Hyun Ryu

Seoul, Korea | ryuhyun1905@kaist.ac.kr | Github | Google Scholar

RESEARCH INTERESTS

LLM Reasoning and Argumentation, RL on Non-Verifiable Tasks, Natural Language Processing

EDUCATION

Carnegie Mellon University — Pittsburgh, PA
Visiting Student in Language Technologies Institute — Aug. 2025 - Sep. 2026
- Co-Advisors: Prof. Carolyn Rose, Prof. Sean Welleck

Korea Advanced Institute of Science and Technology (KAIST) — Daejeon, Korea
Master's Student in Graduate School of AI — Mar. 2025 - Present
- GPA: Overall **4.3** / **4.3**
- Advisor: Prof. Eunho Yang

B.S. in School of Electrical Engineering — Mar. 2018 - Feb. 2024
- GPA: Overall **4.23 / 4.3**, Major **4.26 / 4.3**, Ranked **1st** in the class
- Advisor: Prof. Young-Gyu Yoon

SELECTED PUBLICATIONS

- **H. Ryu*,** J. Kim*, Y. Xie, S. Kim, C. Rose, E. Yang. Eat Your Precision: Precision Objective Mitigates Reward Hacking in Rubric-based RL. arXiv 2026. (Submitted to ICLR 2027) [Paper] [Code]
- **H. Ryu*,** G. Chu*, G. Betz, E. Yang, C. Rose, S. Welleck. Argument Reconstruction as Supervision for Critical Thinking in LLMs. arXiv 2026. (Submitted to EACL 2027) [Paper] [Code] [Dataset]
- **H. Ryu**, D. Jang, H. S. Lee, …, G. Betz, E. Yang. ReviewScore: Misinformed Peer Review Detection with Large Language Models. **EMNLP 2026 Findings**. [Paper] [Code]
- **H. Ryu**, G. Kim, H. S. Lee, E. Yang. Divide and Translate: Compositional First-Order Logic Translation and Verification for Complex Logical Reasoning. **ICLR 2025**. [Paper] [Code]

FULL PUBLICATIONS

Preprint — *: equal contributions
- **H. Ryu*,** J. Kim*, Y. Xie, S. Kim, C. Rose, E. Yang. Eat Your Precision: Precision Objective Mitigates Reward Hacking in Rubric-based RL. arXiv 2026. (Submitted to ICLR 2027) [Paper] [Code]
- **H. Ryu*,** G. Chu*, G. Betz, E. Yang, C. Rose, S. Welleck. Argument Reconstruction as Supervision for Critical Thinking in LLMs. arXiv 2026. (Submitted to EACL 2027) [Paper] [Code] [Dataset]

Published
- **H. Ryu**, D. Jang, H. S. Lee, …, G. Betz, E. Yang. ReviewScore: Misinformed Peer Review Detection with Large Language Models. **EMNLP 2026 Findings**. [Paper] [Code]
- D. Jang*, Y. Kim*, C. Park, **H. Ryu**, E. Yang. Reasoning model is Stubborn: Diagnosing Instruction Overriding in Reasoning Models. **EMNLP 2026 Findings**. [Paper] [Code] [Project] [Dataset]
- **H. Ryu**, G. Kim, H. S. Lee, E. Yang. Divide and Translate: Compositional First-Order Logic Translation and Verification for Complex Logical Reasoning. **ICLR 2025**. [Paper] [Code]
- **H. Ryu**, S. Yoon, H. S. Yoon, E. Yoon, C. D. Yoo. SimPSI: A Simple Strategy to Preserve Spectral Information in Time Series Data Augmentation. **AAAI 2024**. [Paper] [Code] [Video]
- **H. Ryu**, and J. Choi. EMC^2-Net: Joint Equalization and Modulation Classification based on Constellation Network. **ICASSP 2023**. [Paper] [Code] [Video]
- C. Shin*, **H. Ryu***, E.-S. Cho, S. Han, K.-H. Lee, C.-H. Kim, Y.-G. Yoon. Three-dimensional fluorescence microscopy through virtual refocusing using a recursive light propagation network. **Medical Image Analysis 2022**. [Paper]
- C. Shin*, **H. Ryu***, E.-S. Cho, Y.-G. Yoon. RLP-Net: A Recursive Light Propagation Network for 3-D Virtual Refocusing. **MICCAI 2021**. (Selected as KAIST's Research Highlights, Young Scientist Award, Student Travel Award, and Oral Presentation) [Paper] [Code] [Press1] [Press2]

