Soyeong Jeong

Updated November 11, 2025

Email: starsuzi@kaist.ac.kr **Homepage**: https://starsuzi.github.io **Google Scholar**: /Soyeong Jeong

Research Interest

My research interests are mainly on Retrieval-Augmented Generation (RAG) for solving open-domain language tasks and interpretation of Large Language Models (LLMs) to enhance their interpretability in real-world applications. Not limited to, I am interested in broad topics on natural language understanding.

Education

KAIST — Daejeon, Korea
Ph.D. in Graduate school of AI (Advisor: Prof. Sung Ju Hwang) — Mar 2022 – Present
M.S. in School of Computing — Mar 2020 – Feb 2022
Thesis: Information Retrieval by Augmenting Document Representation

Korea University — Seoul, Korea
B.S. in Computer Science and Engineering (Graduated with Honor) — Mar 2016 – Feb 2020
B.E. in Software Technology and Enterprise Program (Interdisciplinary Program)

Anyang Foreign Language High School — Gyeonggi, Korea
Prestigious high school for talented students (Major in English) — Mar 2013 – Feb 2016

Employment

Applied Scientist Intern, Amazon — Bellevue, WA, USA
• Conducted research on Long-Context LMs and their reasoning capabilities. — Jun 2025 - Oct 2025

Applied Scientist Intern, Amazon — Bellevue, WA, USA
• Conducted research on a multi-agent framework for conversational tasks. — Jul 2024 - Oct 2024

Undergrad. Research Assistant, Korea University (Advisor: Prof. Jaewoo Kang) — Seoul, Korea
• Participated in the major recommendation project by embedding curriculum vectors. — Mar 2019 - Feb 2020

Research Intern, Seoul SW-SoC Convergence R&BD Center, ETRI — Seoul, Korea
• Participated in the Speech Emotion Recognition project. — Jul 2019 - Aug 2019

Research Intern, Artificial Intelligence Research Laboratory, ETRI — Daejeon, Korea
• Participated in the Artificial Social Intelligence for Human-Care Robots project. — Jan 2019 - Feb 2019

Publications

International Publications

[37] When Thoughts Meet Facts: Reusable Reasoning for Long-Context LMs
Soyeong Jeong, Taehee Jung, Sung Ju Hwang, Joo-Kyung Kim, and Dongyeop Kang
Under review

[36] UniversalRAG: Retrieval-Augmented Generation
over Multiple Corpora with Diverse Modalities and Granularities
Woongyeong Yeo*, Kangsan Kim*, Soyeong Jeong, Jinheon Baek, and Sung Ju Hwang
Under review

[35] PRISM: Fine-Grained Paper-to-Paper Retrieval with Multi-Aspect-Aware Query Optimization
Sangwoo Park, Jinheon Baek, Soyeong Jeong, and Sung Ju Hwang
Under review

[34] Adaptive Multi-Agent Response Refinement in Conversational Systems
Soyeong Jeong, Aparna Elangovan, Emine Yilmaz, and Oleg Rokhlenko
Linguistic and Cognitive Approaches to Dialog Agents Workshop at AAAI **(LaCATODA @ AAAI)**, 2026. **(Oral)**

[33] Database-Augmented Query Representation for Information Retrieval
Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang, and Jong C. Park
Empirical Methods in Natural Language Processing **(EMNLP)**, 2025. **(Oral)**

[32] The RAG Paradox: A Black-Box Attack Exploiting Unintentional Vulnerabilities
in Retrieval-Augmented Generation Systems
Chanwoo Choi, Jinsoo Kim, Sukmin Cho, Soyeong Jeong, and Buru Chang
Findings of the Empirical Methods in Natural Language Processing **(Findings of EMNLP)**, 2025.

[31] CaMMT: Benchmarking Culturally Aware Multimodal Machine Translation
Emilio Villa-Cueva, Sholpan Bolatzhanova, Diana Turmakhan, Kareem Elzeky…,
Jinheon Baek, …, Soyeong Jeong, …, Injy Hamed, Atnafu Lambebo Tonja, and Thamar Solorio
Findings of the Empirical Methods in Natural Language Processing **(Findings of EMNLP)**, 2025.

[30] Upcycling Candidate Tokens of Large Language Models for Query Expansion
Jinseok Kim, Sukmin Cho, Soyeong Jeong, Sangyeop Kim, and Sungzoon Cho
The Conference on Information and Knowledge Management **(CIKM)**, 2025.

[29] VideoRAG: Retrieval-Augmented Generation over Video Corpus
Soyeong Jeong*, Kangsan Kim*, Jinheon Baek*, and Sung Ju Hwang
Findings of the Association for Computational Linguistics **(Findings of ACL)**, 2025.

