

Jaehyung Seo

KOREA UNIVERSITY · NLP & AI LAB

145 Anam-ro, Seongbuk-gu, Seoul, 02841, Republic of Korea

☎ (+82) 10-7375-7552 | ✉ seojae777@korea.ac.kr | 🏠 j-seo.github.io/ | 📺 J-Seo | 🌐 [jaehyungseo-datascientist](https://jaehyungseo-datascientist.github.io/)

Research Interests

I received my **Ph.D in the Department of Computer Science and Engineering at Korea University** (Advisor: [Heuiseok Lim](#)) in August 2025. I am a co-founder of KU-NMT Group. I have published more than **30 publications** in the field of **Natural Language Processing, Computer Science, and Artificial Intelligence**, along with 3 granted patents.

- Natural Language Processing
- Artificial Intelligence
- Commonsense Reasoning
- Hallucination in NLP
- Knowledge Editing
- Natural Language Generation
- Language Resource and Evaluation
- Language Modeling
- Neural Machine Translation
- Data Management and Engineering
- Semantic Search and Information Retrieval

Education

Korea University

PH.D. IN COMPUTER SCIENCE AND ENGINEERING, (GPA: 4.47 / 4.5)

- Advisor: Heuiseok Lim
- Natural Language Processing & Artificial Intelligence Lab

Seoul, South Korea

Sep. 2020 - Aug. 2025

Korea University

B.A. IN ENGLISH LANGUAGE AND LITERATURE & BUSINESS SCHOOL, (GPA: 4.19 / 4.5)

- Advisor: Donguk Kim

Seoul, South Korea

Mar. 2014 - Exp. Aug. 2020

Academic Services

Reviewer

NAACL / ICLR / EMNLP / AAAI

- Knowledge Editing & Hallucination & Bayesian & Language Generation & Large Language Model & Agent Systems

OpenReview

Jan - Aug. 2025

Reviewer

ACL / EMNLP

- Language Resource and Evaluation & Language Generation & Large Language Model

OpenReview

Jan - Oct. 2024

Reviewer

ACL / EMNLP

- Language Resource and Evaluation & Multilingual & Language Generation

OpenReview

Jan - Dec. 2023

Selected Publications

† (Corresponding author)

The Impact of Negated Text on Hallucination with Large Language Models

Jaehyung Seo, Hyeonseok Moon, Heuiseok Lim†

EMNLP 2025 - Proceedings of the Conference on Empirical Methods in Natural Language Processing

KoLEG: On-the-Fly Korean Legal Knowledge Editing with Continuous Retrieval

Jaehyung Seo, Dahyun Jung, Yongchan Chun Jaewook Lee, Dongjun Kim, Hwijung Ryu, Donghoon Shin, Heuiseok Lim†

EMNLP 2025 - Findings of the Association for Computational Linguistics

MultiDocFusion: Hierarchical and Multimodal Chunking Pipeline for Enhanced RAG on Long Industrial Documents

Joong Min Shin, Chanjun Park, Jeongbae Park, **Jaehyung Seo**[†], Heuseok Lim[†]

EMNLP 2025 - Proceedings of the Conference on Empirical Methods in Natural Language Processing

Metric Calculating Benchmark: Complicate Instruction Following Benchmark for Large Language Models

Hyeonseok Moon, Seongtae Hong, **Jaehyung Seo**[†], Heuseok Lim[†]

EMNLP 2025 - Proceedings of the Conference on Empirical Methods in Natural Language Processing

LimaCost: Data Valuation for Instruction Tuning of Large Language Models

Hyeonseok Moon, **Jaehyung Seo**, Seonmin Koo, Jinsung Kim, Youngkyoung Ham, Jiwon Moon, Heuseok Lim[†]

EMNLP 2025 - Findings of the Association for Computational Linguistics

Call for Rigor in Reporting Quality of Instruction Tuning Data

Hyeonseok Moon, **Jaehyung Seo**, Heuseok Lim[†]

ACL 2025 - Annual Meeting of the Association for Computational Linguistics

K-HALU: Multiple Answer Korean Hallucination Benchmark for Large Language Models

Jaehyung Seo, Heuseok Lim[†]

