Jonghyun (Jong) Song

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

Research interest

- Optimizing query-document interactions in information retrieval (IR) systems
- Representation Learning for retrieval-augmented generation (RAG) and large multi-modal models (LMM)
- Enhancing the multi-modal capabilities of large language models to understand and process document images.
- 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: Jay-Yoon Lee
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. **Comparing Neighbors Together Makes it Easy: Jointly Comparing Multiple Candidates for Efficient and Effective Retrieval**
   In EMNLP Main Track (Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing), 2024 / Spotlight Talk at 9th Workshop on Representation Learning for NLP on ACL 2024
   Jonghyun Song, Cheyon Jin, Wenlong Zhao, Andrew McCallum and Jay-Yoon Lee
2. **Redefining Information Extraction from Visually Rich Documents as Token Classification**
   In IJCAI Competition of Visually Rich Form Document Intelligence and Understanding (VRDIU), 2024
   Jonghyun Song, Eunyi Lyou
3. **Intention Detection Model for Wearable Robots Using sEMG Signal**
   In International Conference on X+Artificial Intelligence (XAICON), Virtual Event, 2021
   Jonghyun Song

Work Experience

Research Assistant (Ph.D. Student) under Professor Jay-Yoon Lee 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 enhance the retrieve-and-rerank pipeline.
  - Utilized shallow self-attention layers to compare query and candidate embeddings jointly, enabling scalable and efficient multiple comparisons.
  - Demonstrated robust performance in entity linking, passage ranking, and dialogue ranking tasks with improved latency and memory efficiency.
  - One paper accepted at **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 control compression based on human intention to prevent work-related musculoskeletal disorders (WMSD) in the wrist. In detail, I focused on:
  - Conducted physical modeling for cable routing to maximize power transmission efficiency
  - Designed a silicone component embedded with bearings and fabric to enhance mobility and portability
  - Developed Arduino-based robotic control systems using force (FSR) sensors

Founder & Software Engineer Sep. 2019 - Jun. 2020

Hakwongo Corp. Seongnam, Korea

- A startup that connects working moms with private education institutes using deep learning technology. (Funded from Seongnam-si and Yonsei University)
  - Developed an NLP model to recommend private education institutes to working moms
  - Built the front end of an Android application using the Flutter framework
  - Pre-processing data on private education institutes using SQL and pandas

Research Internship under Professor Joo-Young Lee Sep. 2019 - Oct. 2019

Seoul National University, Seoul, Korea

- Experiment staff to verify the effects of skin-cooling wearable devices

Awards and Honors

2nd Place, VRDIU Competition (Track A) on IJCAI 2024 sponsored by Google Research Jul. 2024
– Task: predicting the RoIs that can provide correct answer to given questions
– Fine-tuned 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, Nov. 2023
Ministry of Science and ICT, Korea
– Presented *Multi-modal and Multi-view Patent Search System*, a patent search engine with CLIP embeddings of drawings and text
– Utilized self-supervised learning, using 'prior art' section in patents as pseudo-label
– Implemented a chatbot interface with LangChain and Streamlit

– Served as a team leader

Park Min-Chul Data Science Challenge Scholarship, Mar. 2022
Seoul National University, Korea

Cum Laude, Seoul National University, Korea Feb. 2022

Sanhak (Industrial-Educational Cooperation) Foundation Scholarship, Mar. 2021 – Dec. 2021
Korea Sanhak Foundation, Korea

Merit-Based Scholarship, Seoul National University, Korea Dec. 2018 – Dec. 2019

TEACHING EXPERIENCE

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

Technical Skills

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

Personal Information

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