SUNGHO PARK

shpark@dblab.postech.ac.kr, LinkedIn Profile

SUMMARY

I am a Ph.D. student at Data Systems Lab @ Pohang University of Science and Technology (POSTECH), under the supervision of Professor Wook-Shin Han. My research interests range from neural information retrieval to multi-modal open-domain question answering (such as open question answering over tables and text), and to Agentic AI. Recently, I have focused on designing algorithms that retrieve and integrate tables and passages relevant to a query by leveraging semantic relationships across modalities.

EDUCATION

Ph.D.	Feb. 2023 - Present	POSTECH, South Korea
		Graduate School of Artificial Intelligence
		(Supervisor: Professor Wook-Shin Han)
B.S.	Feb. 2019 - Feb. 2023	POSTECH, South Korea
		Bachelor of Electrical Engineering
		Summa Cum Laude (GPA 3.98/4.3; Major GPA 4.10/4.3)
HSD.	Feb. 2017 - Feb. 2019	Chungnam Science High School, South Korea
		Early graduation

PUBLICATIONS

1. **Park, S.,** Kim, J., and Han, W., "SPARTA: Scalable and Principled Benchmark of Tree-Structured Multi-hop QA over Text and Tables" *(submitted)*
Key Idea: SPARTA is a fully automated SQL-centric pipeline that constructs a tree-structured multi-hop QA benchmark by unifying structured and unstructured evidence from tables and text into a single relational representation.

2. Lee, S., **Park, S.**, and Han, W., "SAFE: Schema-Driven Approximate Distance Join for Efficient Knowledge Graph Querying" *(EMNLP 2025 Main)*
Key Idea: SAFE introduces a schema-driven approximate distance join algorithm that refines noisy LLM-generated query graphs using schema-level constraints and efficiently aligns them with large knowledge graphs, enabling robust and scalable knowledge graph querying.

3. **Park, S.,** Yun, J., Lee, J., and Han, W., "HELIOS: Harmonizing Early Fusion, Late Fusion, and LLM Reasoning for Multi-Granular Table-Text Retrieval" *(ACL 2025 Main)*
Key Idea: HELIOS formulates retrieval as finding a query-relevant subgraph within a bipartite data graph built via early fusion of table segments and passages, and introduces a three-stage pipeline integrating early fusion, late fusion, and LLM reasoning.

4. **Park, S.,** Seok, J., Lee, J., Yun, J., and Lee, W., "KDD Cup Meta CRAG 2024 Technical Report: Three-step Question-Answering Framework" *(2024 KDD Cup Workshop for Retrieval Augmented Generation)*
Key Idea: A three-step RAG framework that minimizes unnecessary retrievals by leveraging LLMs' inherent knowledge and introduces a verification stage to prevent error propagation, improving both accuracy and efficiency in question answering.

EXPERIENCE

Internship @ Oracle Labs Jul. 2023 - Oct. 2023

Developed an NL2SQL generation model for Oracle Database.

Internship @ Samsung Electronics' Mobile Experience Business Jun. 2022 - Sep. 2022

Developed an image-processing algorithm

Internship @ AMI Laboratory Jun. 2021 - Jun. 2022

Developed an end-to-end method for reconstructing an animatable 3D human mesh from a single cropped 2D image of a person.

Internship @ Sellerhub Jun. 2020 - Sep. 2020

Served as a product manager at Sellerhub, a startup that provides a platform for managing online shopping mall integrations.

PROJECTS

Table-Text Retrieval Jan. 2024 - Feb. 2025

· Figured out the strengths and weaknesses of both early and late fusion approaches, and developed a graph-based retrieval method that integrates the advantages of both techniques for table-text retrieval.
· Observed that traditional retrieval methods, which rely solely on semantic similarity, struggle with tasks requiring advanced logical inference (such as column-wise aggregation or multi-hop reasoning). Enhanced the retrieval process by incorporating LLM reasoning.
· Analyzed and determined the appropriate granularity for each retrieval stage, optimizing the overall performance. The optimized retrieval method outperforms existing approaches.
· Accepted at ACL 2025 Main

Retrieval-augmented Generation for Open Question Answering Mar. 2024 - Jun. 2024

· Identified issues with the baseline RAG, including unnecessary retrievals and error propagation from incorrect retrieval results.
· Added two steps to the baseline RAG to address these problems: utilizing latent knowledge within large language models to reduce unnecessary retrievals, and incorporating a verification phase to ensure the factual accuracy of the generated responses.
· Won prizes for the Comparison question in KDD Cup 2024 Meta CRAG Challenge Tasks 1, 2, and 3, as well as for the post-processing question in Task 1.

