

Jeonghoon Kim

Team Lead

NAVER Cloud South Korea

jeonghoon.samuel@gmail.com Google Scholar LinkedIn

About

I am a research scientist and team lead at NAVER Cloud, working on large-scale foundation models. My research focuses on understanding how pre-training, model architecture, and data shape the learning dynamics and capabilities of Transformer models. My work spans foundation model pre-training, multilingual and multimodal learning, and model compression. I aim to bridge fundamental research and large-scale model production, turning empirical insights into more efficient AI systems.

Work Experience

NAVER Cloud

Dec.2021 - Present

NAVER's cloud and AI company, building the HyperCLOVA X foundation models and NAVER Cloud Platform.

Foundation Research, Team Leader

FOCUS AREAS

- **LLM Fundamental Research** — Pre-training dynamics & Transformer architecture [1, 2, 3]
  - (Early 2026) Designed the team's first mid-training stage, introduced it into model production, and established it as a process continuously applied in production.
  - Pathfinding of HyperCLOVA X modeling (Technical Report: 2024, 2025, 2026)
- **Korean Expert Development** — In charge of improving the Korean-language capabilities of HyperCLOVA X. [8, 11, 13, 17]
- **AudioLLM Mission** — Project Manager for the development of duplex and interaction models. [12, 14, 15]

EARLIER WORK

- (2025) Omni-foundation models and multilinguality of large-scale Transformers [9]
  - Served as Mission Lead for Omni-Pathfinding.
- (2021–2023) Model compression & quantization [4, 5, 6, 7]

PROMOTIONS

Team Leader 2025
Foundation Research Team, Hyperscale AI Division

Technical Leader 2024
Foundation Research Team

Technical Leader 2023
Hyperscale AI Efficiency Team

LG Energy Solution
Nov.2020 - Nov.2021

Business Strategy, Data Scientist

Nepes
Jan.2018 - Oct.2020

Future Intelligence, Researcher

- Military service (South Korea).

Education

Ph.D. in Artificial Intelligence
Mar.2025 - 2027 (expected)

Korea Advanced Institute of Science and Technology (KAIST)

Advisor: Prof. Jinwoo Shin

Alongside my full-time role at NAVER Cloud, I am pursuing a Ph.D. in AI at KAIST under Prof. Jinwoo Shin, supported by a NAVER Cloud grant.

Publications

*equal contribution · †corresponding author

LEAD AUTHOR

1 NeurIPS 2026

The Lexical Imprint of Pre-training Data

Jeonghoon Kim*, Woojin Chung*, Woomin Song, Cheonbok Park, Nako Sung, Jinwoo Shin

2 ICML 2025

Peri-LN: Revisiting Normalization Layer in the Transformer Architecture

Jeonghoon Kim, Byeongchan Lee, Cheonbok Park, Yeontaek Oh, Beomjun Kim, Taehwan Yoo, Seongjin Shin, Dongyoon Han, Jinwoo Shin, Kang Min Yoo

3 NeurIPS 2025

Exploiting Vocabulary Frequency Imbalance in Language Model Pre-training

Woojin Chung*, **Jeonghoon Kim***

4 NAACL 2025

LRQ: Optimizing Post-Training Quantization for Large Language Models by Learning Low-Rank Weight-Scaling Matrices

Jung Hyun Lee*, **Jeonghoon Kim***, June Yong Yang, Se Jung Kwon, Eunho Yang, Kang Min Yoo, Dongsoo Lee

5 ICLR 2024 *Mentorship program*

Rethinking Channel Dimensions to Isolate Outliers for Low-bit Weight Quantization of Large Language Models

Jung Hwan Heo*, **Jeonghoon Kim***, Beomseok Kwon, Byeongwook Kim, Se Jung Kwon, Dongsoo Lee

6 NeurIPS 2023

Memory-Efficient Fine-Tuning of Compressed Large Language Models via sub-4-bit Integer Quantization

Jeonghoon Kim*, Jung Hyun Lee*, Sungdong Kim, Joonsuk Park, Kang Min Yoo, Se Jung Kwon, Dongsoo Lee

7 ICML 2023

FlexRound: Learnable Rounding based on Element-wise Division for Post-Training Quantization

Jung Hyun Lee*, **Jeonghoon Kim***, Se Jung Kwon, Dongsoo Lee

8 Arxiv (preprint, 2025)

Cross-lingual Collapse: How Language-Centric Foundation Models Shape Reasoning in Large Language Models

Cheonbok Park*, **Jeonghoon Kim***, Joosung Lee, Sanghwan Bae, Jaegul Choo, Kang Min Yoo

