

Juyoung Suk

Member of Technical Staff at Trillion Labs · Seoul, South Korea

juyoung.suk@trillionlabs.co | juyoung.site | [Google Scholar](https://scholar.google.com/citations?user=juyoung.suk) | [GitHub](https://github.com/juyoung-suk) | [LinkedIn](https://www.linkedin.com/in/juyoung-suk)

Education

Korea Advanced Institute of Science and Technology (KAIST) *Mar. 2024 - Feb. 2026*
M.S. in Artificial Intelligence · Advisor: Minjoon Seo
Thesis: Building Reliable Open-Source Language Model Evaluators through Weight Merging and Continual Training

Korea Advanced Institute of Science and Technology (KAIST) *Mar. 2019 - Feb. 2024*
B.S. in Computer Science

Work Experience

Member of Technical Staff, Trillion Labs — Seoul, South Korea *Nov. 2024 - Present*

- Core developer on Trillion’s foundation model lineup — the **Gravity** (MoE) and **Tri** (0.5B–70B, including Tri-21B-Think) series. Models on [HuggingFace](https://huggingface.co/trillionlabs).
- Work across pre-training and post-training: data, infrastructure, long-context training, and evaluation.

Machine Learning Engineer, [ThetaOne](https://thetaone.com) — Seoul, South Korea *Feb. 2023 - Jul. 2023*

- Engineered end-to-end ML pipeline for [Metabuddy](https://metabuddy.com), implementing RAG with LangChain and custom models (grammar error detection, reranker) for enhanced user interactions.

Machine Learning Engineer Intern, NAVER Corp. — Seoul, South Korea *Aug. 2022 - Feb. 2023*

- Enhanced hate-speech detection for AI Clean Bot 2.0 (40M+ users) using active learning.

Honors & Awards

Minister’s Award, Ministry of Science and ICT (과학기술정보통신부 장관상), Republic of Korea *Aug. 2025*

Best Paper Award, NAACL 2025 — for *The BiGGen Bench* *2025*

Selected Talks

Prometheus 2 Reward Model, Meta (RAM Talk series, hosted by Ilia Kulikov) May 2024

LLMs as Evaluators, Microsoft Research India (SNLP reading group, hosted by Sanchit Ahuja and Varun Gumma) Jun. 2024

Open-Source

[Prometheus-Eval](#), Core Developer May 2024 - May 2025

- Led development of the prometheus-eval open-source toolkit for evaluating language models with specialized LMs (1.1k+ GitHub stars).
- Architected and shipped evaluation pipelines, training scripts, and the prometheus-eval Python package.

Teaching

Teaching Assistant, Introduction to Programming with Python, KAIST MOOC Platform — Jan. 2022 - Jun. 2024
Online

Teaching Assistant, Introduction to Programming (CS101), KAIST — Daejeon, South Korea Aug. 2021 - Jun. 2022

Publications

*Reverse chronological. * denotes equal contribution. Full list on [Google Scholar](#).*

[1] On the Limits and Opportunities of AI Reviewers: Reviewing the Reviews of Nature-Family Papers with 45 Expert Scientists

Seungone Kim, Dongkeun Yoon, Kiril Gashteovski, **Juyoung Suk**, Jinheon Baek, Pranjali Aggarwal, Ian Wu, et al.

arXiv preprint, May 2026

[2] Predicting LLM Reasoning Performance with Small Proxy Model

Woosung Koh, **Juyoung Suk**, Sungjun Han, Se-Young Yun, Jamin Shin

ICLR 2026

[3] Trillion 7B Technical Report

Sungjun Han, **Juyoung Suk**, Suyeong An, Hyunguk Kim, Kyuseok Kim, Wonsuk Yang, Seungtaek Choi, Jay Shin

Technical Report, April 2025

[4] Evaluating Language Models as Synthetic Data Generators

Seungone Kim, **Juyoung Suk**, Xiang Yue, Vijay Viswanathan, Seongyun Lee, Yizhong Wang, Kiril Gashteovski, Carolin Lawrence, Sean Welleck, Graham Neubig

ACL 2025

[5] LLM-as-an-Interviewer: Beyond Static Testing Through Dynamic LLM Evaluation

Eunsu Kim, **Juyoung Suk**, Seungone Kim, Niklas Muennighoff, Dongkeun Kim, Alice Oh

Findings of ACL 2025

[6] MM-Eval: A Multilingual Meta-Evaluation Benchmark for LLM-as-a-Judge and Reward Models

Guijin Son, Dongkeun Yoon, **Juyoung Suk**, Javier Aula-Blasco, Mano Aslan, Vu Trong Kim, Shayekh Bin Islam, Jaume Prats-Cristià, Lucía Tormo-Bañuelos, Seungone Kim

arXiv preprint, October 2024

[7] The BiGGen Bench: A Principled Benchmark for Fine-grained Evaluation of Language Models with Language Models

Seungone Kim, **Juyoung Suk**, Ji Yong Cho, Shayne Longpre, Chaeun Kim, Dongkeun Yoon, Guijin Son, Yejin Cho, Sheikh Shafayat, et al.

NAACL 2025 (Best Paper Award)

[8] Prometheus 2: An Open Source Language Model Specialized in Evaluating Other Language Models

Seungone Kim*, **Juyoung Suk***, Shayne Longpre, Bill Yuchen Lin, Jamin Shin, Sean Welleck, Graham Neubig, Moontae Lee, Kyungjae Lee, Minjoon Seo

EMNLP 2024

[9] CLiCK: A Benchmark Dataset of Cultural and Linguistic Intelligence in Korean

Eunsu Kim, **Juyoung Suk**, Philhoon Oh, Haneul Yoo, James Thorne, Alice Oh

LREC-COLING 2024

Last updated: May 2026 · Korean (native) · English (professional)