

Jaehyung Seo

KONKUK UNIVERSITY · ELION LAB

120 Neungdong-ro, Gwangjin-gu, Seoul 05029, South Korea

✉ seojae777@konkuk.ac.kr | 🌐 elionlab.github.io | 🗣️ J-Seo | 🌐 [jaehyungseo-elionlab](https://jaehyungseo-elionlab.github.io)

Research Interests

I am currently an **Assistant Professor** in the **Department of Computer Science and Engineering at Konkuk University**, leading the **ELION Lab**. I received my **Ph.D. in Computer Science and Engineering from Korea University** (Advisor: [Heuiseok Lim](#)) in August 2025. I am a co-founder of the **KU-NMT Group**. My research has resulted in **30+ publications** in **Natural Language Processing, Computer Science**, and **Artificial Intelligence**, as well as **four granted patents**.

- Natural Language Processing
- Language Modeling
- Commonsense Reasoning
- Hallucination Detection and Mitigation
- Model Editing
- Language Resource and Evaluation
- Multilingual & Multimodal Systems
- AI Agent
- Information Retrieval & RAG

Professional Experience

Konkuk University, Department of Computer Science and Engineering

Seoul, South Korea

ASSISTANT PROFESSOR

Mar. 2026 – Present

- Leading the **ELION Lab (Elucidating Language Intelligence & Representation)**
- Research areas: NLP, Language Models, Interpretation, Reasoning, Hallucination, Model Editing, and World Models

Education

Korea University

Seoul, South Korea

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

Sep. 2020 - Aug. 2025

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

Korea University

Seoul, South Korea

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

Mar. 2014 - Exp. Aug. 2020

- Advisor: Donguk Kim

Academic Services

REVIEWER

- **AI / Machine Learning**: ICLR (2025, 2026), AAAI (2026)
- **Natural Language Processing**: ACL (2023–2025), EMNLP (2023–2025), NAACL (2025), COLING (2022)
- **Computer Vision**: CVPR (2026)

Selected Publications

[†] (Corresponding author)

PILAR: A Page-Grounded Unified Evidence Representation via an Entity-Linked Assertion Graph for Open-Domain QA Agents over Multimodal Document Corpora

Joongmin Shin, Gyuho Shim, Junghun Lee[†], [Jaehyung Seo](#)[†]

EMNLP 2026 - Findings of the Association for Computational Linguistics

The Geometry of Hallucination Detection: Why Simple Probes Are Sufficient

Jungseob Lee, [Jaehyung Seo](#)[†], Heuiseok Lim[†]

EMNLP 2026 - Proceedings of the Conference on Empirical Methods in Natural Language Processing

CultureConverse: A Multilingual Multi-turn Simulation Harness for Culturally Grounded Assistance in East and Southeast Asia

Bryan Chen Zhengyu Tan, Weihua Zheng, Thong T. Doan, Bich Ngoc Doan, Jia Wang Peh, Xiaoyuan Yi, Jing Yao, Xing Xie, Nancy F. Chen, Zhengyuan Liu, JinYeong Bak, Wafi Shamdi, Soo Kai Chie, Liew Yu Siong, Aina Azyyati Binti Mohamad Rezal, Lew Yan Yan Vanessa, Huadan Wu, Dylan Raharja, Nadya Yuki Wangsajaya, Akane Fukushima, Kazushi Kato, Koji Inoue, Tatsuya Kawahara, **Jaehyung Seo**, Dongjun Kim, Seungyoon Lee, Zi Haur Pang, Rui Yang Tan, Charibeth Ko Cheng, Maria Regina Justina Estuar, Jann Railey Montalan, Pham Minh Duc, Roy Ka-Wei Lee

EMNLP 2026 - Proceedings of the Conference on Empirical Methods in Natural Language Processing

SERA: Self-referential Assessment Framework for Bidirectional Generative Commonsense Reasoning

Jaehyung Seo, Hyeonseok Moon, Yoonna Jang, Heuseok Lim[†]

Knowledge-Based Systems 345 (2026) p. 116152. Elsevier

HiKEY: Hierarchical Multimodal Retrieval for Open-Domain Document Question Answering

Joongmin Shin, Gyuho Shim, Jeongbae Park, **Jaehyung Seo**[†], Heuseok Lim[†]

ACL 2026 - Annual Meeting of the Association for Computational Linguistics (Oral)

No Reader Left Behind: Multi-Agent Summaries Everyone Can Understand

Jimin Jung, MyoungJin Kim, **Jaehyung Seo**[†], Heuseok Lim[†]

