

Email: [starsuzi@kaist.ac.kr](mailto:starsuzi@kaist.ac.kr)Homepage: <https://starsuzi.github.io>Google Scholar: [/Soyeong Jeong](#)RESEARCH  
INTEREST

My research interests are mainly on information retrieval for solving open domain language tasks and interpretation of large language models for making them interpretable when deployed on real-world applications. Not limited to, I am interested in broad topics on natural language understanding.

## EDUCATION

**KAIST**

Ph.D. in School of Computing

Daejeon, Korea

M.S. in School of Computing

Mar 2022 – Present

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

Major in English

Mar 2013 – Feb 2016

## EMPLOYMENT

**Graduate student, KAIST** (Advisor: Prof. Jong Cheol Park)

Mar 2020 - Present

- Conducted research on the question-answering task with test-time adaption by investigating the capabilities of smaller self-adaptive LMs, only with unlabeled test data.
- Conducted research on the open conversational question-answering task by addressing the problem of an expensive retriever-reader structure with a phrase retrieval model.
- Conducted research on the realistic conversational question-answering task by replacing an unrealistic ground-truth history evaluation scheme with confidence- and uncertainty-based answer filtering schemes, which are further calibrated to obtain reliable predictions.
- Conducted research to tackle the problem of insufficient labeled training data by augmenting dense document representations with interpolation and perturbation on existing dense retrieval methods.
- Conducted research on solving the vocabulary mismatch problem in information retrieval by expanding documents unsupervisedly through diversely generated summaries with pre-trained language models.
- Conducted research on interpreting pre-trained language models based on calibration and analyzed whether the models can capture semantic changes by identifying grammatical characteristics as humans.

**Undergrad. Research Assistant, Korea University** (Advisor: Prof. Jaewoo Kang)

Mar 2019 - Feb 2020

- Participated in the major recommendation project by embedding curriculum vectors.
- Participated in the food ingredient & drug graph construction project.

**Research Intern, Seoul SW-SoC Convergence R&BD Center, ETRI**

Jul 2019 - Aug 2019

- Participated in the Speech Emotion Recognition project and developed a neural network architecture to classify emotions from the input voice.

**Research Intern, Artificial Intelligence Research Laboratory, ETRI**

Jan 2019 - Feb 2019

- Participated in the AIR project (Developing Artificial Social Intelligence for Human-Care Robots) and developed a web crawler for gathering, annotating, and analyzing images and metadata.

## PUBLICATIONS

**International Publications** (\*: equal contribution; †: corresponding author)

- [17] 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
- [16] Learning to Verify Knowledge-Augmented Language Models  
Jinheon Baek, Soyeong Jeong, Minki Kang, Jong C. Park, and Sung Ju Hwang  
Empirical Methods in Natural Language Processing (**EMNLP**), 2023
- [15] Alleviating Hallucination in Large Language Models as Zero-shot Reader  
for 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
- [14] 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
- [13] 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.
- [12] 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
- [11] 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**)
- [10] Query Generation with External Knowledge for Dense Retrieval  
 Sukmin Cho, Soyeong Jeong, Wonsuk Yang, Jong C. Park  
 Deep Learning Inside Out at Association for Computational Linguistics (**DeeLIO @ ACL**), 2022.
- [9] 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**)
- [8] 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

- [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**

Won 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 <b>TALLIP 2022</b> (ACM Transactions on Asian and Low-Resource Language Information Processing)	2022
	Reviewer of <b>ACL-IJCNLP Demo Track 2021</b>	2021
TEACHING	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	