Hyeonseok Moon

☎ (+82) 10-5124-2557 • ✉ glee889@korea.ac.kr • 🌐 https://hyeonseokk.github.io/
hyeonseokk • in hyeonseok-moon-nlp • queGQ5UAAAAJ

Research Area

- ○ Language Resource
- ○ Model Evaluation
- ○ Benchmark
- ○ Data Evaluation
- ○ Large Language Model
- ○ Machine Translation
- ○ Language Generation
- ○ Data Engineering

Education

Korea University **2021.03 – 2026.02(expected)**
Ph.D Candidate
Major in Computer Science and Engineering - Advisor: Prof. Heuiseok Lim

Korea University **2015.03 – 2021.02**
Bachelor of Science and Engineering
Major in Mathematics and Artificial Intelligence (Double Majors) - Advisor: Prof. Euisung Park

Selected Publications

International Conference

Call for Rigor in Reporting Quality of Instruction Tuning Data **2025**
Hyeonseok Moon, Jaehyung Seo, Heuiseok Lim
ACL 2025

Cross-Lingual Optimization for Language Transfer in Large Language Models **2025**
Jungseob Lee, Seongtae Hong, Hyeonseok Moon, Heuiseok Lim
ACL 2025

Semantic Aware Linear Transfer by Recycling Pre-trained Language Models for Cross-lingual Transfer **2025**
Seungyoon Lee, Seongtae Hong, Hyeonseok Moon, Heuiseok Lim
ACL 2025 Findings

FLEX: A Benchmark for Evaluating Robustness of Fairness in Large Language Models **2025**
Dahyun Jung, Seungyoon Lee, Hyeonseok Moon, Chanjun Park, Heuiseok Lim
NAACL 2025 Findings

MIRAGE: A Metric-Intensive Benchmark for Retrieval-Augmented Generation Evaluation **2025**
Chanhee Park, Hyeonseok Moon, Chanjun Park, Heuiseok Lim
NAACL 2025 Findings

Find the Intention of Instruction: Comprehensive Evaluation of Instruction Understanding for Large Language Models **2025**
Hyeonseok Moon, Jaehyung Seo, Seungyoon Lee, Chanjun Park, Heuiseok Lim
NAACL 2025 Findings

MIGRATE: Cross-Lingual Adaptation of Domain-Specific LLMs through Code-Switching and Embedding Transfer **2025**
Seongtae Hong, Seungyoon Lee, Hyeonseok Moon, Heuiseok Lim
COLING 2025

Leveraging Pre-existing Resources for Data-Efficient Counter-Narrative Generation in Korean **2024**
Seungyoon Lee, Chanjun Park, DaHyun Jung, Hyeonseok Moon, Jaehyung Seo, Sugyeong Eo, Heuiseok Lim
LREC-COLING 2024

Detecting Critical Errors Considering Cross-Cultural Factors in English-Korean Translation **2024**
Sugyeong Eo, Jungwoo Lim, Chanjun Park, DaHyun Jung, Seonmin Koo, Hyeonseok Moon,
Jaehyung Seo, Heuiseok Lim
LREC-COLING 2024

Translation of Multifaceted Data without Re-Training of Machine Translation Systems **2024**
Hyeonseok Moon, Seungjun Lee, Seongtae Hong, Seungjum Lee, Chanjun Park, Heuiseok Lim
EMNLP 2024 Findings

Length-aware Byte Pair Encoding for Mitigating Over-segmentation in Korean Machine Translation **2024**
Jungseob Lee, Hyeonseok Moon(*equal contribution*), Seungjun Lee, Chanjun Park, Sugyeong Eo, Hyunwoong Ko, Jaehyung Seo, Seungyoon Lee, Heuiseok Lim
ACL 2024 Findings

Generative Interpretation: Toward Human-Like Evaluation for Educational Question-Answer Pair Generation **2024**
Hyeonseok Moon, Jaewook Lee, Sugyeong Eo, Chanjun Park, Jaehyung Seo, Heui-Seok Lim
EACL 2024 Findings

Hyper-BTS Dataset: Scalability and Enhanced Analysis of Back TranScription (BTS) for ASR Post-Processing **2024**
Chanjun Park, Jaehyung Seo, Seolhwa Lee, Junyoung Son, Hyeonseok Moon, Sugyeong Eo, Chanhee Lee, Heui-Seok Lim
EACL 2024 Findings

Leveraging Pre-existing Resources for Data-Efficient Counter-Narrative Generation in Korean **2024**
Seungyoon Lee, Chanjun Park, DaHyun Jung, Hyeonseok Moon, Jaehyung Seo, Sugyeong Eo, Heui-Seok Lim
LREC-COLING 2024