ACL 2026 - Annual Meeting of the Association for Computational Linguistics (Oral, Best Paper Award Nomination)

MMAC: A Multilingual, Multimodal Alignment Framework for Cultural Grounding Evaluation

Weihua Zheng, Zhengyuan Liu, Tanmoy Chakraborty, Weiwen Xu, Xiaoxue Gao, Bryan Chen Zhengyu Tan, Bowei Zou, Chang Liu, Yujia Hu, Xing Xie, Xiaoyuan Yi, Jing Yao, Chaojun Wang, Long Li, Rui Liu, Huiyao Liu, Koji Inoue, Ryuichi Sumida, Tatsuya Kawahara, Fan Xu, Lingyu Ye, Wei Tian, Dongjun Kim, Jimin Jung, **Jaehyung Seo**, Nadya Yuki Wangsajaya, Pham Minh Duc, Ojasva Saxena, Palash Nandi, Xiyao Tao, Wiwik Karlina, Tuan Luong, Keertana Arun Vasani, Roy Ka-Wei Lee, Nancy F. Chen

ACL 2026 - Annual Meeting of the Association for Computational Linguistics

Evidential Transformation Network: Turning Pretrained Models into Evidential Models for Uncertainty Estimation

Yongchan Chun, Chanhee Park, Jeongho Yoon, **Jaehyung Seo**[†], Heuseok Lim[†]

CVPR 2026 - IEEE/CVF Conference on Computer Vision and Pattern Recognition (Highlight)

M3DocDep: Multi-modal, Multi-page, Multi-document Dependency Chunking with Large Vision-Language Models

Joongmin Shin, Jeongbae Park, **Jaehyung Seo**[†], Heuseok Lim[†]

CVPR 2026 - IEEE/CVF Conference on Computer Vision and Pattern Recognition

The Impact of Negated Text on Hallucination with Large Language Models

Jaehyung Seo, Hyeonseok Moon, Heuseok 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, Heuseok 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: Code-Verifiable 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

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

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

Jaehyung Seo, Heuseok Lim[†]

ICLR 2025 - The Thirteenth International Conference on Learning Representations

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

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

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

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

Automatic Knowledge Augmentation for Generative Commonsense Reasoning

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

NeurIPS 2021 - Data-centric AI (DCAI) workshop

Study on Decoding Strategies in Neural Machine Translation

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

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

All Publications

[†] (Corresponding author)

PILAR: A Page-Grounded Unified Evidence Representation via an Entity-Linked Assertion Graph for Open-Domain QA Agents over Multimodal Document Corpora

Joongmin Shin, Gyuho Shim, Junghun Lee[†], **Jaehyung Seo**[†]

EMNLP 2026 - Findings of the Association for Computational Linguistics

The Geometry of Hallucination Detection: Why Simple Probes Are Sufficient

Jungseob Lee, **Jaehyung Seo**[†], Heuiseok Lim[†]

EMNLP 2026 - Proceedings of the Conference on Empirical Methods in Natural Language Processing

CultureConverse: A Multilingual Multi-turn Simulation Harness for Culturally Grounded Assistance in East and Southeast Asia

Bryan Chen Zhengyu Tan, Weihua Zheng, Thong T. Doan, Bich Ngoc Doan, Jia Wang Peh, Xiaoyuan Yi, Jing Yao, Xing Xie, Nancy F. Chen, Zhengyuan Liu, JinYeong Bak, Wafi Shamdi, Soo Kai Chie, Liew Yu Siong, Aina Azyyati Binti Mohamad Rezal, Lew Yan Yan Vanessa, Huadan Wu, Dylan Raharja, Nadya Yuki Wangsajaya, Akane Fukushima, Kazushi Kato, Koji Inoue, Tatsuya Kawahara, **Jaehyung Seo**, Dongjun Kim, Seungyoon Lee, Zi Haur Pang, Rui Yang Tan, Charibeth Ko Cheng, Maria Regina Justina Estuar, Jann Railey Montalan, Pham Minh Duc, Roy Ka-Wei Lee

EMNLP 2026 - Proceedings of the Conference on Empirical Methods in Natural Language Processing

SERA: Self-referential Assessment Framework for Bidirectional Generative Commonsense Reasoning

Jaehyung Seo, Hyeonseok Moon, Yoonna Jang, Heuiseok Lim[†]

Knowledge-Based Systems 345 (2026) p. 116152. Elsevier

DART: Draft-Agreement Routing for Training-Free Adaptive Thinking Budgets in Hybrid Reasoning Models