ICLR 2025 - The Thirteenth International Conference on Learning Representations

CoME: An Unlearning-based Approach to Conflict-free Model Editing

Dahyun Jung, **Jaehyung Seo**, Jaewook Lee, Chanjun Park[†], Heuseok Lim[†]

NAACL 2025 - Proceedings of the North American Chapter of the Association for Computational Linguistics

Find the Intention of Instruction: Comprehensive Evaluation of Instruction Understanding for Large Language Models

Hyeonseok Moon, **Jaehyung Seo**, Seungyoon Lee, Chanjun Park, Heuseok Lim[†]

NAACL 2025 - Findings of the North American Chapter of the Association for Computational Linguistics

An analysis on language transfer of pre-trained language model with cross-lingual post-training

SuhYune Son, Chanjun Park, Jungseob Lee, Midan Shim, Chanhee Lee, Yoonna Jang, **Jaehyung Seo**, Jungwoo Lim, Heuseok Lim[†]

ESWA 2025 - Experts Systems with Application

Intelligent Predictive Maintenance RAG framework for Power Plants: Enhancing QA with StyleDFS and Domain Specific Instruction Tuning

Seongtae Hong, Joong Min Shin, **Jaehyung Seo**, Taemin Lee, Jeongbae Park[†], Cho Man Young, Byeongho Choi, Heuseok Lim[†]

EMNLP 2024 - Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track

Post-negation Text Induce New Hallucinations in Large Language Models

Jaehyung Seo, Aram So, Heuseok Lim[†]

[Best Paper] HCLT 2024 - Annual Conference on Human and Language Technology

KoCommonGEN v2: A Benchmark for Navigating Korean Commonsense Reasoning Challenges in Large Language Models

Jaehyung Seo, Jaewook Lee, Chanjun Park, SeongTae Hong, Seungjun Lee, Heuseok Lim[†]

ACL 2024 - Findings of the Annual Meeting of the Association for Computational Linguistics

Length-aware Byte Pair Encoding for Mitigating Over-segmentation in Korean Machine Translation

Jungseob Lee, Hyeonseok Moon, Seungjun Lee Chanjun Park[†], Sugyeong Eo, Hyunwoong Ko, **Jaehyung Seo**, Seungyoon Lee, Heuseok Lim[†]

ACL 2024 - Findings of the Annual Meeting of the Association for Computational Linguistics

Leveraging Pre-existing Resources for Data-Efficient Counter-Narrative Generation in Korean

Seungyoon Lee, Chanjun Park[†], DaHyun Jung, Hyeonseok Moon, **Jaehyung Seo**, Sugyeong Eo, Heuseok Lim[†]

LRG-COLING 2024 - The Joint International Conference on Computational Linguistics, Language Resources and Evaluation

Hyper-BTS Dataset: Scalability and Enhanced Analysis of Back Transcription (BTS) for ASR Post-Processing

Chanjun Park, **Jaehyung Seo**, Seolhwa Lee, Junyoung Son, Hyeonseok Moon, Sugyeong Eo, Chanhee Lee, Heuseok Lim[†]

EACL 2024 - Findings of the Association for Computational Linguistics

KNOTICED: Augmentative and Alternative Communication Software for Language Developmental Disabilities

Sugyeong Eo, Jungwoo Lim, Chanjun Park, Hyeonseok Moon, **Jaehyung Seo**, Heuseok Lim[†]

LRG-COLING 2024 - The Joint International Conference on Computational Linguistics, Language Resources and Evaluation

Generative Interpretation: Toward Human-Like Evaluation for Educational Question-Answer Pair Generation

Hyeonseok Moon, Jaewook Lee, Sugyeong Eo, Chanjun Park, **Jaehyung Seo**, Heuseok Lim[†]

EACL 2024 - Findings of the Association for Computational Linguistics

CHEF in the Language Kitchen: A Generative Data Augmentation Leveraging Korean Morpheme Ingredients

Jaehyung Seo, Hyeonseok Moon, Jaewook Lee, Sugyeong Eo, Chanjun Park, Heuseok Lim[†]

EMNLP 2023 - Proceedings of the Conference on Empirical Methods in Natural Language Processing

