developing a safe and proficient decision-making agent in real-world(e.g., robots). To this end, I aim to develop a method that can extract behavioral rules from existing data efficiently and help the agent continuously self-improve. Relevant topics include imitation learning on human data and developing scalable reinforcement learning (RL) algorithms that can work both on/offline.

Keywords: imitation learning, hierarchical RL, explorations in RL, robot learning

Education

Korea Advanced Institute of Science and Technology. Feb 2025 – Present
MS in Artificial Intelligence Advisor: Kimin Lee

B.S. Degree Examination for Self-Education Aug 2020 – Nov 2021
B.S. in Computer Science.

GPA: 4.3/4.3
Grade distributed: 0.01∼3.00%
Calculated Score: 100/100

Relevant courseworks: Algorithm, Data structure, Computer network, Operating system, Database, System programming, Computer systems, Logic circuits, Artificial intelligence.

Experience

KAIST AI
Research intern (advisor: Kimin Lee) March 2024 – Feb 2025, Full-time

- Developed a method to detect misleading human feedback in RLHF by analyzing LLM-based reward models using influence functions [1].
- Developed a framework to train reinforcement learning agent in Mobile Android Platform [2].

Publications

[1] Taywon Min, **Haeone Lee**, Hanho Ryu, Youngchan Kwon, and Kimin Lee. "Understanding Impact of Human Feedback via Influence Functions". In submission to ICLR 2025, 2024.

[2] Juyong Lee, Taywon Min, Minyong Ahn, Dongyoon Hahm, **Haeone Lee**, Changyeon Kim, and Kimin Lee. "Benchmarking Mobile Device Control Agents across Diverse Configurations." ICLR 2024 Workshop on Generative Models for Decision Making **(Spotlight)**, 2024.

[3] **Haeone Lee**. "ComGAN: Toward GANs Exploiting Multiple Samples." arXiv preprint arXiv:2304.12098, 2023.

Skills

Programming languages: Python, C/C++

Technologies: PyTorch, Linux, Latex

Knowledge: Reinforcement learning, Robot learning, Computer vision, Machine learning, Deep learning

Language: Korean(native), English(highly proficient)
I have fluent working level proficiency in English without any difficulties