Jungseob Lee, Seongtae Hong, Seungjun Lee, **Jaehyung Seo**, Junyoung Son, Sugyeong Eo, Chanjun Park, Hyeongju Park, Hyeonseok Moon, Heuiseok Lim

EMNLP 2026 - Findings of the Association for Computational Linguistics

ThinkFuse: Trajectory-Aware Test-Time Fusion for Small Reasoning Models

Myunghoon Kang, Jungseob Lee, **Jaehyung Seo**, Heuiseok Lim

EMNLP 2026 - Findings of the Association for Computational Linguistics

HiKEY: Hierarchical Multimodal Retrieval for Open-Domain Document Question Answering

Joongmin Shin, Gyuho Shim, Jeongbae Park, **Jaehyung Seo**[†], Heuiseok Lim[†]

ACL 2026 - Annual Meeting of the Association for Computational Linguistics (Oral)

No Reader Left Behind: Multi-Agent Summaries Everyone Can Understand

Jimin Jung, MyoungJin Kim, **Jaehyung Seo**[†], Heuiseok Lim[†]

ACL 2026 - Annual Meeting of the Association for Computational Linguistics (Oral, Best Paper Award Nomination)

MMAC: A Multilingual, Multimodal Alignment Framework for Cultural Grounding Evaluation

Weihua Zheng, Zhengyuan Liu, Tanmoy Chakraborty, Weiwen Xu, Xiaoxue Gao, Bryan Chen Zhengyu Tan, Bowei Zou, Chang Liu, Yujia Hu, Xing Xie, Xiaoyuan Yi, Jing Yao, Chaojun Wang, Long Li, Rui Liu, Huiyao Liu, Koji Inoue, Ryuichi Sumida, Tatsuya Kawahara, Fan Xu, Lingyu Ye, Wei Tian, Dongjun Kim, Jimin Jung, **Jaehyung Seo**, Nadya Yuki Wangsajaya, Pham Minh Duc, Ojasva Saxena, Palash Nandi, Xiyan Tao, Wiwik Karlina, Tuan Luong, Keertana Arun Vasan, Roy Ka-Wei Lee, Nancy F. Chen

ACL 2026 - Annual Meeting of the Association for Computational Linguistics

Evidential Transformation Network: Turning Pretrained Models into Evidential Models for Uncertainty Estimation

Yongchan Chun, Chanhee Park, Jeongho Yoon, **Jaehyung Seo**[†], Heuiseok Lim[†]

CVPR 2026 - IEEE/CVF Conference on Computer Vision and Pattern Recognition (Highlight)

M3DocDep: Multi-modal, Multi-page, Multi-document Dependency Chunking with Large Vision-Language Models

Joongmin Shin, Jeongbae Park, **Jaehyung Seo**[†], Heuiseok Lim[†]

CVPR 2026 - IEEE/CVF Conference on Computer Vision and Pattern Recognition

VAETKI Technical Report

NC-AI Consortium

LimaCost: Data Valuation for Instruction Tuning of Large Language Models

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

EMNLP 2025 - Findings of the Association for Computational Linguistics

Call for Rigor in Reporting Quality of Instruction Tuning Data

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

ACL 2025 - Annual Meeting 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, Heuiseok 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, Heuiseok 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, Heuiseok Lim[†]

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

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, Heuiseok 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, Heuiseok Lim[†]

LREC-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, Heuiseok 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**, Heuiseok Lim[†]

LREC-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**, Heuiseok Lim[†]

EACL 2024 - Findings of the Association for Computational Linguistics

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, Heuiseok 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, Heuiseok Lim[†]

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

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

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

IJCNLP-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, Heuiseok 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, Heuiseok Lim[†]

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

Knowledge Graph-Augmented Korean Generative Commonsense Reasoning

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

Mathematics 11.4 (2023) p. 1006. MDPI

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

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

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

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

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

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

Dealing with the Paradox of Quality Estimation

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

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

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

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

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

IEEE Access 10 (2022) pp. 44964–44973. IEEE

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

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[†]

