Eunseong Choi, Ph.D. Candidate

eunseongchoi.com | eunseong@skku.edu | 010-4918-1115
Google Scholar | in/eunseong | eunseongc

Summary

I'm Eunseong Choi, a Ph.D. candidate in the Data Intelligence and Learning Lab, Sungkyunkwan University (advisor: Prof. Jongwuk Lee), expected to complete my degree in February 2026. My research lies in **Information Retrieval** and **Natural Language Processing**, with an emphasis on building **robust and efficient Retrieval-Augmented Generation** frameworks. My work has received 200+ citations and has been published in top-tier conferences in NLP and IR. Representative research includes:

- **Retrieval-Augmented Generation:**
  – Context–memory conflicts [C1]
  – Prompt compression for efficient LLMs [C4]
- **Information Retrieval:**
  – Passage re-ranking [C2]
  – Learned sparse representation [C8]
  – Generative recommendation [C3]
  – Query reduction [C7]
- **Natural Language Processing:**
  – Metaphor detection [C10]
- **Machine Learning:**
  – Knowledge tracing [C6]
  – Extreme multi-label classification [C9]

In addition to my academic research, I gained industry experience as a research intern at NAVER, working on large-scale real-world search problems.

Education

2020.3 – current **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

B.S., *Samsung Convergence Software Course*, Sungkyunkwan University

Experience

2025.3 – 2025.4 **Research Intern**, Search & Ranking Modeling, NAVER (Supervisor: Young-In Song)
• Worked on continual fine-tuning for recency-aware dense retrieval

2021.7 – 2021.8 **Research Intern**, Search CIC, NAVER (Supervisor: Young-In Song)
• Developed a sparse retrieval method for efficient first-stage retrieval

Publications

[C1] Conflict-Aware Soft Prompting for Retrieval-Augmented Generation *To appear in EMNLP 2025*
Eunseong Choi, June Park, Hyeri Lee, Jongwuk Lee
• Mitigated context–memory conflicts in retrieval-augmented generation using adversarial soft prompting

[C2] Multi-view-guided Passage Reranking with Large Language Models *To appear in EMNLP 2025*
Jeongwoo Na, Jun Kwon, **Eunseong Choi**, Jongwuk Lee
• Proposed efficient and robust LLM-based passage reranker using multi-view embeddings

[C3] GRAM: Generative Recommendation via Semantic-aware Multi-granular Late Fusion *ACL 2025*
Sunkyung Lee, Minjin Choi, **Eunseong Choi**, Hye-young Kim, Jongwuk Lee
• Efficiently leveraging multi-granular semantic information for sequential recommendation

[C4] From Reading to Compressing: Exploring the Multi-document Reader for Prompt Compression *EMNLP 2024 Findings*
Eunseong Choi, Sunkyung Lee, Minjin Choi, June Park, Jongwuk Lee
• Prompt compression using cross-attention, reducing prompt length by 80% while preserving global context

[C5] Multi-granularity Guided Fusion-in-Decoder *NAACL 2024 Findings*
Eunseong Choi, Hyeri Lee, Jongwuk Lee
• Improved open-domain QA by leveraging evidence at multiple levels of granularity

[C6] Forgetting-aware Linear Bias for Attentive Knowledge Tracing *CIKM 2023 (short paper)*
Yoonjin Im*, **Eunseong Choi***, Heejin Kook, Jongwuk Lee
• Modeled forgetting behavior with a linear bias in attention-based models

[C7] ConQueR: Contextualized Query Reduction using Search Logs *SIGIR 2023 (short paper)*
Hye-young Kim*, Minjin Choi*, Sunkyung Lee, **Eunseong Choi**, Young-In Song, Jongwuk Lee
• Query reduction by combining core term extraction and sub-query selection from real-world search logs

[C8] SpaDE: Improving Sparse Representations using a Dual Document Encoder for First-stage Retrieval *CIKM 2022*
Eunseong Choi*, Sunkyung Lee*, Minjin Choi, Hyeseon Ko, Young-In Song, Jongwuk Lee
• Improved sparse retrieval by jointly learning term weighting and semantic expansion with a dual encoder

[C9] Long-tail Mixup for Extreme Multi-label Classification *CIKM 2022 (short paper)*
Sangwoo Han, **Eunseong Choi**, Chan Lim, Hyunjung Shim, Jongwuk Lee
• Addressed label sparsity in extreme classification with mixup-based augmentation

[C10] MelBERT: Metaphor Detection via Contextualized Late Interaction using Metaphorical Identification Theories *NAACL 2021*
Minjin Choi, Sunkyung Lee, **Eunseong Choi**, Heesoo Park, Junhyuk Lee, Dongwon Lee, Jongwuk Lee
• Metaphor detection using contextualized word representations and linguistic theories

Academic Service

Reviewer — ARR 2025 May (EMNLP 2025)

External Reviewer — EMNLP 2022; ACL 2023; SIGIR 2023–2024; KDD 2022, 2026; WSDM 2022; ARR 2025 Feb (ACL 2025)

Honors & Awards

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

2023 **2nd Place (Minister's award), AI Grand Challenge for Policy Support AI**, IITP.
Developed a multi-hop retriever and re-ranker for generating policy-supportive documents

2022 **1st Place (Minister's award), AI Grand Challenge for Math Word Problem Solving**, IITP.
Developed a framework 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 AI, Sungkyunkwan University

2019 **Academic Excellence Scholarship**, Sungkyunkwan University

2017 – 2019 **National Scholarship for Science and Engineering Students**, Korea Student Aid Foundation

2015 – 2016 **Academic Excellence Scholarship**, Sungkyunkwan University.

Teaching Assistant

Regular Courses

Spring 2023 **Introduction to Recommender Systems**

Fall 2022 **Fundamentals of Machine Learning**

Fall 2021 **Deep Neural Networks**

Spring 2020 – 2022 **Introduction to Database**

Extracurricular Courses

Machine Learning **LG Electronics** May 2020, Oct 2020, May 2021, Oct 2021, Oct 2022, Oct 2024
Samsung SDS Jun 2020
SK Innovation Sep 2020
Machine learning basics, including regression and neural networks

NLP Project **Samsung SDS** Sep 2022
Covered topic classification and machine reading comprehension

Industry–Academia Projects

2025.09 – 2026.02 **NL2SQL-based Intelligent Question Answering System Development**
WeZON *Project Manager* - Developed an NL2SQL system enabling non-experts to query databases in natural language

2021.07 – 2022.06 **Document Retrieval via Pre-trained Language Models and Inverted Index**
NAVER *Team Leader* - Designed a dual-document encoder for effective sparse representations

2020.05 – 2021.04 **Topic-aware Query–Document Matching with Deep Neural Networks**
NAVER Leveraged VAE-based topic modeling to represent query–document semantics

References

Available upon request