SEUNGYEON JWA

✉seungyeonjwaresearch@gmail.com seungyeonlj seungyeonlj.github.io Google Scholar

Education

Seoul National University, Seoul, Korea Sep 2025 – Present
Master and Ph.D. integration in Interdisciplinary Program in Artificial Intelligence
Advisor: Jonghyun Choi

Seoul National University, Seoul, Korea Mar 2015 – Feb 2022
B.S. in Computer Science and Engineering
B.S. in Computational Sciences
B.S. in Food and Nutrition

Research Interests

I am broadly interested in *training and evaluating language models with systematic, data-centric approaches*. Specific research directions include:

- Automated and reliable evaluation methods for LLMs
- Data synthesis for training LLMs

Research Works

Seungyeon Jwa, Daechul Ahn, Reokyoung Kim, Dongyeop Kang, Jonghyun Choi. **Becoming Experienced Judges: Selective Test-Time Learning for Evaluators.** *Preprint, 2025. (under review)*

Junsoo Park*, Seungyeon Jwa*, Meiying Ren, Daeyoung Kim, Sanghyuk Choi. **OffsetBias: Leveraging Debiased Data for Tuning Evaluators.** *Findings of EMNLP 2024*

Seungyeon Jwa*, Young-rok Cha*, Moonsu Han, Donghoon Shin. **Mitigating Hate Speech in Korean Open-domain Chatbot using CTRL.** *HCLT 2021*

Industrial Experience

NCSOFT, Seongnam, Korea
NLP Researcher, Application LM Team (Alignment Team) Feb 2024 – Feb 2025

- Identified systematic biases in LLM-based evaluation by analyzing failure patterns of generative judges, and designed a debiasing data construction pipeline to mitigate these biases.
- Trained a bias-robust generative judge model using the constructed data, improving the reliability of automatic LLM evaluation.

NLP Researcher, Dialogue Model Team Jan 2022 – Jan 2024

- Constructed instruction sets for training dialogue models and trained them to align with user intentions and system personas, resulting in more controllable and consistent dialogue behavior.
- Developed an emotional state simulator for a dialogue system, enabling emotion-aware response generation.

NLP Research Intern, Dialogue Model Team Jan 2021 – Aug 2021

- Conducted experiments on controllable dialogue generation and proposed a control-token-based method to mitigate hate speech, reducing unsafe outputs in dialogue models.

Academic Experience

Machine Perception and Reasoning Lab, Seoul National University Mar 2025 – Aug 2025
Research Intern
Advisor: Prof. Jonghyun Choi

- Explored sequential test-time scaling approaches for LLM-based evaluation.

Cognitive Computing Lab, Seoul National University Jul 2020 – Aug 2020
Research Intern
Advisor: Prof. Gahgene Gweon

- Collected math problem data and trained a classifier that distinguishes problem solution types.

Scholarships

SNU Development Fund Scholarship 2019
Merit-based scholarship by Donggok Lee Yong-hee Foundation Funds

Academic Excellence Scholarship, Seoul National University 2018
Merit-based scholarship

Entrance Scholarship, Seoul National University 2015
Merit-based scholarship