Detecting Critical Errors Considering Cross-Cultural Factors in English-Korean Translation **2024**
Sugyeong Eo, Jungwoo Lim, Chanjun Park, Dahyun Jung, Seonmin Koo, Hyeonseok Moon, Jaehyung Seo, Heui-Seok Lim
LREC-COLING 2024

CHEF in the Language Kitchen: A Generative Data Augmentation Leveraging Korean Morpheme Ingredients **2023**
Jaehyung Seo, Hyeonseok Moon, Jaewook Lee, Sugyeong Eo, Chanjun Park, Heui-Seok Lim
EMNLP 2023

KEBAP: Korean Error Explainable Benchmark Dataset for ASR and Post-processing **2023**
Seonmin Koo, Chanjun Park, Jinsung Kim, Jaehyung Seo, Sugyeong Eo, Hyeonseok Moon, Heui-Seok Lim
EMNLP 2023

Post-hoc Utterance Refining Method by Entity Mining for Faithful Knowledge Grounded Conversations **2023**
Yoonna Jang, Suhyune Son, Jeongwoo Lee, Junyoung Son, Yuna Hur, Jungwoo Lim, Hyeonseok Moon, Kisu Yang, Heuiseok Lim
EMNLP 2023

PEEP-talk: A situational dialogue-based chatbot for English education **2023**
Seungjun Lee, Yoonna Jang, Chanjun Park, Jungseob Lee, Jaehyung Seo, Hyeonseok Moon, Sugyeong Eo, Seounghoon Lee, Bernardo Yahya, Heui-Seok Lim
ACL 2023 - Demo

Towards diverse and effective question-answer pair generation from children storybooks **2023**
Sugyeong Eo, Hyeonseok Moon(*equal contribution*), Jinsung Kim, Yuna Hur, Jeongwook Kim, Songeun Lee, Changwoo Chun, Sungsoo Park, Heuiseok Lim
ACL 2023 Findings

Improving Formality-Sensitive Machine Translation Using Data-Centric Approaches and Prompt Engineering **2023**
Seungjun Lee, Hyeonseok Moon, Chanjun Park, Heuiseok Lim
IWSLT 2023

QUAK: A synthetic quality estimation dataset for korean-english neural machine translation **2022**
Sugyeong Eo, Chanjun Park, Hyeonseok Moon, Jaehyung Seo, Gyeongmin Kim, Jungseob Lee, Heuiseok Lim
COILING 2022

A dog is passing over the jet? a text-generation dataset for korean commonsense reasoning and evaluation **2022**
Jaehyung Seo, Seounghoon Lee, Chanjun Park, Yoonna Jang, Hyeonseok Moon, Sugyeong Eo, Seonmin Koo, Heuiseok Lim
NAACL 2022 Findings

KU X upstage's submission for the WMT22 quality estimation: Critical error detection shared task 2022
Sugyeong Eo, Chanjun Park, Hyeonseok Moon, Jaehyung Seo, HeuiSeok Lim
WMT 2022

Priming ancient Korean neural machine translation 2022
Chanjun Park, Seolhwa Lee, Jaehyung Seo, Hyeonseok Moon, Sugyeong Eo, Heui-Seok Lim
LREC 2022

Empirical Analysis of Noising Scheme based Synthetic Data Generation for Automatic Post-editing 2022
Hyeonseok Moon, Chanjun Park, Seolhwa Lee, Jaehyung Seo, Jungseob Lee, Sugyeong Eo, HeuiSeok Lim
LREC 2022

A Self-Supervised Automatic Post-Editing Data Generation Tool 2022
Hyeonseok Moon, Chanjun Park, Sugyeong Eo, Jaehyung Seo, SeungJun Lee, Heuiseok Lim
ICML 2022 - DataPerf Workshop

BTS: Back TranScription for Speech-to-Text Post-Processor using Text-to-Speech-to-Text 2021
Chanjun Park, Jaehyung Seo, Seolhwa Lee, Chanhee Lee, Hyeonseok Moon, Sugyeong Eo, Heuiseok Lim
WAT2021 - ACL Workshop

Should we find another model?: Improving neural machine translation performance with one-piece tokenization method without model modification **2021**
Chanjun Park, Sugyeong Eo, Hyeonseok Moon, HeuiSeok Lim
NAACL 2021 - industry track

International Journal

Doubts on the reliability of parallel corpus filtering 2023
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*

PU-GEN: Enhancing generative commonsense reasoning for language models with human-centered knowledge 2022
Jaehyung Seo, Dongsuk Oh, Sugyeong Eo, Chanjun Park, Kisu Yang, Hyeonseok Moon, Kinam Park, Heuiseok Lim
Knowledge-Based Systems

