Eunseong Choi, Ph.D. Candidate

eunseongchoi.com | eunseong@skku.edu
Google Scholar | in/eunseong | eunseongc

Summary

I'm Eunseong Choi, a Ph.D. candidate in the DIAL research group (advisor: Prof. Jongwuk Lee). My research lies in **Information Retrieval** and **Natural Language Processing**, focusing on **robust and efficient Retrieval-Augmented Generation** frameworks. In particular, I study context–memory conflicts, prompt compression, and reasoning-intensive retrieval to develop scalable and reliable real-world applications.

- **Retrieval-Augmented Generation:** Mitigating context–memory conflicts, incorporating evidentiality, and guiding robust LLM reasoning
- **LLM Efficiency & Compression:** Reducing computational cost through hard and soft prompt compression while preserving semantic fidelity
- **Document Retrieval:** From efficiency-focused sparse retrieval to reasoning-intensive retrieval for complex information needs

Keywords: Retrieval-Augmented Generation, Efficient LLMs, Information Retrieval, NLP

Education

2020.3 – 2026.2 **M.S./Ph.D.**, *Artificial Intelligence*, Sungkyunkwan University
Advisor: Prof.Jongwuk Lee
Thesis: *Improving Evidentiality and Compression for Retrieval-Augmented Generation*

2012.3 – 2020.2 **B.S.**, *Architecture*, Sungkyunkwan University
GPA: 3.92/4.50

B.S., *Samsung Convergence Software Course*, Sungkyunkwan University
GPA: 4.30/4.50

Publications

Under Review

1. Leveraging Lexical Identifiers for Log Anomaly Detection.
Youngbin Kim, Hyunsoo Kim, Jubong Park, **Eunseong Choi**, Jongwuk Lee.
Under Review, 2025.
Improved log anomaly detection by combining LLM-based semantic enrichment with compact lexical identifiers

International Conferences

1. Multi-view-guided Passage Reranking with Large Language Models.
Jeongwoo Na, Jun Kwon, **Eunseong Choi**, Jongwuk Lee.
The 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2025.
Proposed efficient and robust LLM-based passage reranker using multi-view embeddings

2. Conflict-Aware Soft Prompting for Retrieval-Augmented Generation.
Eunseong Choi, June Park, Hyeri Lee, Jongwuk Lee.
The 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2025.
Mitigated context-memory conflicts in retrieval-augmented generation using adversarial soft prompting

3 GRAM: Generative Recommendation via Semantic-aware Multi-granular Late Fusion.
Sunkyung Lee, Minjin Choi, **Eunseong Choi**, Hye-young Kim, Jongwuk Lee.
The 63rd Annual Meeting of the Association for Computational Linguistics (ACL), 2025.
Efficiently leveraging multi-granular semantic information for sequential recommendation

4 From Reading to Compressing: Exploring the Multi-document Reader for Prompt Compression.
Eunseong Choi, Sunkyung Lee, Minjin Choi, June Park, Jongwuk Lee.
Findings of the Association for Computational Linguistics: EMNLP 2024, 2024.
Prompt compression using cross-attention, reducing prompt length by 80% while preserving global context

5 Multi-granularity Guided Fusion-in-Decoder.
Eunseong Choi, Hyeri Lee, Jongwuk Lee.
Findings of the Association for Computational Linguistics: NAACL 2024, 2024.
Improved open-domain QA by leveraging evidence at multiple levels of granularity

6 Forgetting-aware Linear Bias for Attentive Knowledge Tracing.
Yoonjin Im*, **Eunseong Choi***, Heejin Kook, Jongwuk Lee.
In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM) short paper, 2023.
Modeled forgetting behavior with a linear bias in attention-based models

7 ConQueR: Contextualized Query Reduction using Search Logs.
Hye-young Kim*, Minjin Choi*, Sunkyung Lee, **Eunseong Choi**, Young-In Song, Jongwuk Lee.
The 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR) short paper, 2023.
Query reduction by combining core term extraction and sub-query selection from real-world search logs

8 SpaDE: Improving Sparse Representations using a Dual Document Encoder for First-stage Retrieval.
Eunseong Choi*, Sunkyung Lee*, Minjin Choi, Hyeseon Ko, Young-In Song, Jongwuk Lee.
In Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM), 2022.
Improved sparse retrieval by jointly learning term weighting and semantic expansion with a dual encoder