Doubts on the reliability of parallel corpus filtering

Hyeonseok Moon, Chanjun Park, Seonmin Koo, Jungseob Lee, Seungjun Lee, **Jaehyung Seo**, Sugyeong Eo, Yoonna Jang, Hyunjoong Kim, Hyoung-gyu Lee, Heuseok Lim[†]

Expert Systems with Applications 233 (2023) p. 120962. Elsevier

KEBAP: Korean Error Explainable Benchmark Dataset for ASR and Post-processing

Seonmin Koo, Chanjun Park, Jinsung Kim, **Jaehyung Seo**, Sugyeong Eo, Hyeonseok Moon, Heuseok Lim[†]

EMNLP 2023 - Proceedings of the Conference on Empirical Methods in Natural Language Processing

CReTIHC: Designing Causal Reasoning Tasks about Temporal Interventions and Hallucinated Confoundings

Changwoo Chun, SongEun Lee, **Jaehyung Seo**, Heuseok Lim[†]

EMNLP 2023 - Findings of the Association for Computational Linguistics

Informative Evidence-guided Prompt-based Fine-tuning for English-Korean Critical Error Detection

Dahyun Jung, Sugyeong Eo, Chanjun Park, Hyeonseok Moon, **Jaehyung Seo**, Heuseok Lim[†]

JCNLP-AACL 2023 - Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics

Uncovering the Risks and Drawbacks Associated with the Use of Synthetic Data for Grammatical Error Correction

Seonmin Koo, Chanjun Park, Seolhwa Lee, **Jaehyung Seo**, Sugyeong Eo, Hyeonseok Moon, Heuseok Lim[†]

IEEE Access (2023). IEEE

PEEP-Talk: A Situational Dialogue-based Chatbot for English Education

Seungjun Lee, Yoonna Jang, Chanjun Park, Jungseob Lee, **Jaehyung Seo**, Hyeonseok Moon, Sugyeong Eo, Seounghoon Lee, Bernardo Yahya, Heuseok Lim[†]

ACL 2023 - Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)

Synthetic Alone: Exploring the Dark Side of Synthetic Data for Grammatical Error Correction

Chanjun Park, Seonmin Koo, Seolhwa Lee, **Jaehyung Seo**, Sugyeong Eo, Hyeonseok Moon, Heuseok Lim[†]

ICML-DMLR workshop (2023)

Knowledge Graph-Augmented Korean Generative Commonsense Reasoning

Dahyun Jung, **Jaehyung Seo**, Jaewook Lee, Chanjun Park, Heuseok Lim[†]

ICML-DMLR workshop (2023)

A Survey on Evaluation Metrics for Machine Translation

Seungjun Lee, Jungseob Lee, Hyeonseok Moon, Chanjun Park, **Jaehyung Seo**, Sugyeong Eo, Seonmin Koo, Heuseok Lim[†]

Mathematics 11.4 (2023) p. 1006. MDPI

PU-GEN: Enhancing generative commonsense reasoning for language models with human-centered knowledge

Jaehyung Seo, Dongsuk Oh, Sugyeong Eo, Chanjun Park, Kisu Yang, Hyeonseok Moon, Kinam Park, Heuseok Lim[†]

Knowledge-Based Systems (2022) p. 109861. Elsevier

Plain Template Insertion: Korean-Prompt-Based Engineering for Few-Shot Learners

Jaehyung Seo, Hyeonseok Moon, Chanhee Lee, Sugyeong Eo, Chanjun Park, Jihoon Kim, Changwoo Chun, Heuseok Lim[†]

IEEE Access 10 (2022) pp. 107587–107597

A Dog Is Passing Over The Jet? A Text-Generation Dataset for Korean Commonsense Reasoning and Evaluation

Jaehyung Seo, Seounghoon Lee, Chanjun Park, Yoonna Jang, Hyeonseok Moon, Sugyeong Eo, Seonmin Koo, Heuseok Lim[†]

NAACL 2022 - Findings of the Association for Computational Linguistics, Seattle, United States

BTS: Back TranScripton for Speech-to-Text Post-Processor using Text-to-Speech-to-Text

Chanjun Park, **Jaehyung Seo**, Seolhwa Lee, Chanhee Lee, Hyeonseok Moon, Sugyeong Eo, Heuseok Lim[†]