An empirical study on automatic post editing for neural machine translation 2021
Hyeonseok Moon, Chanjun Park, Sugyeong Eo, Jaehyung Seo, Heuiseok Lim
IEEE Access

An automatic post editing with efficient and simple data generation method 2022
Hyeonseok Moon, Chanjun Park, Jaehyung Seo, Sugyeong Eo, Heuiseok Lim
IEEE Access

Exploiting Hanja-based Resources in Processing Korean Historic Documents written by Common Literati 2024
Hyeonseok Moon, Myunghoon Kang, Jaehyung Seo, Sugyeong Eo, Chanjun Park, Yeongwook Yang, Heuiseok Lim
IEEE Access

AI for patents: A novel yet effective and efficient framework for patent analysis 2022
Junyoung Son, Hyeonseok Moon, Jeongwoo Lee, Seolhwa Lee, Chanjun Park, Wonkyung Jung, Heuiseok Lim
IEEE Access

Plain template insertion: korean-prompt-based engineering for few-shot learners 2022
Jaehyung Seo, Hyeonseok Moon, Chanhee Lee, Sugyeong Eo, Chanjun Park, Jihoon Kim, Changwoo Chun, Heuiseok Lim
IEEE Access

Mimicking infants' bilingual language acquisition for domain specialized neural machine translation 2022
Chanjun Park, Woo-Young Go, Sugyeong Eo, Hyeonseok Moon, Seolhwa Lee, Heuiseok Lim
IEEE Access

A survey on evaluation metrics for machine translation 2023
Seungjun Lee, Jungseob Lee, Hyeonseok Moon, Chanjun Park, Jaehyung Seo, Sugyeong Eo, Seonmin Koo, Heuiseok Lim
Mathematics

Comparative analysis of current approaches to quality estimation for neural machine translation **2021**
Sugyeong Eo, Chanjun Park, Hyeonseok Moon, Jaehyung Seo, Heuiseok Lim
Applied Sciences

Return on Advertising Spend Prediction with Task Decomposition-Based LSTM Model **2021**
Hyeonseok Moon, Taemin Lee, Jaehyung Seo, Chanjun Park, Sugyeong Eo, Imatitikua D Aiyanyo, Jeongbae Park, Aram So, Kyoungwha Ok, Kinam Park
Mathematics

Word-level quality estimation for Korean-English neural machine translation **2022**
Sugyeong Eo, Chanjun Park, Hyeonseok Moon, Jaehyung Seo, Heuiseok Lim
IEEE Access

Collaborative Project

LLM Assistant for Teaching Human Consultant **2024.07 – now**
Supported by **Creative Digital Lab** - *Project Manager at Korea University*
- Training Large Language Models with Data Curation
- Data augmentation with a few human-annotated labels

Legal Domain Vertical LLM **2024.07 – now**
Supported by **KT** - *Project Manager at Korea University*
- Training Large Language Models with Data Curation
- Data quality check for building domain specialized LLM
- Constructing data processing pipeline

NLP for Ancient Korean Common Literati Document **2022.06 – 2024.07**
Supported by **National Research Foundation** - *Project Manager at Korea University*
- Named entity recognition and document analysis for ancient Korean documents
- Engaged in data construction process and setup annotation standard
- Related Publication: `Exploiting Hanja-Based Resources in Processing Korean Historic Documents Written by Common Literati` (IEEE Access)

Domain Specialized Parallel Corpus Construction for Machine Translation **2022.06 – 2023.11**
Supported by **NIA (with Minigate Corporation)** - *Project Manager at Korea University*
- Data evaluation and supervision in curation process
- Engaged in data construction process and setup annotation standard

Automated Question-Answer pair Data Generation System **2022.03 – 2023.02**
Supported by **Hyundai Mortors** - *Head Technician at Korea University*
- Automated question-Answer pair generation framework, especially tailored to the educational purpose
- QA generation, Education domain
- Related Publication: `Towards Diverse and Effective Question-Answer Pair Generation from Children Storybooks` (ACL 2023 - findings)

User Query based Recommendation System **2022.03 – 2023.01**
Supported by **FLES corporation** - *Head Technician at Korea University*
- Commercial item recommendation systems based on the user preference
- Recommendation system, Information retrieval

Fortune Telling Generation AI Project **2022.03 – 2023.01**
Supported by **FLES corporation** - *Project Manager at Korea University*
- Fortune telling AI module. Encoder-Decoder generator along with LLM based generator system
- Language generation, Decoding strategy, Large language models
- Related Publication: `SaJuTeller: Conditional Generation Deep-Learning based Fortune Telling Model` (HCLT 2022)

