Youngjune Lee

ML Engineer @ Naver

yjlee511@gmail.com

Education

Korea Advanced Institute of Science and Technology (KAIST) *Seoul, S.Korea*

M.S., Software graduate program *Mar. 2020 - Feb. 2022*

- AIPR Lab (Advised by Prof. Kee-Eung Kim)
- Main Research field: Recommender System, Information Retrieval

Hanyang University *Seoul, S.Korea*

B.S. in Mathematics *Mar. 2012 - Feb. 2019*

Work Experience

Naver *Bundang-gu, S.Korea*

Machine Learning Engineer - Ranking & Recommendation *Jan. 2022 - present*

- Naver is the South Korea's leading search platform. (MAU: about 40 million)

- **Naver Search** - **recommendation** (2022.01 - 2023.12)
  - Develop ranking features and model architecture that can simultaneously satisfy relevance and personalization in search engines.
  - Contributed to the application of personalized, categorized document blocks to facilitate navigation of search results.

- **Feature set optimization** (2023.01 - 2023.12)
  - By calculating influence of each feature, optimize feature set for fast inference, performance and system level efficiency.
  - Feature selection method is published at CIKM 2023 (First author)

- **Naver Short-form** - **recommendation** (2023.06 - 2023.12)
  - Develop ranking features and ranking model for short-form recommendation

- **Naver Search - ranking** (2024.01 - 2024.12)
  - Develop LLM distillation method to develop sLM re-ranker to improve long-tail queries ranking quality
  - Improved long-tail queries search quality significantly
  - Long-tail queries re-ranking method is accepted at EMNLP 2024 (Co-first author)
  - The overall method is presented at DAN 24 (Naver conference)

- **Naver Search - ad recommendation** (2024.04 -)
  - Develop retriever to improve query-advertisement relevance quality in search engine
  - Develop multi-modal retriever for item-to-item and query-to-item recommendation

- **Naver Search - feed recommendation** (2024.05 - 2025.04)
  - Develop global & personalized retriever to improve query-document relevance quality in feed recommendation
  - Personalized retrieval method is accepted at SIGIR 2025 (Co-first author)

- **Naver Search - query filtering (safe model)** (2025.01 - 2025.04)
  - Develop light-weight unsafe-query filtering model for LLM summarization

- Construct and maintain data pipeline for serving and training

TmaxData *Bundang-gu, S.Korea*

R&D Researcher *Feb. 2019 - Feb. 2020*

- Design and develop the ERP System
  Design Personnel and Financial System roadmaps
  Develop the logic of Personal information inquiry and Income Statement modules

Publications

SIGIR 2025 - Industry Track

Youngjune Lee*, Haeyu Jeong*, Changgeon Lim, Jeong Choi, Hongjun Lim, Hangon Kim, Jiyoon Kwon, Saehun Kim. IRA: Adaptive Interest-aware Representation and Alignment for Personalized Multi-interest Retrieval

- Personalization with multi-vector retrieval

EMNLP 2024 - Industry Track
Nayoung Choi*, **Youngjune Lee***, Gyu-Hwung Cho, Haeyu Jeong, Jungmin Kong, Saehun Kim, Keunchan Park, Sarah Cho, Inchang Jeong, Gyohee Nam, Sunghoon Han, Wonil Yang and Jaeho Choi. RRADistill: Distilling LLMs' Passage Ranking Ability for Long-Tail Queries Document Re-Ranking on a Search Engine
- Re-ranker for long-tail queries ranking

CIKM 2023 - Short
Youngjune Lee, Yeongjong Jeong, Keunchan Park and SeongKu Kang . MvFS: Multi-view Feature Selection for Recommender System
- Selecting key features from each data instance in Recommender System

NAACL 2022 - Findings
Haeju Lee*, Oh Joon Kwon*, Yunseon Choi*, Minho Park, Ran Han, Yoonhyung Kim, Jinhyeon Kim, **Youngjune Lee**, Haebin Shin, Kangwook Lee, and Kee-Eung Kim Learning to Embed Multi-Modal Contexts for Situated Conversational Agents.
- Improving transformer's joint-training architecture for Multi-Modal Dialogue system using VR-based shopping dataset

