

MINSANG KIM

+82 10 4026 4271 ◇ kmswin7@gmail.com

<https://kmswin1.github.io/>

RESEARCH INTERESTS

NLP, LLM, Data-Centric AI and Foundation Models

EDUCATION

Korea University, Seoul, South Korea

2021 - 2023

M.S (Master of Science in Engineering)

Department of Computer Science and Engineering

System Intelligence Group (SING)

Research area : Representation Learning for structured data (graphs) and unstructured data (text)

Advisor : professor Seungjun Baek

Korea University, Seoul, South Korea

2013 - 2019

B.S (Bachelor of Science in Engineering)

Department of Computer Science and Engineering

Data mining and visual analytics Lab (DAVIAN) Internship

Research area : Natural Language Processing, Machine Learning

Advisor : professor Jaegul Choo

PUBLICATIONS

- Not All Data are Created Equal: A Compression View to Data Pruning for Training Large Language Models - **Minsang Kim**, Seungjun Baek
Under Review
- QPaug: Question and Passage Augmentation for Open-Domain Question Answering of LLMs - **Minsang Kim**, Cheoneum Park, Seungjun Baek
Under Review
- Exploration of Large Language Models on Cross-Cultural Values - **Minsang Kim**, Seungjun Baek
Under Review
- Hierarchical Position Embedding of Graphs with Landmarks and Clustering for Link Prediction - **Minsang Kim**, Seungjun Baek
The International World Wide Web Conference (WWW), 2024, *Accepted* (20.2% acceptance rate)
PDF
- Persona-Knowledge Dialogue Multi-Context Retrieval and Enhanced Decoding Methods - Minsik Oh*, **Minsang Kim*** (*: equal contributions)
COLING workshop on Customized Chat Grounding Persona and Knowledge (COLING-PKG), 2022, *Oral, Accepted* **PDF**
Workshop challenge 1st place award (Both retrieval & generation tasks)
- OASYS: Domain-Agnostic Automated System for Constructing Knowledge Base from Unstructured Text - **Minsang Kim***, Sang-Hyun Je*, Eunjoo Park (*: equal contributions)
ACM SIGKDD Workshop on Mining and Learning with Graphs (KDD-MLG), 2022, *Poster, Accepted* **PDF**

- ComDensE : Combined Dense Embedding of Relation-aware and Common Features for Knowledge Graph Completion - **Minsang Kim**, Seungjun Baek
International Conference on Pattern Recognition (ICPR), 2022, *Poster, Accepted PDF*

AWARDS

- SK Telecom and Broadband AI Hackathon challenge, 2023
SKT-B AI Challenge - 1st place award in customer experience innovation part.
- COLING workshop challenges on Customized Chat Grounding Persona and Knowledge, 2022
COLING-PKG - All shared tasks 1st place award (1.Retrieval & 2.Generation).

WORK EXPERIENCE

SK Telecom, South Korea

January 2023 - current

Machine Learning Research Engineer

- Research area : NLP, Graphs, and Foundation Models.

Kakao Enterprise AI Lab, South Korea

December 2019 - December 2022

Machine Learning Research Engineer

- Research area : NLP, Graphs, and Foundation Models.

Fast Campus, South Korea

September 2022 - November 2022

Lecturer

- I lectured on several NLP categories - Machine translation, Sentence classification, Passage retrieval, Dialogue system, Knowledge graph, and Information extraction. **Lecture**

Kakao Corp. AI Lab, South Korea

July 2019 - December 2019

Machine Learning Research Engineer

- Research area : NLP, Graphs, and Foundation Models.

PROJECTS

2023.01 - current : SK Telecom

In particular, I am modeling and developing Large-scale pre-trained language and multimodal models. e.g. pre-train, instruction fine-tuning, etc.
Modeling *Foundation Models*.

2021.10 - 2022.12 : Kakao Enterprise

I had developed Open-Domain QA (Passage Retrieval, MultiModal Retrieval), NLU tasks (Assistant, Bot classifier, Intent Classifier, Machine Learning Entity Tagger) for Kakao-i assistant system, and cross-document language modeling. (estimate semantic network and post-train LMs.)

2021.10 - 2022.12 : Kakao Enterprise

I had developed graph-based models such as link prediction, and community detection for network analysis project.

2021.07 - 2021.12 : Kakao Enterprise

I had developed Sequence to sequence model for triple (knowledge) to text generation task.

2020.08 - 2021.12 : Kakao Enterprise

I had developed a knowledge graph auto-construction system without labeled data where our company requires some automatic training service system for B2B business. We also opened Korean text to

triple benchmark dataset for contributing KG construction research environment.

2020.05 - 2021.02 : Kakao Enterprise

I developed personalized member ranking system for KakaoWork that B2B messenger service in our company.

2019.07 - 2020.3 : Kakao Corp.

I had developed Daum search engine that is one of the most famous search engine in South Korea contents search ranking system.

LEADERSHIP EXPERIENCES

2023.10 - 2023.12 : SK Telecom

SK Telecom and Broadband AI Challenge.

As a team leader, I participated in AI hackathon with an Android developer and service planner. Finally, My team achieved **1st award** of the AI Idea Hackathon Challenge - Our team developed an on-device voice-pishing classifier and chatbot based on Telco-LLM.

2023.06 - 2023.12 : SK Telecom

Mentor of SKT AI Fellowship.

I advised mentees of SKT AI Fellowship -topic extraction project based on korean GPT-3 and lead this project.

2019.07 - 2022.12 : Kakao Corp & Kakao Enterprise

Lead several projects and team research seminar.

I had led several service projects in my companies. First, I had developed Daum search engine that is oone of the most famous search engines in South Korea. Second, I developed a personalized member ranking system for KakaoWork that B2B messenger service in our company. In addition, I had developed a knowledge graph auto-construction system without labeled data where our company requires some automatic training service system for B2B business. Third, I have developed graph-based models such as link prediction, and community detection for network analysis project (such as how workers in specific company collaborate with, who would collaborate in the future etc.). Finally, I have developed Open-Domain QA for Kakao-i system, and cross-document language modeling (estimate semantic network and post-train LMs).

I also led and planned a team research seminar for reviewing AI conferences. Every member read the paper and prepared for the presentation. During the seminar, each member presented their papers and discussed these papers.

2017.07 - 2017.08 : Draper University, San Mateo & Industrial Technology research institute, Berlin

KIC college R2BD (Research to Business Development) program : Complete Entrepreneurship program.

In the summer of 2017, I attended to KIC college R2BD program, which was participated by several universities in South Korea, and I was on behalf of our school for cultivating entrepreneurship. I and My colleague introduced a pill scanner that classifies what kind of pills by using a deep learning model and then made a business plan for this project. During this program, I learned about teamwork, design-thinking, presentation skills, etc. from various activities. Furthermore, I had progressed our business plan combined with methods that were taught from various activities. Finally, I presented my project result in front of a lot of international students and Tim Draper who is a famous venture capital investor in Silicon Valley.

TECHNICAL STRENGTHS

Research Area	NLP, Graphs, Foundation Models
Software, Tools	python(advanced), pytorch(advanced), L ^A T _E X(advanced), slurm(advanced)

LANGUAGE PROFICIENCY

- Korean (Native proficiency)
- English (Advanced proficiency)