[28] EXIT: Context-Aware Extractive Compression for Enhancing Retrieval-Augmented Generation
Taeho Hwang, Sukmin Cho, Soyeong Jeong, Hoyun Song, SeungYoon Han, and Jong C. Park
Findings of the Association for Computational Linguistics **(Findings of ACL)**, 2025.

[27] Temporal Information Retrieval via Time-Specifier Model Merging
SeungYoon Han, Taeho Hwang, Sukmin Cho, Soyeong Jeong, Hoyun Song, Huije Lee, and Jong C. Park
Towards Knowledgeable Foundation Models Workshop at ACL **(KnowFM @ ACL)**, 2025.

[26] Unified Multi-Modal Interleaved Document Representation for Information Retrieval
Jaewoo Lee*, Joonho Ko*, Jinheon Baek*, Soyeong Jeong, and Sung Ju Hwang
Vector Databases Workshop at ICML **(VecDB @ ICML)**, 2025.

[25] Lossless Acceleration of Large Language Models with Hierarchical Drafting
based on Temporal Locality in Speculative Decoding
Sukmin Cho, Sangjin Choi, Taeho Hwang, Jeongyeon Seo, Soyeong Jeong,
Huije Lee, Hoyun Song, Jong C. Park, and Youngjin Kwon
Findings of the Nations of the Americas Chapter of the Association for Computational Linguistics
(Findings of NAACL), 2025.

[24] CVQA: Culturally-diverse Multilingual Visual Question Answering Benchmark
David Romero, Chenyang Lyu, Haryo Akbarianto Wibowo, Teresa Lynn, Injy Hamed,
Aditya Nanda Kishore, …, Jinheon Baek, …, Soyeong Jeong, …, Thamar Solorio, and Alham Fikri Aji
Neural Information Processing Systems Datasets and Benchmarks Track **(NeurIPS D&B)**, 2024. **(Oral)**

[23] Typos that Broke the RAG's Back: Genetic Attack on RAG Pipeline
by Simulating Documents in the Wild via Low-level Perturbations
Sukmin Cho, Soyeong Jeong, Jeongyeon Seo, Taeho Hwang, and Jong C. Park
Findings of the Empirical Methods in Natural Language Processing **(Findings of EMNLP)**, 2024.

[22] Ask LLMs Directly, "What shapes your bias?": Measuring Social Bias in Large Language Models
Jisu Shin, Hoyun Song, Huije Lee, Soyeong Jeong, and Jong C. Park
Findings of the Association for Computational Linguistics **(Findings of ACL)**, 2024.

[21] DSLR: Document Refinement with Sentence-Level Re-ranking and Reconstruction
to Enhance Retrieval-Augmented Generation
Taeho Hwang, Soyeong Jeong, Sukmin Cho, SeungYoon Han, and Jong C. Park
Knowledge Augmented Methods for NLP Workshop at ACL **(KnowledgeNLP @ ACL)**, 2024.

[20] Adaptive-RAG: Learning to Adapt Retrieval-Augmented Large Language Models
through Question Complexity
Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang, and Jong C. Park
North American Chapter of the Association for Computational Linguistics **(NAACL)**, 2024.

[19] Self-Knowledge Distillation for Learning Ambiguity
Hancheol Park, Soyeong Jeong, Sukmin Cho, and Jong C. Park
arXiv preprint, 2024

[18] Test-Time Self-Adaptive Small Language Models for Question Answering
Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang, and Jong C. Park
Findings of the Empirical Methods in Natural Language Processing **(Findings of EMNLP)**, 2023.

[17] Knowledge-Augmented Language Model Verification
Jinheon Baek, Soyeong Jeong, Minki Kang, Jong C. Park, and Sung Ju Hwang
Empirical Methods in Natural Language Processing **(EMNLP)**, 2023.

[16] Improving Zero-shot Reader by Reducing Distractions
from Irrelevant Documents in Open-Domain Question Answering

Sukmin Cho, Jeongyeon Seo, Soyeong Jeong, Jong C. Park
Findings of the Empirical Methods in Natural Language Processing **(Findings of EMNLP)**, 2023.

[15] Phrase Retrieval for Open-Domain Conversational Question Answering
with Conversational Dependency Modeling via Contrastive Learning
Soyeong Jeong, Jinheon Baek, Sung Ju Hwang, and Jong C. Park
Findings of the Association for Computational Linguistics **(Findings of ACL)**, 2023.

[14] Discrete Prompt Optimization via Constrained Generation for Zero-shot Re-ranker
Sukmin Cho, Soyeong Jeong, Jeong yeon Seo, Jong C. Park
Findings of Association for Computational Linguistics **(Findings of ACL)**, 2023.

[13] Realistic Conversational Question Answering with Answer Selection
based on Calibrated Confidence and Uncertainty Measurement
Soyeong Jeong, Jinheon Baek, Sung Ju Hwang, and Jong C. Park
Conference of the European Chapter of the Association for Computational Linguistics **(EACL)**, 2023.

