DAEYOUNG KIM

Google Scholar
github.com/cyc1am3n
daeyoung.k03@gmail.com

RESEARCH INTERESTS

Multimodal AI, Reliable Machine Learning, Natural Language Processing

EDUCATION

Korea Advanced Institute of Science and Technology *March 2021 - Feb 2023*
Master's Degree

- Master of Artificial Intelligence (Kim Jaechul Graduate School of AI)
- Advisor: Prof. Edward Choi
- GPA 3.98 / 4.3

Korea University *Mar 2015 - Feb 2021*
Bachelor's Degree

- Bachelor of Science in Computer Science and Engineering
- Bachelor of Integrated Major in Information Security Convergence
- GPA 3.81 / 4.5

PUBLICATIONS

OffsetBias: Leveraging Debiased Data for Tuning Evaluators
Junsoo Park*, Seungyeon Jwa*, Meiying Ren, **Daeyoung Kim**, and Sanghyuk Choi
In Findings in Empirical Methods in Natural Language Processing (**EMNLP**), 2024 [pdf] [code] [dataset] [model]

Towards the Practical Utility of Federated Learning in the Medical Domain
Seongjun Yang*, Hyeonji Hwang*, **Daeyoung Kim**, Radhika Dua, Jong-Yeup Kim, Eunho Yang, and Edward Choi
In Proc. of Conference on Health, Inference, and Learning (**CHIL**), 2023 [pdf] [code]

Revisiting the Importance of Amplifying Bias for Debiasing
Jungsoo Lee*, Jeonghoon Park*, **Daeyoung Kim***, Juyoung Lee, Edward Choi, and Jaegul Choo
In Proc. of Association for the Advancement of Artificial Intelligence (**AAAI**), 2023 (**Oral Presentation**) [pdf] [code]

Uncertainty-Aware Text-to-Program for Question Answering on Structured Electronic Health Records
Daeyoung Kim, Seongsu Bae, Seungho Kim, and Edward Choi
In Proc. of Conference on Health, Inference, and Learning (**CHIL**), 2022 [pdf] [code]

Question Answering for Complex Electronic Health Records Database using Unified Encoder-Decoder Architecture
Seongsu Bae, **Daeyoung Kim**, Jiho Kim, and Edward Choi
In Proc. of Machine Learning for Health (**ML4H**), 2021 (**Oral Presentation**) [pdf]

VARCO-VISION: Expanding Frontiers in Korean Vision-Language Models
Jeongho Ju*, **Daeyoung Kim***, SunYoung Park*, and Youngjune Kim
Technical Report, 2024 [pdf] [model]

Empowering Sentence Encoders with Prompting and Label Retrieval for Zero-shot Text Classification
Jimin Hong*, Jungsoo Park*, **Daeyoung Kim***, Seongjae Choi, Bokyung Son, and Jaewook Kang
Preprint, 2022 [pdf]

RESEARCH EXPERIENCES

Edlab, Korea Advanced Institute of Science and Technology *Seongnam, South Korea*
Master's Degree Researcher *March 2021 - Feb 2023*

- Led the development of an EHR question-answering system, progressing from a foundational encoder-decoder model to an advanced uncertainty-aware text-to-program model that improved reliability
- Proposed and validated a bias-amplification-and-removal approach to address the limitations of conventional debiasing methods
- Investigated the challenges and validated the practical utility of applying theoretical federated learning to real-world medical environments for secure data handling

Data Mining and Information Systems Lab, Korea University *Seoul, South Korea*
Undergraduate Researcher *Mar 2020 - Jul 2020*

- Researched and implemented Zero-shot Summarization techniques for COVID-19 academic papers

WORK EXPERIENCES

KT (Korea Telecom) *Seoul, South Korea*
Research Engineer *Dec 2024 - Present*

- Constructed training datasets for a Vision Language Model (VLM), focusing on synthetic document and instruction tuning
- Led the development of an omni-modal model (image, text, and speech), from core architecture design to training

NCSOFT *Seongnam, South Korea*
Research Engineer *Feb 2023 - Dec 2024*

- Managed the full training pipeline for VARCO-VISION, an image-text-to-text (IT2T) model, including pre-training, instruction tuning, and alignment
- Developed VARCO-Text, an instruction-tuned model for an LLM-based writing assistant, and constructed datasets for text continuation and summarization tasks
- Executed alignment tuning for VARCO LLM 2.0 with custom-built, Korean-specific datasets
- Investigated LLM-as-an-Evaluator biases by designing and building a novel benchmark and dataset for bias identification and mitigation

NAVER Corporation *Seongnam, South Korea*
Research Intern *Jul 2022 - Jan 2023*

- Developed and scaled an LLM-based sentence embedding model (137M to 7B parameters) using Contrastive Learning
  - Optimized the large-scale model training process through multi-GPU and multi-node configurations
- Developed a Zero-shot Text Classification method leveraging sentence encoders and retrieval-augmented prompts

NAVER Corportation *Seongnam, South Korea*
Research Intern *Jul 2020 - Aug 2020*

- Developed a sentiment and intent classification model for online comments, enhancing its performance with custom Korean data augmentation techniques

HONORS AND AWARDS

Graduation Project Competition at Korea University **3rd Place**
Text Summarization of Biomedical paper related to COVID-19 [demo] *Jun 2020*

- Employed Zero-shot Summarization techniques to generate summaries from biomedical papers on COVID-19

HeLP Challenge 2019 at Asan Medical Center **1st Place**
Breast Cancer Classification on Frozen Pathology [code] *Apr 2020*

- Developed a two-stage model to predict cancer metastasis and measure tumor length from frozen pathology slides

Kakao Arena **3rd Place**
Article Recommendation Task [code] *Feb 2019*

- Developed an article recommendation system for the 'Brunch' blogging platform

ACADEMIC SERVICE

- **Workshop Reviewer**: Learning from Time Series for Health@NeurIPS (2025)

LANGUAGE PROFICIENCY

Native **Korean**; Proficient in **English**

- iBT TOEFL: 106 (Reading: 30, Listening: 27, Speaking: 25, Writing: 24)