DSTC10 Workshop at AAAI 2022
Haeju Lee*, Oh Joon Kwon*, Yunseon Choi*, Jinhyeon Kim, **Youngjune Lee**, Ran Han, Yoonhyung Kim, Minho Park, Kangwook Lee, Haebin Shin and Kee-Eung Kim. Tackling Situated Multi-Modal Task-Oriented Dialogs with a Single Transformer Model.
- Improving transformer's joint-training architecture for Multi-Modal Dialogue system using VR-based shopping dataset

CIKM 2021 - Short
Youngjune Lee and Kee-Eung Kim. Dual Correction Strategy for Ranking Distillation in Top-N Recommender System.
- Distillation focused on the parts that are particularly difficult to follow using discrepancy on both the user side and the item side of the teacher model and the student model.

Data Centric AI Workshop at NeurIPS 2021
Youngjune Lee, Oh Joon Kwon, Haeju Lee, Joonyoung Kim, Kangwook Lee and Kee-Eung Kim. Augment & Valuate : A Data Enhancement Pipeline for data-centric AI Learning.
- Development of a data quality improvement pipeline, that utilizes augmentation and hard negative dataset extraction and modification, to improve image classification performance with only a small dataset.

Competitions

Data Centric AI Competition

Andrew Ng, DeepLearning.AI and Landing AI — *Jul. 2021 - Sep. 2021*

- 6th place (6/489) & Honorable Mention in Innovative
- About competition: https://https-deeplearning-ai.github.io/data-centric-comp/
- Blog post of winners: https://www.deeplearning.ai/data-centric-ai-competition-kaist-aiprlab/
- Task: **Improve model performance by only improving the dataset (with fixed hyperparameters)**
- Tech stacks: Vision model, Auto Augmentation, hypergradient based influence function, Contrastive Learning

The Tenth Dialog System Technology Challenge (DSTC 10)

Facebook Research — *June. 2021 - Oct. 2021*

- Winner or runner-up in each sub-tasks
- About competition: https://github.com/facebookresearch/simmc2
- Sub-tasks: **1. Disambiguation classification 2. Multimodal Co-ref Resolution 3. Dialog State Tracking, 4. Response Generation/Retrieval**
- Tech stacks: Language Model, Dialog System, Vision-Linguistic Model

Honors & Awards

2025	**3rd place in the 2024 In-house Innovation R&D Awards for developing long-tail query re-ranking**, Naver
2021	**Winner or runner-up in each sub-tasks**, DSTC10 (Track3-Facebook Research)
2021	**Honorable Mention and rank #6 (6/489)**, Data Centric AI Competition (Andrew Ng., DeepLearning.AI, Landing.AI)
2021	**SIGIR Student Travel Award**, ACM SIGIR
2020	**LG Electronics industrial Scholarship**, LG Electronics
2018	**Bonsol Kim Jong-han Scholarship**, Bonsol Kim Jong-han Scholarship Foundation
2012	**National Science & Technology Scholarship**, Korea Student Aid Foundation

Additional Activities

W&B Seoul Meetup

Weights & Biases — *January. 2025*

- Representation of long-tail queries re-ranking in NAVER Search Engine
- Related post(Korean) hyperlink: Meetup post

Naver conference (DAN 24)

NAVER *November. 2024*

- Naver conference is the biggest IT conference in Korea
- Introduction about LLM distillation for long-tail queries re-ranking in NAVER Search engine
- Conference page hyperlink: DAN 24

Technical interview in a Korean news article

NAVER *April. 2024*

- Article about LLM distillation for long-tail queries re-ranking in NAVER Search engine
- Related post(Korean) hyperlink: News Article

Tech talk in the Korean data community

Superb AI *Jan. 2022*

- Issued by **Superb AI**
- Talk about progress and idea for the Data-Centric AI competition
- Related post(Korean) hyperlink: Data Round Table