Development of NL2SQL Generation Model for Oracle Database Jul. 2023 - Oct. 2023

· Identified that representative NL2SQL benchmarks like Spider follow the SQLite dialect, and more than half of the queries cannot be executed on Oracle databases.
· Built an Oracle-dialect-specific dataset by translating SQLite queries into the Oracle dialect using SQLGlot. SQL queries with SQLite-specific features not supported in Oracle, were excluded from the dev and training sets.
· Improved execution accuracy by 27.4% over the baseline model through various fine-tuning techniques, including hyperparameter tuning, prompt engineering (zero-shot or few-shot), parameter-efficient fine-tuning, and dataset augmentation.

Detecting Anomalies in DB Log Data Oct. 2023

· Implemented a model for detecting anomalies across multiple DB log datasets.
· Improved the generalization performance of the anomaly transformer (ICLR 2022) by combining scores calculated using Mahalanobis distance, based on multivariate time series statistics, with the scores from the anomaly transformer.

Uncertainty-based Clarification Request for Ambiguous NL Queries Nov. 2022 - Dec. 2022

· Developed a system to assess the reliability of generated SQL queries based on their confidence scores, determining whether the generated SQL query is trustworthy. If the confidence score is low, the system requests clarification from the user to provide a more precise natural language query.
· Implemented a technique using Integrated Gradients to identify which words in the natural language query contribute to higher uncertainty, highlighting them to guide the user in refining the query.

Constructing Animatable 3D Human Mesh from Cropped 2D Image   Aug. 2022 - Dec. 2022

· Implemented an end-to-end framework for reconstructing animatable 3D human meshes from cropped 2D images.
· The framework consists of two stages: synthesizing full-body images from cropped 2D images, and constructing animatable 3D human meshes from the synthesized images.
· Verified that the accuracy of the 3D mesh construction varies depending on the pose of the synthesized full-body image and optimized the pose to maximize the reconstruction accuracy.
· Awarded the Grand Prize in the Newcomb Lim Ki-Hong Design Challenge.

HONORS AND AWARDS

KDD Cup Meta CRAG 2024: Head member of team dRAGonRAnGers   Jul. 2024

· First place in comparison question in Task 1, 2 and 3
· First place in post-processing question in Task 1

Newcomb Lim Ki-Hong Design Challenge Grand Prize   Feb. 2023

Awarded to students of POSTECH EE who presented outstanding research results in a design challenge

Millitech Military Academy Commander's Award   Feb. 2022

Awarded to a ROND(Benchmarking of Israel's Talpiot)'s novice who showed outstanding research results

National Science and Technology Scholarship   Aug. 2021

Scholarship awarded to students with outstanding academic ability in the field of science and engineering

TEACHING EXPERIENCE

Teaching Assistant: AI Application Specialist   Aug. 2024 - Aug. 2025
Samsung Electronics Co.

Delivered sessions on language model applications and Retrieval-Augmented Generation (RAG) to software developers.

Teaching Assistant: Level 4 Data Science Expert Program   Jul. 2024
Samsung Electronics Co.

Led hands-on sessions on Retrieval-Augmented Generation (RAG) for data science professionals, with 90% of participants holding Ph.D. degrees.

Teaching Assistant: Research Project   Jun. 2024
POSTECH

Provided mentorship to students Juen Kim and Sohyeon Kim on topics related to Retrieval-Augmented Generation. Juen Kim received the Popularity Award for their project.

Teaching Assistant: Programming & Problem Solving   Feb. 2024
POSTECH

Assisted with course instruction and student support in programming and problem-solving techniques.

SKILLS

Programming Languages	Python, C, SQL
Python Packages	PyTorch
Tools	PostgreSQL, Docker

AREA OF INTEREST

Retrieval-augmented Generation, Open-domain QA, Multi-modal QA, Information Retrieval, Text2SQL

REFERENCES

Wook-Shin Han   wshan@dblab.postech.ac.kr

Full Professor, Pohang University of Science and Technology (POSTECH), Pohang, South Korea

Rhicheek Patra   rhicheek.patra@oracle.com

Research Director, Oracle Labs, Zurich, Switzerland