



# Jungseob Lee (이정섭)

NLP Researcher · Graduate Student, Korea University

🏠 1995.10.11    ✉ omanma1928@korea.ac.kr    🌐 js-lee-ai.github.io  
🔄 js-lee-AI    🔗 LinkedIn    🎓 Google Scholar

## Research Interests

---

Large Language Models, Natural Language Processing, Agentic LLM, Machine Translation, Information Retrieval, Model Merging

## Education

---

**M.S./Ph.D. in Computer Science**  
Korea University, Seoul, Korea

2021 – Present  
Advisor: Prof. Heuseok Lim

- NLP Lab — Research on machine translation, cross-lingual transfer learning, and large language models

## Publications

---

### 2026

- “*Improving Semantic Proximity in Information Retrieval through Cross-Lingual Alignment*”  
S. Hong, Y. Jang, **J. Lee**, H. Moon, H. Lim  
**ICLR 2026** (Poster), The Fourteenth International Conference on Learning Representations  
[\[PDF\]](#)

### 2025

- “*Cross-lingual Optimization for Language Transfer in Large Language Models*”  
**J. Lee**, S. Hong, H. Moon, H.S. Lim  
**ACL 2025** (Oral), Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics  
[\[PDF\]](#)
- “*VAETKI Technical Report*”  
NCAI Consortium  
Technical Report, 2025  
[\[PDF\]](#)
- “*An Analysis on Language Transfer of Pre-trained Language Model with Cross-lingual Post-training*”  
S. Son, C. Park, **J. Lee**, M. Shim, C. Lee, Y. Jang, J. Seo, J. Lim, H. Lim  
**Expert Systems With Applications**, 267, 125841  
[\[DOI\]](#)

### 2024

- “*Length-aware Byte Pair Encoding for Mitigating Over-segmentation in Korean Machine Translation*”  
**J. Lee**, H. Moon, S. Lee, C. Park, S. Eo, H. Ko, J. Seo, S. Lee, H.S. Lim

Findings of ACL 2024, pp. 2287–2303

[\[PDF\]](#)

- “*Analysis of the Effectiveness of Model, Data, and User-Centric Approaches for Chat Application: A Case Study of BlenderBot 2.0*”  
C. Park, **J. Lee**, S. Son, K. Park, J. Jang, H. Lim  
**Applied Sciences** 14(11), 4821  
[\[DOI\]](#)

## 2023

- “*A Survey on Evaluation Metrics for Machine Translation*”  
S. Lee, **J. Lee**, H. Moon, C. Park, J. Seo, S. Eo, S. Koo, H. Lim  
**Mathematics** 11(4), 1006  
[\[PDF\]](#)
- “*PEEP-Talk: A Situational Dialogue-based Chatbot for English Education*”  
S. Lee, Y. Jang, C. Park, **J. Lee**, J. Seo, H. Moon, S. Eo, S. Lee, B. Yahya, et al.  
**ACL 2023** (System Demonstrations)  
[\[PDF\]](#)
- “*Doubts on the Reliability of Parallel Corpus Filtering*”  
H. Moon, C. Park, S. Koo, **J. Lee**, S. Lee, J. Seo, S. Eo, Y. Jang, H. Kim, H. Lee, et al.  
**Expert Systems with Applications** 233, 120962  
[\[DOI\]](#)
- “*KoCheckGPT: Korean LLM Written Document Detector*”  
M. Kang, **J. Lee**, S. Lee, S. Hong, J. Park, H. Lim  
Annual Conference on Human and Language Technology (HCLT), pp. 432–436  
[\[PDF\]](#)

## 2022

- “*K-NCT: Korean Neural Grammatical Error Correction Gold-standard Test Set Using Novel Error Type Classification Criteria*”  
S. Koo, C. Park, J. Seo, S. Lee, H. Moon, **J. Lee**, H. Lim  
**IEEE Access** 10, pp. 118167–118175  
[\[PDF\]](#)
- “*Korean and Multilingual Language Models Study for Cross-lingual Post-training (XPT)*”  
S. Son, C. Park, **J. Lee**, M. Shim, C. Lee, K. Park, H. Lim  
**Journal of the Korea Convergence Society** 13(3), pp. 77–89  
[\[PDF\]](#)
- “*QUAK: A Synthetic Quality Estimation Dataset for Korean-English Neural Machine Translation*”  
S. Eo, C. Park, H. Moon, J. Seo, G. Kim, **J. Lee**, H.S. Lim  
**COLING 2022**, Proceedings of the 29th International Conference on Computational Linguistics  
[\[PDF\]](#)
- “*There is No Rose Without a Thorn: Finding Weaknesses on BlenderBot 2.0*”  
**J. Lee**, M. Shim, S. Son, C. Park, Y. Kim, H. Lim  
arXiv preprint arXiv:2201.03239  
[\[PDF\]](#)
- “*Focus on FoCus: Is FoCus Focused on Context, Knowledge and Persona?*”  
S.Y. Lee, **J. Lee**, C. Park, S. Eo, H. Moon, J. Seo, J. Park, H.S. Lim  
Proceedings of the 1st Workshop on Customized Chat Grounding Persona and Knowledge  
[\[PDF\]](#)
- “*Language Chameleon: Transformation Analysis between Languages Using Cross-lingual Post-training*”  
S. Son, C. Park, **J. Lee**, M. Shim, C. Lee, Y. Jang, J. Seo, H. Lim

arXiv preprint arXiv:2209.06422

[\[PDF\]](#)

- “*Empirical Analysis of Noising Scheme based Synthetic Data Generation for Automatic Post-editing*”  
H. Moon, C. Park, S. Lee, J. Seo, **J. Lee**, S. Eo, H.S. Lim  
**LREC 2022**, Proceedings of the Thirteenth Language Resources and Evaluation Conference  
[\[PDF\]](#)
- “*SaJuTeller: Conditional Generation Deep-Learning based Fortune Telling Model*”  
H. Moon, **J. Lee**, J. Seo, S. Eo, C. Park, W. Kim, J. Park, H. Lim  
Annual Conference on Human and Language Technology (HCLT), pp. 277–283  
[\[PDF\]](#)
- “*Korean Language Model Construction and Comparative Analysis with Cross-lingual Post-Training (XPT)*”  
S. Son, C. Park, **J. Lee**, M. Shim, S. Lee, J.W. Lee, A. So, H. Lim  
Annual Conference on Human and Language Technology (HCLT), pp. 295–299  
[\[PDF\]](#)
- “*DART: Data Augmentation using Retrieval Technique*”  
S. Lee, J. Seo, **J. Lee**, M. Kang, H. Moon, C. Park, D. Jung, J. Lee, K. Park, et al.  
Annual Conference on Human and Language Technology (HCLT), pp. 313–319  
[\[PDF\]](#)
- “*Policy-based Performance Comparison Study of Real-time Simultaneous Translation*”  
**J. Lee**, H. Moon, C. Park, J. Seo, S. Eo, S. Lee, S. Koo, H. Lim  
Journal of the Korea Convergence Society 13(3), pp. 43–54  
[\[PDF\]](#)

## 2021

- “*Empirical Study on BlenderBot 2.0’s Errors Analysis in Terms of Model, Data and Dialogue*”  
**J. Lee**, S. Son, M. Shim, Y. Kim, C. Park, A. So, J. Park, H. Lim  
Journal of the Korea Convergence Society 12(12), pp. 93–106  
[\[PDF\]](#)

Last updated: April 6, 2026