Parallel Corpus Filtering and Mining Research Project **2021.12 – 2022.07**
Supported by **Naver Papago** - *Head Technician at Korea University*
- Analysis on parallel corpus filtering methods targeting Korean-English machine translation
- Parallel corpus filtering, Machine translation
- Related Publication: `Doubts on the reliability of parallel corpus filtering` (Expert Systems with Applications)

Persona-based Dialogue with k-Nearest-Neighbor Approach **2023.05 – 2023.12**
Supported by **NC Soft** - *Head Technician at Korea University*
- Research on the applicability of k-nearest neighbor approach in persona dialogue
- k-nearest neighbor, persona dialogue, language generation

Korean-Prompt-based Engineering for Few-shot Research Project **2022.05 – 2022.07**

Supported by **Hyundai Motors** - *Researcher at Korea University*
- Few-shot prompting strategy for enhancing Korean understanding task performance
- Prompt engineering, Few shot, Language understanding
- Related Publication: `https://ieeexplore.ieee.org/abstract/document/9913979`
  `Plain Template Insertion: Korean-Prompt-Based Engineering for Few-Shot Learners` (IEEE Access)

Information Retrieval system for Industrial Frequently Asked Question **2021.07 – 2022.03**
Supported by **Data Voucher (O2O corporation)** - *Researcher at Korea University*
- Information retrieval system for frequently-asked QA systems
- Keyword Extraction, Information Extraction, Question Answering module

Patent document processing Research Project **2021.06 – 2022.10**
Supported by **LG Innotek** - *Head Technician at Korea University*
- Sentence extraction and key phrase extraction module for patent documents
- Automatic Summarization, Sentence classification, Information Extraction
- Related Publication: `https://ieeexplore.ieee.org/abstract/document/9779775`
  `AI for Patents: A Novel Yet Effective and Efficient Framework for Patent Analysis` (IEEE Access)

Return on Advertising Spend (ROAS) Prediction Project **2021.07 – 2022.03**
Supported by **Data Voucher (BizSpring corporation)** - *Head Technician at Korea University*
- Regression module for return-on-advertising-spend prediction
- Keyword Extraction, Return on Advertising Spend, Regression Model
- Related Publication: `https://www.mdpi.com/2227-7390/10/10/1637`
  `Return on Advertising Spend Prediction with Task Decomposition-Based LSTM Model` (Mathematics)

Teaching

Teaching Assistant at Korea University **2023.09 – 2024.06**
- (DFE610-00) NLP for digital finance engineering
- (BDC101-00) Introduction to Natural Language Processing In Big Data
- (COSE461-02) Natural Language Processing

Honors & Awards

Best Paper Award **2024**
- The 36th Annual Conference on Human & Cognitive Language Technology (HCLT2024)

Best Paper Award **2023**
- The 35th Annual Conference on Human & Cognitive Language Technology (HCLT2023)

1st place in WMT 2022 QE Task 3, 2022 **2022**
- Seventh Conference on Machine Translation (WMT22) Quality Estimation Shared Task

Best Paper Award **2022**
- The 34th Annual Conference on Human & Cognitive Language Technology (HCLT2022)

Best Paper Award **2021**
- The 33rd Annual Conference on Human & Cognitive Language Technology (HCLT2021)

Patent

Device and Method For Generation of Diverse Question-Answer Pair
- U.S. Patent Application No. 18/585,166

Device and Method for Generating Fortune Telling Model Based n Conditional Generation Deep-Learning
- South Korea Patent Granted No. 10-2790031
- South Korea Patent Application No. 10-2022-0158977

Task Decomposition Method based Prediction of Return on Advertising Spend and Device Performing the same
- South Korea Patent Granted No. 10-2593447
- South Korea Patent Application No. 10-2021-0156657

Device and Method for Parallel Corpus Filtering Based On Semantic Similarity
- South Korea Patent Granted No. 10-2593448
- South Korea Patent Application No. 10-2022-0151593

Device and Method for Generating of Training Data for Quality Estimation In Machine Translation
- South Korea Patent Granted No. 10-2593447
- South Korea Patent Application No. 10-2021-0156657

Diverse and Effective Question-Answer Pair Generation System for Education

○ South Korea Patent Application No. 10-2023-0024355

Device And Method For Assessment Of Educational Question-Answering

○ South Korea Patent Application No. 10-2023-0162875

Performance Evaluation Method For Large Language Models Based On Intention Catching Capability

○ South Korea Patent Application No. 10-2024-0169352

Device and Method for Generating Training Data For Post Editing

○ South Korea Patent Application No. 10-2021-0118924