Submitted

- S. Eom*, **H. Ryu***, S. Yoon, C. D. Yoo. AugST: A Spectro-Temporal Contrastive Data Augmentation Method for Automatic Speech Recognition. (Submitted to ICML 2024)
- **H. Ryu***, K.-Y. Park*, C. Shin, Y.-G. Yoon. Separable Fourier Convolution. (Submitted to NeurIPS 2023)

HONORS & AWARDS

- Leadership Award, College of Engineering, 2023
- 3rd Place, FameLab Korea, The Ministry of Science and ICT, 2022
- **KAIST's Research Highlights**, KAIST, 2022
- **MICCAI Young Scientist Award**, The MICCAI Society, 2021
- Dean's List, College of Engineering, 2019 Spring, 2019 Fall, 2020 Spring
- Undergraduate Fellowship (3,500 USD), Korea Foundation for Advanced Studies, 2020 - 2023
- National Fellowship for Science and Engineering (4,000 USD), Korea Student Aid Foundation, 2020 - 2023
- **KAIST Honor Fellowship**, School of Electrical Engineering, 2020
- **KAIST Presidential Fellowship (8,000 USD)**, KAIST, 2020 - 2023
- **Alumni Academic Fellowship (14,000 USD)**, KAIST Alumni Association, 2019 - 2023
- Dean's List, School of Freshmen, 2018 Spring, 2018 Fall

RESEARCH TALKS AND SEMINARS

- Invited Speaker: "Training AI Reviewers", LG AI Research, Korea, Nov. 2026
- Seminar Speaker: "Mitigating Reward Hacking in Rubric-based RL", Teledia Lab @CMU LTI, USA, Sep. 2026
- Seminar Speaker: "Improving LLM's Critical Thinking Ability", L3 Lab @CMU LTI, USA, Feb. 2026
- Seminar Speaker: "Logical Reasoning with LLM", Teledia Lab @CMU LTI, USA, Oct. 2025
- Invited Speaker: "Logical Reasoning with LLM", Bilateral AI, Austria, Apr. 2025

CURRENT RESEARCH EXPERIENCE

Teledia Lab & L3 Lab @CMU LTI Pittsburgh, PA
Visiting Student (Co-Advisors: Prof. Carolyn Rose, Sean Welleck) Aug. 2025 - Sep. 2026

- **Mitigating Reward Hacking in Rubric-based RL**
  - A new reward function which combines standard rubric coverage (i.e., recall objective) with a new precision objective, which requires each unit of a model response to be both necessary and factually correct.
  - Achieves Pareto-optimal performance that balances rubric coverage and overall response quality on medical and science domains.
- **Holistic Framework that Improves LLM's Critical Thinking Ability**
  - To think critically about arguments, human learners are trained to *reconstruct* arguments to make arguments' underlying inferences explicit.
  - We (1) propose an engine that automatically *reconstruct* arbitrary arguments (GAAR), (2) synthesize a new high-quality *argument reconstruction* dataset (Arguinas) using the GAAR engine, and (3) investigate whether LLMs can enhance their critical thinking ability by learning to *reconstruct* arguments.
  - Models trained to learn argument reconstruction outperform models that do not across seven critical thinking tasks.

Superintelligence Lab @LG AI Research Seoul, Korea
Research Intern & External Collaborator (Advisor: Dr. Moontae Lee) Jul. 2025 - Oct. 2026

- **Training AI Reviewers to Assess the Soundness of Papers**
  - A framework that trains a paper soundness evaluator, where the learning objective is to jointly maximize (1) soundness and substantiveness of review points and (2) coverage of high-quality reference review points.
  - Our AI reviewer model, PaperCritic, outperforms frontier models and models trained with other AI reviewer training methods on an in-domain test set, and it also effectively generalizes to out-of-domain reviewing tasks.

Machine Learning and Intelligence Lab @KAIST Seoul, Korea
Research Assistant (Advisor: Prof. Eunho Yang) Mar. 2024 - Present

- **Misinformed Paper Review Detection with LLMs**
  - A new evaluation of peer review quality, focusing on detecting *misleading* reviews, which are either weaknesses containing incorrect premises and questions that can be already answered by the paper.
  - For automatic ReviewScore evaluation, SOTA LLMs show moderate human-model agreements (0.4-0.5 F1, 0.3-0.4 Cohen's Kappa).

