

EDUCATION

Pohang University of Science and Technology
Master of Science in Artificial Intelligence

Sep 2023 - Aug 2025
Pohang, South Korea

Hanyang University ERICA
Bachelor of Science in Electrical Engineering

Mar 2018 - Aug 2023¹
Ansan, South Korea

EXPERIENCE

NAVER Cloud
Research Intern (HyperCLOVA Team)

Feb 2026 - Current
Seongnam, South Korea

- * *HyperCLOVA* foundation model performance enhancement through large-scale data processing and synthetic reasoning generation, including exploration, evaluation, and optimization of diverse improvement strategies via systematic experimentation.

Natural Language Processing Group
Graduate Researcher (Advisor: Prof. Gary Geunbae Lee & Hyounghun Kim)

Sep 2023 - Aug 2025
Pohang, South Korea

- * **EMNLP 2025 Findings:** Explored self-correction mechanisms in code generation and proposed a reinforcement learning-based approach that enables multi-turn code correction through fine-grained and accumulated reward optimization.
- * **ACL 2025:** Developed training-based approaches to visual programming language generation for industrial automation, including a two-stage strategy that combines retrieval-augmented fine-tuning and preference learning.
- * **NAACL 2025:** Explored knowledge-injected prompt compression for retrieval-augmented QA, focusing on automatically generating domain knowledge and integrating it into the context compression process.

Efficient Learning Lab
Undergraduate Researcher (Advisor: Prof. Jaeho Lee)

Jan 2023 - Feb 2023
Pohang, South Korea

- * Investigated sparse neural networks to study how pruning affects accuracy-efficiency trade-offs; implemented a toy experiment inspired by *SMC-Bench*, highlighting the limitations of standard pruning methods.

RESEARCH INTERESTS

In general, my research focuses on language models, specifically on logical reasoning grounded in human feedback. I am interested in leveraging reinforcement learning and symbolic representations to better align models with humans. I aim to explore how to establish RL environments that enable LM assistants to not only engage in continuous learning and exploration but also cultivate the self-evolving behaviors needed for open-ended reasoning.

SELECTED PROJECTS

Logical Reasoning Jul 2024 - Present
In our *Self-Correction* paper, we proposed a novel online RL-based approach that generates Python code and, without external guidance, verifies and corrects errors in the iteratively generated code. This intrinsic capacity for self-correction was typically only observed in large models, yet we trained small models to acquire this ability, enhancing initial response quality and achieving substantial improvements through self-correction across diverse Python-based benchmarks.

Reinforcement Learning Jul 2024 - Present
Preference Learning: In our *Industrial Program Generation* paper, we developed a language-model assistant for factory automation with support from Hyundai Mobis. Because programmers in factories implement ladder programs

¹Leave of absence for mandatory military service (Jan 2019 - Aug 2020)

in widely varying styles based on individual preferences, we trained the assistant to help them write programs more consistently via preference learning. To our knowledge, this project was the first to apply language models to ladder diagram programs.

Human-in-the-Loop Learning: To enable iterative interaction with the user, we explored and developed a reward function tailored for multi-turn scenarios. My contribution involved extending the REINFORCE algorithm to multi-turn settings and implementing an accumulated reward function that aggregates rewards across the entire trajectory. The code is available *here*.

Retrieval-Augmented Generation

Sep 2023 - Sep 2024

In our *Prompt Compression* paper, we implemented a novel entity-based prompt compression method that combines predicting masked spans from the encoder with the advantage of causal generation from the decoder to mitigate hallucinations in medical-domain QA. This approach was computationally efficient and robust, even in unseen settings.

PUBLICATIONS (* → Equal contribution)

Self-Correcting Code Generation Using Small Language Models

Jeonghun Cho, Deokhyung Kang, Hyounghun Kim, and Gary Geunbae Lee

Findings of EMNLP 2025. [paper][code]

Retrieval-Augmented Fine-Tuning With Preference Optimization For Visual Program Generation

Deokhyung Kang*, Jeonghun Cho*, Yejin Jeon, Sunbin Jang, Minsub Lee, Jawoon Cho, and Gary Geunbae Lee

Proceedings of ACL 2025. [paper][press]

K-COMP: Retrieval-Augmented Medical Domain Question Answering With Knowledge-Injected Compressor

Jeonghun Cho and Gary Geunbae Lee

Proceedings of NAACL 2025. [paper][code]

Medical Domain Retrieval-Augmented Question Answering through Entity-based Context Extraction

Jeonghun Cho and Gary Geunbae Lee

Proceedings of Korea Computer Congress 2024. [paper]

*Best paper (Top 1.7%)

PATENTS

Data processing device for user prompt-based code automatic generation

Sunbin Jang, Gary Geunbae Lee, Deokhyung Kang, and Jeonghun Cho

Patent No: 10-2025-0072930, Publication Date: 2025-06-04

Method and Apparatus for Entity-Based Retrieval-Augmented Question Answering

Jeonghun Cho and Gary Geunbae Lee

Patent No: 10-2024-0133650, Publication Date: 2024-10-02

HONORS

Best Paper Award

Korea Computer Congress 2024

Jun 2024

Hanyang Brain Scholarship

Hanyang University ERICA

Spring 2023

Scholarship for Academic Excellence

Hanyang University ERICA

Fall 2022

TEACHING EXPERIENCE

Principles of Software Construction

Teaching Assistant

Spring 2025

POSCO Youth AI & Big Data Academy

Teaching Assistant

Fall 2024