

Minji Kim

Research Interests: Large Language Models, Alignment, Safety, Multimodal Learning

Website: mz-kim.github.io
Email: mzkim@postech.ac.kr
minji.kim@markr.ai

EDUCATION

Pohang University of Science and Technology (POSTECH) M.S. in Artificial Intelligence, Advisor: Prof. Hyounghun Kim Thesis: Exploring Interpretable and Trustworthy Alignment Techniques for Large Language Models Our lab moved from Ulsan National Institute of Science and Technology (UNIST) to POSTECH in Feb 2025.	Pohang, South Korea Feb 2024 – Feb 2026
University of Ulsan B.S. in Computer Science, School of IT Convergence (GPA: 4.30 / 4.5)	Ulsan, South Korea Mar 2022 – Feb 2024

PUBLICATIONS

6. **Refuse without Refusal: A Structural Analysis of Safety-Tuning Responses for Reducing False Refusals in Language Models**
[Minji Kim](#) and Hyounghun Kim. Proceedings of EMNLP 2026. [\[paper\]](#)
5. **Knowing What Not to Answer: Selective Non-Compliance in Vision-Language Models**
[Minji Kim](#), Jihyoung Jang, and Hyounghun Kim. Proceedings of EMNLP 2026. [\[paper\]](#)
4. **Revealing the Inherent Instructability of Pre-Trained Language Models**
Seokhyun An, [Minji Kim](#), and Hyounghun Kim. Findings of EMNLP 2025. [\[paper\]](#)
3. **Enabling Chatbots with Eyes and Ears: An Immersive Multimodal Conversation System for Dynamic Interactions**
Jihyoung Jang, Minwook Bae, [Minji Kim](#), Dilek Hakkani-Tür, and Hyounghun Kim. Proceedings of ACL 2025. [\[paper\]](#)
2. **AVEC: Achieving Instruction-Following via a Single Vector Intervention**
[Minji Kim](#) and Hyounghun Kim. Proceedings of KAIA 2025.
1. **Detection of Sexuality and Violence in Korean News Article Titles Based on KoBERT Model**
[Minji Kim](#), Hwando Kim, Jimin Bong, and Daehwan Kim. Proceedings of ACK 2023.

RESEARCH EXPERIENCE

Markr AI AI Researcher	South Korea 2026 – Present
Language & Intelligence Lab, POSTECH Graduate Student, Advisor: Prof. Hyounghun Kim	Pohang, South Korea Feb 2024 – Feb 2026
Electronics and Telecommunications Research Institute (ETRI) Research Intern	Daejeon, South Korea Jan 2023 – Feb 2023
Digitization and reliability analysis of modern and contemporary artworks: training data construction, hand-writing detection and denoising, and Transformer-based damage detection.	
Computer Vision Lab, University of Ulsan Undergraduate Research Student	Ulsan, South Korea Oct 2022 – Dec 2022

PROJECT EXPERIENCE

DSC , Automated Table Structure Recognition and Data Extraction	Sep 2025 – Feb 2026
--	---------------------

TEACHING EXPERIENCE

Teaching Assistant, POSTECH

Introduction to Natural Language Processing (Fall 2025)

Sep 2025 – Dec 2025

Teaching Assistant, Ulsan ICT Promotion Agency

Artificial Intelligence and Metaverse

Aug 2022 – Sep 2022

OTHER EXPERIENCE

LG Electronics LDC PBL Program Assistant

Jul 2024 – Aug 2024

UNIST AI Novatus Academia PBL Program Assistant

May 2024 – Sep 2024

Naver Connect Artificial Intelligence Mentor

May 2022 – Dec 2022

HONORS & AWARDS

Grand Prize, University Student Mentor, Naver Connect

Jan 2023

Scholarship for Academic Excellence, University of Ulsan

Mar 2022 – Sep 2023

SKILLS

Programming Languages: Python, C, Java

Libraries: PyTorch, TensorFlow, Hugging Face (Transformers, PEFT), vLLM

Others: Linux, Docker, Unity, CAD, OpenCV