- **Compositional First-Order Logic Translation and Verification for Complex Logical Reasoning**
  - A neuro-symbolic approach that enhances LLM's logical reasoning ability, by compositional translation of natural language into first-order logic (FOL) and verification of logical semantics.
  - Set a new SOTA performance on seven logical reasoning benchmarks and resolve nearly all FOL translation errors.

PAST RESEARCH EXPERIENCE

The Tearney Lab @Harvard Medical School Boston, MA
Research Intern (Advisor: Prof. Guillermo J. Tearney) Jun. 2023 - Aug. 2023

- **Advancing Dynamic Micro-Optical Coherence Tomography Imaging through Super-Resolution**
  - An advanced super-resolution algorithm that effectively extends data acquisition time and minimizes domain-specific speckle noise, which are challenges inherent in the physics of optical coherence tomography (OCT).
  - Quadrupled the spatial and temporal resolution, enhancing the capabilities of Dynamic Micro-OCT Imaging.

Artificial Intelligence & Machine Learning Lab @KAIST EE Daejeon, Korea
Undergraduate Researcher (Advisor: Prof. Chang D. Yoo) Dec. 2022 - Feb. 2024

- **SimPSI: A Simple Strategy to Preserve Spectral Information in Time Series Data Augmentation**
  - A data augmentation framework that preserves spectral information by mixing the original and augmented input spectrum, weighted by a preservation map that assigns importance scores of each frequency.
  - Mitigated the data domain dependency in time series data augmentation techniques and significantly enhanced model performance across a range of time series tasks.

Intelligent Communication Systems Lab @KAIST EE Daejeon, Korea
Undergraduate Researcher (Advisor: Prof. Junil Choi) Apr. 2022 - Oct. 2022

- **EMC^2-Net: Joint Equalization and Modulation Classification based on Constellation Network**
  - A deep learning-based, interpretable modulation classification technique that unifies the concepts of conventional signal processing techniques.
  - Achieved SOTA performance in multipath fading channels while maintaining computational efficiency.

Medical Solution Team @AITRICS Seoul, Korea
Research Intern (Advisor: Dr. Jay Heo) Dec. 2020 - May 2021

- **Interpretable Disease Prediction with Distribution-Agnostic AI Modeling**
  - An AI model that reduces discrepancies across varying clinical data distributions and offers explainability.
  - Incorporated into the AITRICS-VC product, assessing patient deterioration risk and sharing insights across hospital sites.

Neuro-Instrumentation & Computational Analysis Lab @KAIST EE Daejeon, Korea
Undergraduate Researcher (Advisor: Prof. Young-Gyu Yoon) Feb. 2020 - May 2023

- **RLP-Net: A Recursive Light Propagation Network for 3-D Virtual Refocusing**
  - An AI model that infers the 3-D volume from two adjacent 2-D wide-field fluorescence images through virtual refocusing, adhering to the physics of light propagation.
  - Faithfully reconstructed the 3-D volume of fluorescence microbeads, C. elegans, and Zebrafish.
- **Separable Fourier Convolution**
  - A versatile and computationally efficient building block for CNNs, designed to expand the receptive field while enabling interaction among distant pixels at full resolution, by exploiting the convolution theorem and the separability of the Fourier transform.
  - Significantly enhanced the performance of convolutional networks across various vision task benchmarks.

TEACHING EXPERIENCE

- **CS101 Introduction to Programming**, Teaching Assistant, 2020 Fall
- **CS101 Introduction to Programming**, Tutor, 2019 Spring, 2019 Fall, 2020 Spring

OUTREACH & PROFESSIONAL DEVELOPMENT

Service and Outreach

- **Korea Foundation for the Advancement of Science and Creativity**, Science Communicator Nov. 2022 - Present
- **Republic of Korea Army**, Signals Intelligence Analyst, Translator, Sergeant Jul. 2021 - Jan. 2023
- **School of Electrical Engineering**, Student President Feb. 2019 - Dec. 2020

- **International Conference for the Integration of Science, Technology and Society @KAIST**, Organizing Committee — Mar. 2018 - Aug. 2019

Peer Review

- ACL ARR 2026 May, August
- COLM 2026
- NeurIPS 2026
- ICML 2026
- ICLR 2026, 2027
- AAAI 2026
- ICASSP 2024, 2025
- GLOBECOM 2022, 2023