

# Guijin Son

Linkedin: [linkedin.com/guijin-son](https://www.linkedin.com/company/guijin-son)

Github: [github.com/guijinSON](https://github.com/guijinSON)

Email: [spthsrbwls123@yonsei.ac.kr](mailto:spthsrbwls123@yonsei.ac.kr)

Mobile: +82-10-4991-1341

## EDUCATION

- Yonsei University** Seoul, South Korea  
*Bachelor of Economics; GPA: 3.63/4.5* February 2020 - 2024

## WORKING EXPERIENCE

- OneLineAI** Seoul, South Korea  
*Co-Founder (Full Time)* February 2023 - Present  
*Post-Trained LLMs for Korean finance. Developed agents for real-time question answering. Funded 500K USD.*
- Qraft Technologies** Seoul, South Korea  
*NLP Model Engineer (Freelancer)* August 2022 - February 2023  
*Researched aspect-based sentiment analysis for financial entities. Implemented event-based and factor-based investment strategies.*
- FuturePlay** Seoul, South Korea  
*Data Analyst (Full-time, Team Member)* January 2022 - July 2022  
*Engineered a data analysis dashboard for early-warning investment insights using AWS S3, Lambda, and Quicksight; contributed to industry-focused reports on Vertical SaaS, Hardware, and Recession trends.*

## PUBLICATIONS

- Conference Publications:**  
*FINKRX: Establishing Best Practices for Korean Financial NLP*  
**Guijin Son**, Hyunwoo Ko, Haneral Jung, Chami Hwang  
ACL 2025 (Industry).  
*Linguistic Generalizability of Test-Time Scaling in Mathematical Reasoning*  
**Guijin Son**, Jiwoo Hong, Hyunwoo Ko, James Thorne  
ACL 2025 (ORAL).  
*On the Robustness of Reward Models for Language Model Alignment*  
Jiwoo Hong, Noah Lee, Eunki Kim, **Guijin Son**, [3 authors], James Thorne.  
ICML 2025.  
*KMMLU: Measuring Massive Multitask Language Understanding in Korean.*  
**Guijin Son**, Hanwool Lee, Sungdong Kim, Seungone Kim, Niklas Muennighoff, Taekyoon Choi, Cheonbok Park, Kang Min Yoo, Stella Biderman.  
NAACL 2025.  
*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**, [20 authors], Sean Welleck, Graham Neubig, Moontae Lee, Kyungjae Lee, Minjoon Seo.  
NAACL 2025  
*MULTI-TASK INFERENCE: Can Large Language Models Follow Multiple Instructions at Once?.*  
**Guijin Son**, Sangwon Baek, Sangdae Nam, Ilgyun Jeong, Seungone Kim.  
ACL 2024  
*HAE-RAE Bench: Evaluation of Korean Knowledge in Language Models.*  
**Guijin Son**, Hanwool Lee, Suwan Kim, Huiseo Kim, Jaecheol Lee, Je Won Yeom, Jihyu Jung, Jung Woo Kim, Songseong Kim.  
LREC-COLING 2024.
- Workshop Papers:**  
*Multi-Step Reasoning in Korean and the Emergent Mirage*  
**Guijin Son**, Hyunwoo Ko, Dasol Choi.  
C3NLP@NAACL2025  
*LLM-as-a-Judge & Reward Model: What They Can and Cannot Do.*  
**Guijin Son**, Hyunwoo Ko, Hoyoung Lee, Yewon Kim, Seunghyeok Hong.  
GoodData@AAAI2025  
*KRX Bench: Automating Financial Benchmark Creation via Large Language Models.*  
**Guijin Son**, Hyunjun Jeon, Chami Hwang, and Hanearl Jung.  
FinNLP@LREC-COLING 2024  
*Beyond Classification: Financial Reasoning in State-of-the-Art Language Models*  
**Guijin Son**, Hanearl Jung, Moonjeong Hahm, Keonju Na, Sol Jin.  
FinNLP@IJCAI2023  
*Removing Non-Stationary Knowledge From Pre-Trained Language Models for Entity-Level Sentiment Classification in Finance.*  
**Guijin Son**, Hanwool Lee, Nahyeon Kang, Moonjeong Hahm.  
MUFFIN@AAAI 2023

- **Preprints:**

*Pushing on Multilingual Reasoning Models with Language-Mixed Chain-of-Thought*

**Guijin Son**, Donghun Yang, Hitesh Laxmichand Patel, Amit Agarwal, Hyunwoo Ko, Chanuk Lim, Srikant Panda, Minhyuk Kim, Nikunj Drolia, Dasol Choi, Kyong-Ha Lee, Youngjae Yu

*When AI co-scientists fail: SPOT-a benchmark for automated verification of scientific research*

**Guijin Son**, Jiwoo Hong, Honglu Fan, Heejeong Nam, Hyunwoo Ko, Seungwon Lim, Jinyeop Song, Jinha Choi, Gonçalo Paulo, Youngjae Yu, Stella Biderman.

*Understand, Solve and Translate: Bridging the Multilingual Mathematical Reasoning Gap*

Hyunwoo Ko, **Guijin Son**, Dasol Choi.

*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.

## INVITED TALKS

---

- LLM FineTuning (SNUH) - September 2025; Seoul, South Korea
- Reasoning LLMs in Korean (KISTI) - September 2025; Daejeon, South Korea
- LLMs in Resource-Constrained Environments (Wandb) - February 2025; Seoul, South Korea
- Ko-R1 (ModuLabs) - February 2025; Seoul, South Korea
- Language Model Evaluation (LG Electronics) - August 2024; Seoul, South Korea
- Best Practices for Korean Evaluation (LangChain KR) - February 2024; Seoul, South Korea
- LLMs for Finance (Korea Financial Telecommunications & Clearings Institute) - July 2024; Seoul, South Korea
- Finance-Specific Language Models (SHINYOUNG SECURITIES) - March 2024; Seoul, South Korea
- Best Practices for Korean Evaluation (MODUCON) - December 2023; Seoul, South Korea
- AI in Finance (ModuLabs) - June 2022; Seoul, South Korea

## LEADERSHIP EXPERIENCE

---

- Election Committee Chairperson at Yonsei University Underwood International College - November, 2021
- Student President at Yonsei University Underwood International College Economics - April, 2021