[12] Augmenting Document Representations for Dense Retrieval with Interpolation and Perturbation
Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang, and Jong C. Park
Annual Meeting of the Association for Computational Linguistics **(ACL)**, 2022. **(Oral)**

[11] Query Generation with External Knowledge for Dense Retrieval
Sukmin Cho, Soyeong Jeong, Wonsuk Yang, Jong C. Park
Deep Learning Inside Out Workshop at ACL **(DeeLIO @ ACL)**, 2022.

[10] Unsupervised Document Expansion for Information Retrieval with Stochastic Text Generation
Soyeong Jeong, Jinheon Baek, ChaeHun Park, and Jong C. Park
Scholarly Document Processing Workshop at NAACL **(SDP @ NAACL)**, 2021. **(Oral)**

[9] Development of Speech Emotion Recognition Algorithm using MFCC and Prosody
Hyejin Koo, Soyeong Jeong, Sungjae Yoon, Wonjong Kim
International Conference on Electronics, Information, and Communication **(ICEIC)**, 2020.

Domestic Publications, mostly written in Korean

[8] Retrieval-Augmented Generation through Zero-shot Sentence-Level Passage Refinement using LLMs
Taeho Hwang, Soyeong Jeong, Sukmin Cho, and Jong C. Park
Korea Computer Congress (KCC), 2024. **(Best Paper)**

[7] Controllable prompt tuning with relation dependent tokens
Jinseok Kim, Sukmin Cho, Soyeong Jeong, and Jong C. Park
Korea Computer Congress (KCC), 2023.

[6] Stopwords Mask Pooling for Dense Retrieval in Medical Domain
Dongho Choi, Hoyun Song, Soyeong Jeong, Sukmin Cho, and Jong C. Park
Korea Computer Congress (KCC), 2022. **(Best Presentation)**

[5] Assessing automatic summarization model as a reading assistant
Aujin Kim, Jisu Shin, Soyeong Jeong, Sukmin Cho, and Jong C. Park
Korea Computer Congress (KCC), 2022.

[4] Calibration of Pre-trained Language Model for the Korean Language
Soyeong Jeong, Wonsuk Yang, ChaeHun Park, Jong C. Park
Journal of KIISE (JOK), 2021.

[3] Calibration of Pre-trained Language Model for Korean
Soyeong Jeong, Wonsuk Yang, ChaeHun Park, Jong C. Park
Human & Cognitive Language Technology (HCLT), 2020. **(Best Paper)**

[2] Embedding Academic Majors and Lectures for Analyzing Departments in University
Jinheon Baek, Gwanghoon Jang, Soyeong Jeong, Donghyeon Park, Kiwon Kwon, Jaewoo Kang
Korea Computer Congress (KCC), 2019.

Thesis

[1] Information Retrieval by Augmenting Document Representation
Soyeong Jeong
Master's Thesis, KAIST, 2022

Honors and Awards

Awarded the Presidential Science Scholarship for Graduate Study — 2025
Selected as One of the Great Reviewers within ACL ARR 2025 Cycles — 2025
Awarded a Ph.D. fellowship from NRF (National Research Foundation) of Korea — 2023
Title: Realistic Open-domain Question Answering System with Large Language Models
Received the Best Paper Award at HCLT 2020 — 2020
Graduated with Honor — 2019
Computer Science and Engineering Department at Korea University
Received the First Prize in the Graduation Project, Competition Among Around 20 Teams — 2019
Computer Science and Engineering Department at Korea University
Nominated as Semester High Honors (Spring 2019) — 2019
Nominated as Semester High Honors (Fall 2018) — 2018
Received the Second Prize for the iOS Hackathon at Korea University — 2018
Awarded as the Finalist, Competition of 21:1 with 735 total submissions — 2017
LG Global Challenger

Academic Services

Reviewer of **NeurIPS** (Conference on Neural Information Processing Systems) — 2025
Reviewer of **COLM** (Conference on Language Modeling) — 2025
Reviewer of **ACL ARR** (Association for Computational Linguistics Rolling Review) — 2023 - 2025
Reviewer of **TALLIP** (ACM Transactions on Asian and Low-Resource Language Information Processing) — 2022
Reviewer of **ACL Demo Track** — 2021

Talks and Seminars

Towards Adaptive and Multimodal Retrieval-Augmented Generation
Google ExploreCSR — Mar 2025

Teaching

Machine Learning (CS376)
KAIST Spring 2024 (Teaching Assistant)
Computational Linguistics (CS579)
KAIST Fall 2022 (Teaching Assistant)
Natural Language Processing with Python (CS372)
KAIST Spring 2020, Spring 2021, Spring 2023 (Teaching Assistant)

Skills

Languages: Korean (mother tongue), English (fluent)
Programming: Python, C, Java, Swift