Soyeong Jeong

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   Anyang, Korea
Prestigious high school for talented students (Major in English)   Mar 2013 – Feb 2016

EMPLOYMENT

Applied Scientist Intern, Amazon   Jun 2025 - Sep 2025
- Will conduct research on LLMs and their general intelligence.

Applied Scientist Intern, Amazon   Jul 2024 - Oct 2024
- Conducting research on a proactive conversational task with LLM-powered multi-agents.

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

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

Research Intern, Artificial Intelligence Research Laboratory, ETRI   Jan 2019 - Feb 2019
- Participated in the AIR project (Developing Artificial Social Intelligence for Human-Care Robots).

PUBLICATIONS

International Publications

[32] VideoRAG: Retrieval-Augmented Generation over Video Corpus
Soyeong Jeong*, Kangsan Kim*, Jinheon Baek*, and Sung Ju Hwang
Under review

[31] Adaptive Multi-Agent Response Refinement in Conversational Systems
Soyeong Jeong, Aparna Elangovan, Emine Yilmaz, and Oleg Rokhlenko
Under review

[30] Database-Augmented Query Representation for Information Retrieval
Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang, and Jong C. Park
Under review

[29] Unified Multi-Modal Interleaved Document Representation for Information Retrieval
Jaewoo Lee*, Joonho Ko*, Jinheon Baek*, Soyeong Jeong, and Sung Ju Hwang
Under review

[28] 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
Under review

[27] EXIT: Context-Aware Extractive Compression for Enhancing Retrieval-Augmented Generation
Taeho Hwang, Sukmin Cho, Soyeong Jeong, Hoyun Song, SeungYoon Han, and Jong C. Park
Under review

[26] Incorporating Discarded Candidate Tokens of Large Language Models for Efficient Query Expansion
Jinseok Kim, Sukmin Cho, Soyeong Jeong, Sangyeop Kim, and Sungzoon Cho
Under review

[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 **(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, <u>Soyeong Jeong</u>, Wonsuk Yang, Jong C. Park
Deep Learning Inside Out at Association for Computational Linguistics **(DeeLIO @ ACL)**, 2022.

[10] Unsupervised Document Expansion for Information Retrieval with Stochastic Text Generation
<u>Soyeong Jeong</u>, 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, <u>Soyeong Jeong</u>, 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, <u>Soyeong Jeong</u>, 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, <u>Soyeong Jeong</u>, and Jong C. Park
Korea Computer Congress (KCC), 2023.

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

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

[3] Calibration of Pre-trained Language Model for Korean
<u>Soyeong Jeong</u>, 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, <u>Soyeong Jeong</u>, Donghyeon Park, Kiwon Kwon, Jaewoo Kang
Korea Computer Congress (KCC), 2019.

Thesis

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

HONORS AND AWARDS	Awarded a Ph.D. fellowship from NRF (National Research Foundation) of Korea Title: Realistic Open-domain Question Answering System with Large Language Models	2023
	Received the Best Paper Award at HCLT 2020	2020
	Graduated with Honor Computer Science and Engineering Department at Korea University	2019
	Received the First Prize in the Graduation Project, Competition Among Around 20 Teams Computer Science and Engineering Department at Korea University	2019
	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 LG Global Challenger	2017

ACADEMIC SERVICES	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

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