JCNLP-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

Honors & Awards

- 2026.07 **Best Paper Award Nomination (Top 1%), 2026**, ACL 2026
- 2026.07 **Two Oral Presentations (Top 5%), 2026**, ACL 2026
- 2026.06 **Highlight Paper (Top 3.7%), 2026**, CVPR 2026
- 2025.10 **BEST Paper Award, 2025**, The 37th Annual Conference on Human & Cognitive Language Technology
- 2025.08 **Selected as a Member of the National Representative AI Research Team**, NC AI
- 2024.10 **BEST Paper Award, 2024**, The 36th Annual Conference on Human & Cognitive Language Technology
- 2023.10 **Outstanding Paper Award, 2023**, The 35th Annual Conference on Human & Cognitive Language Technology
- 2022.12 **NAVER PH.D Fellowship Award, 2022**, Scholarship Award
- 2022.12 **1st place in WMT 2022 QE Task 3, 2022**, WMT 2022 - EMNLP 2022
- 2022.10 **Outstanding Paper Award, 2022**, The 34th Annual Conference on Human & Cognitive Language Technology
- 2021.10 **Outstanding Paper Award, 2021**, The 33rd Annual Conference on Human & Cognitive Language Technology
- 2021.07 **4th Place in CommonGen 1.1 Leaderboard, 2021**, CommonGen 1.1 Leaderboard

Selected Research Experience

(National Initiative) AI Star Fellowship 2026 — Top-Tier Early-Career AI Researcher Support Program

Jeju National University (Lead),
Konkuk University, NC AI, AIVIS,
Metsakuur Company

PROJECT LEADER (PROJECT 1)

Jul 2026 – Dec 2031

- Supported by MSIT and IITP.
- Selected for the AI Star Fellowship 2026, a 6-year national program with total funding of **KRW 11.0B** (≈ \$8.0M), of which **KRW 5.0B** (≈ \$3.6M) is allocated to Konkuk University.
- Serving as **Project 1 Leader (PL)**, in collaboration with Jeju National University (lead institution), NC AI, AIVIS, and Metsakuur Company.

(Research Infrastructure) KETI AI Research Computing Support Project

*Korea Electronics Technology
Institute (KETI) & Konkuk University*

PRINCIPAL INVESTIGATOR

Jul 2026 – Present

- Selected as PI for the KETI AI Research Computing Support Project.
- Awarded dedicated GPU computing resources (**NVIDIA H100 × 8**).

(National R&D Project) Development of a Korean Language Knowledge-Linkage System for Personalized Comprehensive Korean Language Counseling Services

*Konkuk University, Korea University,
Yonsei University, Hankuk University
of Foreign Studies, Ontheit*

PARTICIPATING RESEARCHER

Apr 2026 – Dec 2028

- Supported by the Korea Creative Content Agency (KOCCA).
- Selected as a new project in Apr 2026, with total funding of **KRW 4.1B** (≈ \$3.0M) over 2026–2028.

(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 LLM post-training & dataset construction

(Collaborative Research) MMA-ASIA: A Multilingual and Multimodal Alignment Framework for Culturally-Grounded Evaluation

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.
- Collaboration Partners: A*, Kyoto University, Alibaba DAMO, Jiangxi Normal University, IIT Delhi, Inner Mongolia University

(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

RESEARCH ENGINEER

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

Withmind & Korea University

Jun 2022 – May 2023

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

PROJECT MANAGER & RESEARCH ENGINEER

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

Samsung Research & Korea University

May 2022 – Apr 2023

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

PROJECT MANAGER & RESEARCH ENGINEER

- 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](#).

Hyundai Motor Group & Korea University

Mar 2021 – Apr 2022

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

RESEARCH ENGINEER

- 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](#).

BizSpring & Korea University

Dec 2021 – Mar 2022

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

PROJECT MANAGER & RESEARCH ENGINEER

- 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](#).

O2O & Korea University

Dec 2021 – Mar 2022

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

RESEARCH ENGINEER

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

Samsung SDS & Korea University

May 2021 – Jan 2022

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

RESEARCH ENGINEER

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

NSR & Korea University

Jul 2020 – Mar 2021

Patents

2025.10	DEVICE AND METHOD FOR GENERATING KOREAN COMMONSENSE REASONING DATASET , Granted Patent (10-2875876)	South Korea
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

Konkuk University

INSTRUCTOR

- Algorithms (taught in English)
- Understanding Large Language Models
- Capstone Project 1

South Korea

Sep - Dec. 2026

Konkuk University

INSTRUCTOR

- Algorithms (taught in English)
- Object-Oriented Programming (taught in English)

South Korea

Mar - Jun. 2026

DGIST

INVITED SPEAKER

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

South Korea

Apr. 2025

Dongguk University

INVITED ADVISOR

- Introduction to Natural Language Processing and Commonsense Reasoning

South Korea

Jun. 2023

NongHyup Bank Co., Ltd

INVITED ADVISOR

- Teaching data science for new employees

South Korea

Jul. 2020 - Sep. 2020

Korea University

TEACHING ASSISTANT

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

South Korea

Mar - Jun. 2023

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