Seokhee Hong

seokhee.hong@lgresearch.ai hongcheki@gmail.com hongcheki.github.io

AI Scientist with real hands-on experience helping build and improve foundation LLMs, contributing across multiple model release cycles through post-training, evaluation, and benchmark development.

Research Interests

Large Language Model (LLM) post-training, LLM evaluation and benchmarking, Synthetic data generation for LLMs, LLM reasoning and instruction-following

Education

M.S. in Computer Science and Engineering, Seoul National University Mar. 2021 - Aug. 2023
Advisor: Prof. Gunhee Kim

B.S. in Computer Science, Yonsei University Mar. 2014 - Feb. 2021
(Highest Honor)

Work Experiences

EXAONE Lab, LG AI Research Seoul, Korea

- *AI Scientist* Nov. 2023 - Present

  Contribute to post-training of foundation LLMs in the EXAONE series (1.2B–236B), publicly released on HuggingFace and used widely by researchers

  - **LLM Post-training: Data Synthesis & Curation**
    - Design and generate large-scale synthetic datasets for world knowledge, expert knowledge, instruction-following, and agentic search
    - Work on dataset curation and aggregation including filtering, quality control, and mixture ratio optimization
  - **LLM Evaluation & Benchmarking**
    - Conduct end-to-end model evaluation to inform release decisions, including benchmark selection and internal evaluation framework development
    - Perform cross-model comparative analysis across internal and external models
    - Co-developed and publicly released Korean expert knowledge benchmark (KMMLU-Pro/Redux)
    - Design challenging instruction-following benchmarks to stress-test advanced LLM capabilities

Vision&Learning Lab, Seoul National University Seoul, Korea

- *Research Intern* (Advisor: Gunhee Kim) Jun. 2020 - Feb. 2021

Technical Skills

Languages: Python (primary), C++

ML/LLM: PyTorch, HuggingFace, large-scale data processing

LLM: post-training, evaluation, benchmark design, synthetic data generation

Publications

K-EXAONE Technical Report
LG AI Research (*core contributor*)
Arxiv preprint, 2026

EXAONE 4.0: Unified Large Language Models Integrating Non-reasoning and Reasoning Modes
LG AI Research (*core contributor*)

Arxiv preprint, 2025

From KMMLU-Redux to KMMLU-Pro: A Professional Korean Benchmark Suite for LLM Evaluation
Seokhee Hong*, Sunkyoung Kim*, Guijin Son, Soyeon Kim, Yeonjung Hong, Jinsik Lee
EMNLP 2025 (Findings)

MANTA: A Scalable Pipeline for Transmuting Massive Web Corpora into Instruction Datasets
Heuiyeen Yeen, Seokhee Hong, Hyeongu Yun, Jinsik Lee
EMNLP 2025 (Findings)

EXAONE Deep: Reasoning Enhanced Language Models
LG AI Research (*core contributor*)
Arxiv preprint, 2025

EXAONE 3.5: Series of Large Language Models for Real-world Use Cases
LG AI Research (*core contributor*)
Arxiv preprint, 2024

EXAONE 3.0 7.8B Instruction Tuned Language Model
LG AI Research (*core contributor*)
Arxiv preprint, 2024

Who Wrote this Code? Watermarking for Code Generation
Taehyun Lee*, Seokhee Hong*, Jaewoo Ahn, Ilgee Hong, Hwaran Lee, Sangdoo Yun, Jamin Shin†, Gunhee Kim†
ACL 2024

SQuARe: A Large-Scale Dataset of Sensitive Questions and Acceptable Responses Created Through Human-Machine Collaboration
Hwaran Lee*, Seokhee Hong*, Joonsuk Park, Takyoung Kim, Meeyoung Cha, Yejin Choi, Byoung Pil Kim, Gunhee Kim, Eun-Ju Lee, Yong Lim, Alice Oh, Sangchul Park, Jung-Woo Ha
ACL 2023 **(Oral)**

KoSBi: A Dataset for Mitigating Social Bias Risks Towards Safer Large Model Application
Hwaran Lee*, Seokhee Hong*, Joonsuk Park, Takyoung Kim, Gunhee Kim, Jung-Woo Ha
ACL 2023 (Industry Track)

How Robust are Fact Checking Systems on Colloquial Claims?
Byeongchang Kim*, Hyunwoo Kim*, Seokhee Hong, Gunhee Kim
NAACL 2021

* : equal contribution / † : corresponding authors

Awards and Honors

Outstanding Reviewers, EMNLP 2024 — Nov 2024
Graduated with Highest Honor, Yonsei University, Korea — Feb 2021
National Science and Technology Scholarship (full tuition), Ministry of Science, Korea — Fall 2019, Fall 2020

Teaching Experience

Teaching Assistant of Samsung DS Reinforcement Learning — Spring 2022, Spring 2023
Teaching Assistant of Computer Vision (M1522.001000) — Fall 2021
Teaching Assistant of Graduate Artificial Intelligence (M4190.569) — Spring 2021

Academic Services

ACL ARR Reviewer 2023 - Present