

Jonghyun (Jong) Song

 yc-song |  Jonghyun Song |  hyeongoon11@snu.ac.kr |  +82.10.9418.8227

RESEARCH INTEREST

- Language representation learning for information retrieval, including dense and sparse retrieval methods
- Internal mechanisms and representational dynamics of language models, particularly encoder-based architectures
- Applications of information retrieval in downstream tasks, such as Retrieval-Augmented Generation (RAG) and Task-Aware Language Models (TALM)
- Development and adaptation of personalized language models for user-specific tasks and preferences
- **Keywords:** Natural Language Processing (NLP), Information Retrieval, Retrieval-Augmented Generation, Multi-modal Language Models, Large Language Models, Representation Learning

EDUCATION

Seoul National University, Seoul, Korea

Mar. 2022 - Present

Ph.D. in Data Science

GPA: 3.95/4.3

Advisor: Yohan Jo

Course Highlights: Machine Learning & Deep Learning, Machine Learning for Visual Understanding, Conversational AI for Dialogue System

Seoul National University, Seoul, Korea

Mar. 2017 - Feb. 2022

B.S., Cum Laude, in Mechanical Engineering

GPA: 3.88/4.3

Undergrad thesis: Wrist Wearable Robot for Work-Related Musculoskeletal Disorders Prevention

Advisor: Kyu-Jin Cho

Course Highlights: Machine Learning and Elementary Math, Introduction to Robotics, Introduction to Computer Programming

PAPERS AND PRESENTATIONS

1. Joint Sparse-Dense Optimization for Learned Sparse Text-Image Retrieval

CIKM 2025 Short Research Paper Track (Proceedings of the 34th ACM International Conference on Information and Knowledge Management)

Jonghyun Song, YoungJune Lee, Gyu-hwung Cho, Ilhyeon Song, Saehun Kim and Yohan Jo

2. Comparing Neighbors Together Makes it Easy: Jointly Comparing Multiple Candidates for Efficient and Effective Retrieval

EMNLP 2024 Main Track (Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing) / Spotlight Talk at 9th Workshop on Representation Learning for NLP in ACL 2024

Jonghyun Song, Cheyon Jin, Wenlong Zhao, Andrew McCallum and Jay-Yoon Lee

WORK EXPERIENCE

Research Intern

Jan. 2025 – Mar. 2025

NAVER Corporation, Gyeonggi, Korea

- Designed and implemented a multimodal item-to-item recommender system using both dense and sparse document embeddings (text + image), achieving a **+4.3p** gain in nDCG@1 on real-world clicklog test sets.
- Fine-tuned multilingual CLIP (m-CLIP) with pairwise and triplet contrastive learning on large-scale user interaction data.
- Constructed a high-quality training dataset by leveraging LLM-based filtering to enhance relevance and learning stability.
- Conducting ongoing research on *SPLADE-CLIP* for lightweight and interpretable sparse image representations.

Research Assistant (Ph.D. Student) under Professor Yohan Jo

Jul. 2022 – Present

Seoul National University, Seoul, Korea

- Project: Jointly Comparing Multiple Candidates for Efficient and Effective Retrieval
 - Proposed the Comparing Multiple Candidates (CMC) framework to improve the retrieve-and-rerank pipeline.
 - Employed shallow self-attention layers to jointly compare query and candidate embeddings, enabling scalable and efficient multiple comparisons.
 - Achieved strong performance across entity linking, passage ranking, and dialogue ranking tasks, with improved latency and memory efficiency.
 - One paper accepted to **EMNLP 2024 (main track)**

Research Internship under Professor Kyu-Jin Cho

Jul. 2020 – Dec. 2021

Seoul National University, Seoul, Korea

- Project: Soft Wearable Robot for Preventing Musculoskeletal Disorders at the Wrist
- Developed wearable robotic devices that regulate compression based on human intention to prevent work-related musculoskeletal disorders (WMSDs) in the wrist. Specifically:
 - Performed physical modeling of cable routing to maximize power transmission efficiency
 - Designed a silicone component embedded with bearings and fabric to improve mobility and portability
 - Built Arduino-based robotic control systems using force-sensitive resistors (FSRs)

Founder & Software Engineer

Sep. 2019 – Jun. 2020

Hakwongo Corp. Seongnam, Korea

- Founded a startup that connects working mothers with private education institutes using deep learning technologies. (Funded by Seongnam City and Yonsei University)
 - Developed a natural language processing (NLP) model to recommend educational institutes tailored to working mothers' needs
 - Built the Android application frontend using the Flutter framework
 - Processed and managed educational institute database using SQL and pandas

AWARDS AND HONORS

2nd Place , VRDIU Competition (Track A) on IJCAI 2024 sponsored by Google Research	Jul. 2024
– Task: predicting the Region-of-Interests (RoIs) that can provide correct answer to given questions	
– Fine-tuned SOTA model (LayoutLMv3) with a token classifier for predicting the answer span (97.9 F1)	
– Served as a team leader	
1st Place (Minister's Award) on K-Datascience Hackathon, Ministry of Science and ICT, Korea	Nov. 2023
– Presented <i>Multi-modal and Multi-view Patent Search System</i> , a patent search engine with CLIP embeddings of drawings and text	
– Utilized self-supervised learning, using 'prior art' section in patents as a pseudo-label	
– Implemented a chatbot interface with LangChain and Streamlit	
– Served as a team leader	
Park Min-Chul Data Science Challenge Scholarship , Seoul National University, Korea	Mar. 2022
Cum Laude , Seoul National University, Korea	Feb. 2022
Sanhak (Industrial-Educational Cooperation) Foundation Scholarship , Korea Sanhak Foundation, Korea	Mar. 2021 – Dec. 2021
Merit-Based Scholarship , Seoul National University, Korea	Dec. 2018 – Dec. 2019

TEACHING EXPERIENCE

TA , <i>AI Expert Training Project</i> , Samsung Electronics	July 2024
Head TA , <i>Natural Language Processing with Neural Networks</i> , Seoul National University	Fall 2023
Instructor , <i>Big Data Fintech Specialist Training Course</i> , Ministry of Employment and Labor	Fall 2023
Head TA , <i>Math and Statistics Foundations for Data Science</i> , Seoul National University	Spring 2023
Head TA , <i>Applications of Natural Language Processing</i> , Seoul National University	Fall 2022
TA , <i>Data science Boot Camp</i> , Seoul National University	Fall 2022
Math Tutor , <i>Self-Paced Learning & Tutoring Program</i> , Seoul National University	Winter 2020
Undergraduate TA , <i>Creative Engineering Design</i> , Seoul National University	Fall 2019

TECHNICAL SKILLS

Languages	Python, MATLAB, C++, C, SQL, Arduino
Library & Tools	Pytorch, Huggingface, FAISS, Langchain, Google Cloud Platform, Weights & Biases, Git, LaTex, Solidworks

PERSONAL INFORMATION

- Korean (Native Speaker) and English (Fluent)
- Leadership Role: Leader of the Graduate School Tennis Club