CO-AUTHOR

9 Technical report (Arxiv, 2026) *Lead of the Omni-pathfinding mission.*

HyperCLOVA X 8B Omni

HyperCLOVA X Team

10 Technical report (Arxiv, 2026)

HyperCLOVA X 32B Think

HyperCLOVA X Team

11 EMNLP 2026

From Unverifiable Domain Text to Cross-Lingual RLVR via Code-Switched Task Synthesis

Cheonbok Park*, Jihwan Kim*, Shinyoung Joo, Joosung Lee, Hwiyeol Jo, **Jeonghoon Kim**†, Jaegul Choo

12 EMNLP 2026

OmniACBench: A Benchmark for Evaluating Context-Grounded Acoustic Control in Omni-Modal Models

Seunghee Kim, Bumkyu Park, Kyudan Jung, Joosung Lee, Soyoon Kim, **Jeonghoon Kim**, Taeuk Kim, Hwiyeol Jo

13 Findings of EMNLP 2026

What Models Know, How Well They Know It: Knowledge-Weighted Fine-Tuning for Learning When to Say "I Don't Know"

Joosung Lee, Hwiyeol Jo, Donghyeon Ko, Kyubyung Chae, Cheonbok Park, **Jeonghoon Kim**†

14 Interspeech 2026

ParaPairAudioBench: Paralinguistic Pairwise Audio Benchmark for LALM-as-a-Judge

Jisu Jeon, Seungyeon Jwa, Joosung Lee, Jinhyeon Kim, Woojin Chung, Hwiyeol Jo, **Jeonghoon Kim**, Jonghyun Choi, Soyoon Kim

15 ACL 2026 (Industry Track)

Sommelier: Scalable Open Multi-turn Audio Pre-processing for Full-duplex Speech Language Models

Kyudan Jung, Jihwan Kim, Soyoon Kim, **Jeonghoon Kim**, Jaegul Choo, Cheonbok Park

16 Findings of ACL 2026

ReflexiCoder: Teaching Large Language Models to Self-Reflect on Generated Code and Self-Correct It via Reinforcement Learning

Juyong Jiang, Jiasi Shen, Sunghun Kim, Kang Min Yoo, **Jeonghoon Kim**†, Sungju Kim

17 Findings of ACL 2026

Enhancing Hallucination Detection via Future Context

Joosung Lee, Cheonbok Park, Hwiyeol Jo, **Jeonghoon Kim**, Joonsuk Park, Kang Min Yoo

18 Arxiv (preprint, 2025)

ReGUIDE: Data Efficient GUI Grounding via Spatial Reasoning and Search

Hyunseok Lee, **Jeonghoon Kim**, Beomjun Kim, Jihoon Tack, Chansong Jo, Jaehong Lee, Cheonbok Park, Sookyo In, Jinwoo Shin, Kang Min Yoo

19 Technical report (Arxiv, 2025) *Uses the Peri-LN architecture.*

HyperCLOVA X THINK Technical Report

HyperCLOVA X Team

20 ICLR 2024

LUT-GEMM: Quantized Matrix Multiplication based on LUTs for Efficient Inference in Large-Scale Generative Language Models

Gunho Park, Baeseong Park, Minsub Kim, Sungjae Lee, **Jeonghoon Kim**, Beomseok Kwon, Se Jung Kwon, Byeongwook Kim, Youngjoo Lee, Dongsoo Lee

21 Technical report (Arxiv, 2024)

HyperCLOVA X Technical Report

HyperCLOVA X Team

22 ICLR 2023

Winning Both the Accuracy of Floating Point Activation and the Simplicity of Integer Arithmetic

Yulhwa Kim, Jaeyong Jang, Jehun Lee, Jihoon Park, **Jeonghoon Kim**, Byeongwook Kim, Baeseong Park, Se Jung Kwon, Dongsoo Lee, Jae-Joon Kim

23 Findings of EMNLP 2022

AlphaTuning: Quantization-Aware Parameter-Efficient Adaptation of Large-Scale Pre-Trained Language Models

Se Jung Kwon, **Jeonghoon Kim**, Jeongin Bae, Kang Min Yoo, Jin-Hwa Kim, Baeseong Park, Byeongwook Kim, Jung-Woo Ha, Nako Sung, Dongsoo Lee

Service

REVIEWING

- International Conference on Machine Learning
- Neural Information Processing Systems
- International Conference on Learning Representations
- Association for Computational Linguistics

TECH BLOG

- Technical curator of the CLOVA Tech Blog (NAVER Cloud)