9 Long-tail Mixup for Extreme Multi-label Classification.
Sangwoo Han, **Eunseong Choi**, Chan Lim, Hyunjung Shim, Jongwuk Lee.
In Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM) short paper, 2022.
Addressed label sparsity in extreme classification with mixup-based augmentation

10 MelBERT: Metaphor Detection via Contextualized Late Interaction using Metaphorical Identification Theories.
Minjin Choi, Sunkyung Lee, **Eunseong Choi**, Heesoo Park, Junhyuk Lee, Dongwon Lee, Jongwuk Lee.
Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2021.
Metaphor detection using contextualized word representations and linguistic theories

Domestic Journals & Conferences

1 Improving the Performance of Knowledge Tracing Models using Quantized Correctness Embeddings.
Yoonjin Im, Jaewan Moon, **Eunseong Choi**, Jongwuk Lee.
Journal of KIISE, Vol.50 No.4, 2023.
Knowledge tracing with quantized embeddings encoding correctness and question difficulty

2 Deep Learning based Low-resource Korean Dialect Neural Machine Translation.
Sang-Min Han, **Eunseong Choi**, Jongwuk Lee.
In Proceedings of the Korea Software Congress, 2022.
Neural machine translation for Korean dialects under low-resource settings

3 Pseudo-sentence Representation for Math Word Solving Problem (Best Presentation Award).
Jiwoo Kim, Sunkyung Lee, **Eunseong Choi**, Jongwuk Lee.
In Proceedings of the Korea Computer Congress, 2022.
Math word problem solving with structure-aware pseudo-sentence representations

4 Improving the Accuracy of Knowledge Tracing Models using Quantized Correctness Embeddings (Best paper award).
Yoonjin Im, Jaewan Moon, **Eunseong Choi**, Jongwuk Lee.
In Proceedings of the Korea Computer Congress, 2022.
Knowledge tracing through difficulty-aware correctness embeddings

5 Data Augmentation Methods for Improving thePerformance of Machine Reading Comprehension.
Sunkyung Lee, **Eunseong Choi**, Seonho Jeong, Jongwuk Lee.
Journal of Korean Institute of Information Scientists and Engineers Vol.48 No.12, 2021.
Reading comprehension with word- and sentence-level data augmentation techniques

Experience

2025.3 – 2025.4 **Research Intern, Search & Ranking Modeling**, NAVER Corp.
• Time-dependent user queries require retrievers sensitive to temporal intent.
→ Developed log-based post-training with LLM supervision to better capture temporal signals in dense retrieval

2021.7 – 2021.8 **Research Intern, Search CIC**, NAVER Corp.
• Existing methods that encode queries face efficiency challenges in real-time settings.
→ Designed a uni-encoder-based sparse retrieval approach that improved efficiency while maintaining effectiveness, resulting in a publication at CIKM 2022

Skills

Languages English, Korean
Coding Python, PyTorch, Transformers, FAISS

Honors & Awards

2024 **1st Prize, Best Graduate Research Paper Award**, Sungkyunkwan University.
Presented research on prompt compression for efficient LLM

2023 **2nd Place, AI Grand Challenge for Policy Support AI**, IITP.
Developed a multi-hop retriever and re-ranker for policy-supportive documents

2022 **1st Place, AI Grand Challenge for Math Word Problem Solving**, IITP.
Built a system for math word problem solving using domain-specific operations

2021 **2nd Prize, Best Graduate Research Paper Award**, Sungkyunkwan University.
Presented research on efficient sparse retrieval using term expansion and re-weighting

Scholarships

2020 – 2022 **Graduate School Scholarship**, Department of Artificial Intelligence, SKKU.
2019 **Academic Excellence Scholarship**, Sungkyunkwan University.
2017 – 2019 **National Scholarship for Science and Engineering Students**, KOSAF
2015 – 2016 **Academic Excellence Scholarship**, Sungkyunkwan University.

Teaching Assistant

Spring 2023	**Introduction to Recommender Systems**
Fall 2022	**Fundamentals of Machine Learning**
Fall 2021	**Deep Neural Networks**
Spring 2020 – 2022	**Introduction to Database**

2024	**Machine Learning** (LG Electronics)
2022	**Machine Learning** (LG Electronics, Samsung SDS)
2021	**Machine Learning** (LG Electronics)
2020	**Machine Learning** (LG Electronics, Samsung SDS, SK Innovation)