ACL 2021 - Proceedings of the 8th Workshop on Asian Translation (WAT2021), Online

Comparative Analysis of Current Approaches to Quality Estimation for Neural Machine Translation

Sugyeong Eo, Chanjun Park, Hyeonseok Moon, **Jaehyung Seo**, Heuseok Lim[†]

Applied Sciences 11.14 (2021) p. 6584. Multidisciplinary Digital Publishing Institute

QUAK: A Synthetic Quality Estimation Dataset for Korean-English Neural Machine Translation

Sugyeong Eo, Chanjun Park, Hyeonseok Moon, **Jaehyung Seo**, Gyeongmin Kim, Jungseob Lee, Heuseok Lim[†]

COLING 2022 - Proceedings of the 29th International Conference on Computational Linguistics

An Empirical Study on Automatic Post Editing for Neural Machine Translation

Hyeonseok Moon, Chanjun Park, Sugyeong Eo, **Jaehyung Seo**, Heuseok Lim[†]

IEEE Access 9 (2021) pp. 123754–123763. IEEE

Dealing with the Paradox of Quality Estimation

Sugyeong Eo, Chanjun Park, Hyeonseok Moon, **Jaehyung Seo**, Heuseok Lim[†]

MT Summit - Proceedings of the 4th Workshop on Technologies for MT of Low Resource Languages (LoResMT2021), Virtual

Grounded Vocabulary for Image Retrieval Using a Modified Multi-Generator Generative Adversarial Network

Kuekyeng Kim, Chanjun Park, **Jaehyung Seo**, Heuseok Lim[†]

IEEE Access (2021). IEEE

Automatic Knowledge Augmentation for Generative Commonsense Reasoning

Jaehyung Seo, Chanjun Park, Sugyeong Eo, Hyeonseok Moon, Heuseok Lim[†]

NeurIPS 2021 - Data-centric AI (DCAI) workshop

A New Tool for Efficiently Generating Quality Estimation Datasets

Sugyeong Eo, Chanjun Park, **Jaehyung Seo**, Hyeonseok Moon, Heuseok Lim[†]

NeurIPS 2021 - Data-centric AI (DCAI) workshop

How should human translation coexist with NMT? Efficient tool for building high quality parallel corpus

Chanjun Park, Seolhwa Lee, Hyeonseok Moon, Sugyeong Eo, **Jaehyung Seo**, Heuseok Lim[†]

NeurIPS 2021 - Data-centric AI (DCAI) workshop (2021)

An Automatic Post Editing with Efficient and Simple Data Generation Method

Hyeonseok Moon, Chanjun Park, **Jaehyung Seo**, Sugyeong Eo, Heuseok Lim[†]

IEEE Access (2022). IEEE

Dense-to-Question and Sparse-to-Answer: Hybrid Retriever System for Industrial Frequently Asked Questions

Jaehyung Seo, Taemin Lee, Hyeonseok Moon, Chanjun Park, Sugyeong Eo, Imatitkua D Aiyanyo, Kinam Park, Aram So, Sungmin Ahn, Jeongbae Park[†]

Mathematics 10.8 (2022) p. 1335. MDPI

Word-level Quality Estimation for Korean-English Neural Machine Translation

Sugyeong Eo, Chanjun Park, Hyeonseok Moon, **Jaehyung Seo**, Heuseok Lim[†]

Return on Advertising Spend Prediction with Task Decomposition-Based LSTM Model

Hyeonseok Moon, Taemin Lee, **Jaehyung Seo**, Chanjun Park, Sugyeong Eo, Imatitkua D Aiyanyo, Jeongbae Park, Aram So, Kyoungwha Ok, Kinam Park[†]

Mathematics 10.10 (2022) p. 1637. MDPI

Empirical Analysis of Parallel Corpora and In-Depth Analysis Using LIWC

Chanjun Park, Midan Shim, Sugyeong Eo, Seolhwa Lee, **Jaehyung Seo**, Hyeonseok Moon, Heuseok Lim[†]

Applied Sciences 12.11 (2022) p. 5545. Multidisciplinary Digital Publishing Institute

