

Minkyung Cho

M.S. Student, Multimodal Language Processing Lab, KAIST
minkyung.cho@kaist.ac.kr | minkyung.me | Google Scholar | GitHub

Education

2025 -	M.S. in Culture Technology Natural language processing at the Multimodal Language Processing (MLP) Lab, advised by Prof. KyungTae Lim	KAIST
2020 - 2025	B.S. in Mathematical Sciences B.S. in Computer Science and Engineering College of Liberal Studies	Seoul National University

Publications

Peer-Reviewed Conference Papers

2026	Hidden Threat in Synthetic Data: Covert Targeted Bias Injection through Benign Text Minkyung Cho , Jihyo Kim, SeungWoo Song, Junghun Yuk, Minjoon Kee, Hoyun Song, and KyungTae Lim Conference on Empirical Methods in Natural Language Processing (EMNLP), to appear
2026	MENTOR: Reinforcement Learning via Flexible Teacher-Optimized Rewards for Tool-Use Distillation ChangSu Choi, Hoyun Song, Dongyeon Kim, Minkyung Cho , WooHyeon Jung, SunJin Park, NohHyeob Bae, Seona Yu, and KyungTae Lim Conference on Empirical Methods in Natural Language Processing (EMNLP), to appear
2026	TReX: Tokenizer Regression for Optimal Data Mixture Inho Won, HanGyeol Yoo, Minkyung Cho , Jungyeul Park, Hoyun Song, and KyungTae Lim Conference of the European Chapter of the Association for Computational Linguistics (EACL)

Preprints & Technical Reports

2025	KORMo: Korean Open Reasoning Model for Everyone Minjun Kim, HyeonSeok Lim, HanGyeol Yoo, Inho Won, SeungWoo Song, Minkyung Cho , Junghun Yuk, ChangSu Choi, Dongjae Shin, Huije Lee, Hoyun Song, Alice Oh, and KyungTae Lim arXiv preprint. The first fully open LLM from the Korean open-source community
2024	Linq-Embed-Mistral Technical Report Chanyeol Choi, Junseong Kim, Seolhwa Lee, Jihoon Kwon, Sangmo Gu, Yejin Kim, Minkyung Cho , and Jy-yong Sohn arXiv preprint. Ranked #1 on the MTEB retrieval leaderboard

Experience

Jan - Feb 2024	Research Intern , LinqAlpha Synthetic training data for a Korean and law specialized embedding model
Jul - Aug 2023	Research Intern , MARG, Seoul National University Theme extraction from symbolic music

Jul - Aug 2022

Technical Writer, FriendliAI

Documentation for Orca, a distributed serving system for generative models

Scholarships

2022 - 2024

Forest of Learning Scholarship, Korea Foundation of Advanced Studies

Granted 16M KRW ($\simeq$ 11.8K USD) over two years. Our cohort still meets up monthly in EAST, an interdisciplinary East Asian studies research group we started together.

Languages

Korean (native), English, Mandarin Chinese