
My research focuses on **evaluation and post-training of LLMs on nuanced, non-verifiable tasks**: I have designed a structure that makes decision-making in dilemma situations verifiable [2], and introduced rubrics to assess and improve culture-aware moral decision-making [5] and coding agents [6]. I am also interested in **multi-agent and human-AI collaboration**: I aim to build high-fidelity simulations and interaction strategies that enable LLM agents to work effectively and safely in teams.

EDUCATION

University of Michigan <i>Ph.D. in Computer Science and Engineering (Advisor: Prof. Lu Wang)</i>	Aug 2024 - Apr 2029 (expected)
Seoul National University <i>B.S. in Computer Science and Engineering</i> <i>GPA: 4.18/4.3, Summa cum laude, rank: 1st in the department</i>	Mar 2020 - Feb 2024
Seoul Science High School <i>The school for gifted students in mathematics and science, teaching college-level math and science</i> <i>GPA: 4.18/4.3</i>	Mar 2017 - Feb 2020

SKILLS

Hands-on Experience	Rubrics as rewards, RLVR, SFT, LLM-as-a-judge, Self-improvement, Synthetic data generation, Moral reasoning, Multi-agent simulation
Tools & Frameworks	vLLM, LLaMA-Factory, Tinker, OpenRouter, Slurm

PUBLICATIONS

-
- [6] **Rubrics as Test Cases: Test-Grounded Execution-Free Verification of Coding Agents**
Ayoung Lee, Justus Mattern
Under review at ARR Oct 2026
- [5] **MET: Theory-Grounded and Culture-Aware Multilingual Moral Reasoning**
Ayoung Lee, Ryan Sungmo Kwon, Yunxiang Zhang, Yuxuan Liu, Peter Railton, Lu Wang
COLM 2026 [\[paper\]](#)
- [4] **LiveOIBench: Can Large Language Models Outperform Human Contestants in Informatics Olympiads?**
Kaifan Zou, Aaron Xiong, Yunxiang Zhang, Xinliang Frederick Zhang, Yueqi Ren, Jirong Yang, Ayoung Lee, Shitanshu Bhushan, Lu Wang
ICML 2026 [\[paper\]](#)
- [3] **Logit Arithmetic Elicits Long Reasoning Capabilities Without Training**
Yunxiang Zhang, Muhammad Khalifa, Lechen Zhang, Xin Liu, Ayoung Lee, Xinliang Frederick Zhang, Farima Fatahi Bayat, Lu Wang
Findings of ACL 2026 [\[paper\]](#)
- [2] **CLASH: Evaluating Language Models on Judging High-Stakes Dilemmas from Multiple Perspectives**
Ayoung Lee, Ryan Sungmo Kwon, Peter Railton, Lu Wang
ICLR 2026 [\[paper\]](#)

[1] On Consistency Training for Language-Based Image Editing Interface
Younghwon Lee*, Ayoung Lee*, Yeonjoon Jung, Seung-won Hwang (* denotes equal contribution)
IJCNLP-AACL 2023, Second Workshop on Natural Language Interfaces (Oral) [\[paper\]](#)

HONORS AND AWARDS (SELECTED)

Best Presentation Award, NLP @ Michigan Day	Mar 2025
Excellent Bachelor's Thesis Presentation Award	Feb 2024
Outstanding Tutor Award	Aug 2021
Presidential Science Scholarship	Mar 2020 - Feb 2024

WORK EXPERIENCE

Proximal <i>Research Intern (Host: Justus Mattern)</i>	May 2026 - Aug 2026 <i>San Francisco, United States</i>
NAVER Cloud HealthCare AI <i>Research Intern (Host: Heeyoung Kwak)</i>	Mar 2024 - Jun 2024 <i>Seongnam, South Korea</i>
Language and Data Intelligence Lab <i>Undergraduate Research Intern (Advisor: Prof. Seung-won Hwang)</i>	Mar 2023 - Feb 2024 <i>Seoul, South Korea</i>
DeepMetrics <i>Machine Learning Engineering Intern (Advisor: Prof. Hyun Oh Song)</i>	Jun 2022 - Jan 2023 <i>Seoul, South Korea</i>
Samsung Device Solutions Memory <i>Software Intern</i>	Jul 2021 - Aug 2021 <i>Hwaseong, South Korea</i>

TALKS

Injecting Math Reasoning Abilities in Language Models <i>@ Deepest, Seoul National University</i>	Jun 2024
Study Abroad Presentation for Students <i>@ Seoul Science High School</i>	May 2024

TEACHING

(EECS 497) Human-Centered Software Design and Development <i>Graduate Student Instructor</i>	Jan 2026 - Apr 2026 <i>University of Michigan</i>
(M1522.000700) Logic Design <i>Teaching Assistant</i>	Sep 2021 - Dec 2021 <i>Seoul National University</i>
(034.005) Foundation of Physics 1 <i>Peer Tutor</i>	Mar 2021 - Aug 2021 <i>Seoul National University</i>
(M1522.000700) Logic Design <i>Tutor</i>	Mar 2021 - Aug 2021 <i>Seoul National University</i>