Priming Ancient Korean Neural Machine Translation

Chanjun Park, Seolhwa Lee, **Jaehyung Seo**, Hyeonseok Moon, Sugyeong Eo, Heuseok Lim[†]

LREC 2022 - Proceedings of the Language Resources and Evaluation Conference, Marseille, France

Empirical Analysis of Noising Scheme based Synthetic Data Generation for Automatic Post-editing

Hyeonseok Moon, Chanjun Park, Seolhwa Lee, **Jaehyung Seo**, Jungseob Lee, Sugyeong Eo, Heuseok Lim[†]

LREC 2022 - Proceedings of the Language Resources and Evaluation Conference, Marseille, France

A Self-Supervised Automatic Post-Editing Data Generation Tool

Hyeonseok Moon, Chanjun Park, Sugyeong Eo, **Jaehyung Seo**, SeungJun Lee, Heuseok Lim[†]

ICML 2022 - DataPerf workshop (2022)

BERTOEIC: Solving TOEIC Problems Using Simple and Efficient Data Augmentation Techniques with Pretrained Transformer Encoders

Jeongwoo Lee, Hyeonseok Moon, Chanjun Park, **Jaehyung Seo**, Sugyeong Eo, Heuseok Lim[†]

Applied Sciences 12.13 (2022) p. 6686. Multidisciplinary Digital Publishing Institute

Research on Utilization Strategy of User Engagements in Korean Fake News Detection

Myunghoon Kang, **Jaehyung Seo**, Chanjun Park, Heuseok Lim[†]

IEEE Access (2022) pp. 1–1

PicTalky: Augmentative and Alternative Communication Software for Language Developmental Disabilities

Chanjun Park, Yoonna Jang, Seolhwa Lee, **Jaehyung Seo**, Kisu Yang, Heuseok Lim[†]

IJCNLP-AACL 2022 - The 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics

K-NCT: Korean Neural Grammatical Error Correction Gold-Standard Test Set Using Novel Error Type Classification Criteria

Seonmin Koo, Chanjun Park, **Jaehyung Seo**, Seungjun Lee, Hyeonseok Moon, Jungseob Lee, Heuseok Lim[†]

IEEE Access (2022). IEEE

The ASR Post-Processor Performance Challenges of BackTranScription (BTS): Data-Centric and Model-Centric Approaches

Chanjun Park, **Jaehyung Seo**, Seolhwa Lee, Chanhee Lee, Heuseok Lim[†]

Mathematics 10.19 (2022)

Study on Decoding Strategies in Neural Machine Translation

Jaehyung Seo, Chanjun Park, Sugyeong Eo, Hyeonseok Moon, Heuseok Lim[†]

Journal of the Korea Convergence Society 12.11 (2021) pp. 69–80. Korea Convergence Society

Honors & Awards

2025.08	Selected as a Member of the National Representative AI Research Team (with NC AI), NEWS	South Korea
2024.10	BEST Paper Award, 2024 , The 36th Annual Conference on Human & Cognitive Language Technology	South Korea
2023.10	Outstanding Paper Award, 2023 , The 35th Annual Conference on Human & Cognitive Language Technology	South Korea
2022.12	NAVER PH.D Fellowship Award, 2022 , Scholarship Award	South Korea
2022.12	1st place in WMT 2022 QE Task 3, 2022 , WMT 2022 - EMNLP 2022	Abu Dhabi
2022.10	Outstanding Paper Award, 2022 , The 34th Annual Conference on Human & Cognitive Language Technology	South Korea
2021.10	Outstanding Paper Award, 2021 , The 33rd Annual Conference on Human & Cognitive Language Technology	South Korea
2021.07	4th Place in CommonGen 1.1 Leaderboard, 2021 , CommonGen 1.1 Leaderboard	World Record

Selected Research Experience

(National Initiative) National Representative AI — World Best LLM

NC AI, Korea University, ETRI, KAIST,
SNU, LOTTE Innovate, POSCO DX

RESEARCH ENGINEER

Aug 2025 – Dec 2025

- Participated in Korea's National Representative AI initiative, supported by NIPA, IITP, and NIA.
- Built a foundation model from scratch using 1,000 NVIDIA H100 GPUs.
- Contributed to large-scale LLM post-training, reasoning, Mixture-of-Experts (MoE) modeling, dataset construction, and reinforcement learning.

(Collaborative Research) Benchmark Dataset across Eight Asian Countries

Singapore A* Research & Korea
University

PROJECT MANAGER

May 2025 – Oct 2025

- Led the design and evaluation of a multi-modal East-Asian benchmark for LLMs.
- Coordinated collaboration spanning eight countries: Singapore, South Korea, China, Japan, Mongolia, Vietnam, Indonesia, and India.

(Industry–Academic Joint Project) Unlearning for LLM Reliability

KT & Korea University

PROJECT MANAGER & RESEARCH ENGINEER

Apr 2024 – May 2025

- Developed knowledge editing methods tailored for vertical LLMs.
- Proposed RAG-based editing approaches with lifelong learning capabilities.
- KoLEG: On-the-Fly Korean Legal Knowledge Editing with Continuous Retrieval

(National Initiative) Next-Generation AI Technology Development — Hallucination

Korea University, KAIST, SKKU, KU,
KETI, NCAI, O2O

PROJECT MANAGER & RESEARCH ENGINEER

Jan 2024 – Dec 2026

- Supported by IITP.
- Secured research funding of **4.8B KRW** (≈ \$3.5M) over 3 years.
- Led research on hallucination detection, mitigation, and domain adaptation.
- Received **Best Paper Award**: Post-negation Text Induces New Hallucinations in Large Language Models.
- Published at **ICLR 2025**: [OpenReview Link](#).

(Industry–Academic Joint Project) Large Language Modeling for Special Technical Documents

GAON & Korea University

PROJECT MANAGER & RESEARCH ENGINEER

Dec 2023 – Jun 2024

- Supported by GAON.
- Proposed a RAG-based LLM for predictive maintenance in nuclear power plants.
- Developed an on-premise deployable model system for handling classified government documents.
- Published at **EMNLP 2024 (Industry Track)**: [ACL Anthology Link](#).

(Collaborative Research) Open Ko-LLM Leaderboard

Upstage & Korea University

RESEARCH ENGINEER

Sep 2023 – Dec 2023

- Constructed and deployed the H5 Leaderboard to evaluate commonsense reasoning in LLMs.
- Contributed to the evaluation of 2,000+ open-source LLMs.
- Developed **KoCommonGEN v2**, a benchmark for Korean commonsense reasoning challenges.
- Published at **ACL Findings 2024**: [ACL Anthology Link](#).

(Industry–Academic Joint Project) UI/UX Artificial Intelligence Service Development

Samsung Fire and Marine Insurance
& Korea University

RESEARCH ENGINEER

Jun 2023 – Feb 2024

- Developed an AI-based application service for dementia prevention (**Synerpeace**).
- Implemented memory-enhancement learning via API calls and developed an automatic diary summarization feature.

(Industry–Academic Joint Project) R&D Voucher: AI Contact Center

O2O & Korea University

RESEARCH ENGINEER

Jun 2022 – May 2023

- Developed an AI-based system for evaluating call quality and providing job training for customer service agents.

(Industry–Academic Joint Project) R&D Voucher: AI Interviewing

Withmind & Korea University

RESEARCH ENGINEER

Jun 2022 – May 2023

- Developed an AI-powered interview automation service and established an evaluation system.

(Industry–Academic Joint Project) Data Augmentation with Language Generation Research

Samsung Research & Korea
University

PROJECT MANAGER & RESEARCH ENGINEER

May 2022 – Apr 2023

- Proposed a data augmentation method using language generation to improve training of a phone-scam classifier.

(Industry–Academic Joint Project) Korean Prompt-based Engineering for Few-shot Learning

Hyundai Motor Group & Korea
University

PROJECT MANAGER & RESEARCH ENGINEER

Mar 2021 – Apr 2022

- Proposed the Plain Template Insertion method for Korean prompt-based few-shot learning.
- Developed an instance classifier for low-resource scenarios.
- Published at **IEEE Access (SCIE)**: [IEEE Xplore Link](#).

(Industry–Academic Joint Project) Return on Advertising Spend (ROAS) Prediction Research

BizSpring & Korea University

RESEARCH ENGINEER

Dec 2021 – Mar 2022

- Conducted research on predicting Return on Advertising Spend (ROAS) to optimize advertising effectiveness.
- Project supported by the Data Voucher Program.
- Published at **Mathematics (SCIE)**: [MDPI Link](#).

(Project) Hybrid Retriever System for Industrial Frequently Asked Questions Research

O2O & Korea University

PROJECT MANAGER & RESEARCH ENGINEER

Dec 2021 – Mar 2022

- Proposed a hybrid information retrieval method for customer inquiries by combining dense and sparse embeddings.
- Introduced scaling and weighting functions to improve robustness to domain-specific terminology.
- Published at **Mathematics (SCIE)**: [MDPI Link](#).

(Industry–Academic Joint Project) Korean Multi-turn Response Chatbot Research

Samsung SDS & Korea University

RESEARCH ENGINEER

May 2021 – Jan 2022

- Developed a Korean multi-turn chatbot based on BlenderBot.
- Published at **KCI 2022**: [DOI Link](#).

(Industry–Academic Joint Project) Improving the Understanding of Korean-based Generative Models

NSR & Korea University

RESEARCH ENGINEER

Jul 2020 – Mar 2021

- Project supported by the National Security Research Institute (NSR).
- Developed a GPT-based model to improve classification performance.

Skills

Programming	Python
Machine Learning & Deep Learning	Pytorch
Natural Language Processing	Commonsense Reasoning, Knowledge Editing, Hallucination, Large Language Model
GPU Server Build & Management	Linux system engineering, Kubernetes, NAS, MLOps, GPU parallelization, RAID configuration
Language Skills	Korean, English

Patents

2025.06	APPARATUS OF OBTAINING RESPONSE BASED ON SEMANTIC SIMILARITY AND METHOD THEREOF , Granted Patent (10-2828691)	South Korea
2023.12	DEVICE AND METHOD FOR KOREAN DATA AUGMENTATION BASED ON CONTRASTIVE LEARNING AND MORPHEME COMBINATION , Patent Application (10-2023-0189496)	South Korea
2023.10	APPARATUS AND METHOD FOR GENERATING LANGUAGE BASED ON COMMONSENSE REASONING , Granted Patent (10-2593463)	South Korea
2023.03	GENERATING METHOD OF MODEL MEASURING DOCUMENT SIMILARITY AND MEASURING METHOD FOR DOCUMENT SIMILARITY USING THE SAME , Granted Patent (10-2507192)	South Korea
2022.10	DEVICE AND METHOD FOR GENERATING KOREAN COMMONSENSE REASONING DATASET , Patent Application (10-2022-0132373)	South Korea

Invited Talk & Teaching Experience

DGIST

INVITED SPEAKER

South Korea

Apr. 2025

- Commonsense Reasoning or Hallucination? Unpacking the Plausible Errors of Language Models

Dongguk University

INVITED ADVISOR

South Korea

Jun. 2023

- Introduction to Natural Language Processing and Commonsense Reasoning

NongHyup Bank Co., Ltd

INVITED ADVISOR

South Korea

Jul. 2020 - Sep. 2020

- Teaching data science for new employees

Korea University

TEACHING ASSISTANT

South Korea

Mar - Jun. 2023

- (DFE615-00) Natural Language Processing for Artificial Intelligence

Korea University

TEACHING ASSISTANT

- (DFE610) Natural Language Processing for Digital Finance Engineering
- Lecture materials: <https://github.com/nlpai-lab/nlp-bible-code>

South Korea

Sep - Dec. 2022

Korea University

TEACHING ASSISTANT

- (BDC101) Introduction to Natural Language Processing in Big Data
- Lecture materials: https://github.com/J-Seo/BDC103_IR

South Korea

Mar - Jun. 2022

Korea University

RESEARCH ASSISTANT

- (DFE610) Natural Language Processing for Digital Finance Engineering
- Lecture materials: <https://github.com/nlpai-lab/nlp-bible-code>

South Korea

Sep - Dec. 2021

Korea University

TEACHING ASSISTANT

- (BDC101) Introduction to Natural Language Processing in Big Data
- Lecture materials: https://github.com/J-Seo/BDC103_IR

South Korea

Mar - Jun. 2021