SangYeon Cho

+82 10 4618 8508 whtkddus98@gmail.com Website Github

Master of Science in Engineering @ Department of AI, Chung-Ang University, Seoul 06964, Republic of Korea

Welcome!

"Let's do my best rather than be the best."

Hello, I'm Cho Sangyeon. I am currently a master's degrees at Chung-Ang University in Korea. I am interested in the fields of machine learning and deep learning engineers.

Research Interests

- Deep Learning
- Natural Language Processing
- Knowledge Distillation
- Multi-modal Representation Learning

Education

MS **Chung-Ang Univ**, AI Engineering 2023.03 – 2025.02

- GPA: 4.34/4.5 (Transcripts)
- Coursework: Multi-modal Learning, Knowledge Distillation
- Advisor: Prof. Jungyeong Kim(Labs)
- Graduation Thesis

BS **Kookmin Univ**, Software Engineering 2017.03 – 2023.02

- GPA: 4.09/4.5 (Transcripts)
- Coursework: Algorithms, Data Structure, Artificial Intelligence
- Graduation Work

Publications

Improving Multimodal Data Quality with Unified Filtering Score (UF-Score) 2024.12

Sangyeon Cho, Mingi Kim, Hwang JinKwon, Jaehoon Go, Junyeong Kim*

Association for the Advancement of Artificial Intelligence (AAAI) GoodData, 2025

Multi2Cap: Improving Automated Audio Captioning with Cross-modal Feature Distillation and Large Language Model 2024.11 *Under Review*

Sangyeon Cho, Jangyeong Jeon, Mingi Kim, Jaeho Han, Junyeong Kim*

Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL), 2025

Synergy-CLIP: Extending CLIP with Multi-modal Integration for Robust Representation Learning 2024.09 *Under Review*

Sangyeon Cho, Jangyeong Jeon, Mingi Kim, Junyeong Kim*

IEEE ACCESS, 2024

BioBridge: Unified Bio-Embedding with Bridging Modality in Code-Switched EMR 2024.08

Jangyeong Jeon, ***Sangyeon Cho***, Dongjoon Lee, Changhee Lee, Junyeong Kim*

IEEE ACCESS, Volumn 12, 2024

DSG-KD: Knowledge Distillation from Domain-Specific to General Language Models 2024.08
Sangyeon Cho, Jangyeong Jeon, Dongjoon Lee, Changhee Lee, Junyeong Kim*
IEEE ACCESS, Volumn 12, 2024

ConCSE: Unified Contrastive Learning and Augmentation for Code-Switched Embeddings 2024.07
Jangyeong Jeon, ***Sangyeon Cho***, Minuk Ma, Junyeong Kim*
IEEE International Conference on Pattern Recognition (ICPR)

A pediatric emergency prediction model using natural language process in the pediatric emergency department 2024.06
Under Review
Arum Choi, Chohee Kim, Jisu Ryoo, Jangyeong Jeon ,***Sangyeon Cho***, Dongjoon Lee, Junyeong Kim, Changhee Lee*, Woori Bae*
Nature Scientific Reports, 2024

Experience

Lablup Inc., AI Research Assistants — Seoul, Korea
2022.06 – 2022.09

- Contribute to the research team
- Developed ground truth data pre-processing module
- Predictive modeling of respiratory infectious disease transmission risk based on ground truth data of droplet spread
- Data augmentation modeling using GAN
- Seminar LINK

Projects

119 Intelligent report reception voice recognition data construction 2023.09 - 2024.09

- building data pre-processing and analytics pipeline for speech recognition
- modeling dispatch decision-making for decision aids for guard officers
- Tools Used: Pyhon, SQL

Optical Character Recognition (OCR) model improvements 2023.04 - 2024.09

- Designed a model to classify emergency and critical patients using emergency department EMR data.
- Pre-trained the language model using clinical notes collected from real hospitals.
- we additionally utilize clinical note (text) information to utilize the most potential information.
- By combining tabular encoder and text encoder at the embedding level, we achieved better performance than existing models.

Optical Character Recognition (OCR) model improvements 2022.09 - 2023.02

- AI Competition Organized by Software Center University
- Using Pororo as a baseline, we improved it by utilizing data augmentation and adding the Mix-Style module.
- As a result, the team ranked at the top of the leaderboard and won an honorable mention.
- Presentation LINK

AI speakers for seniors living alone 2022.03 - 2022.06

- Designing an AI speaker that analyzes the emotions of seniors living alone and alerts their caregivers
- Responsible for designing/training deep learning models
- We used two models: Sentiment Classification Model and Sentiment Conversation Model.
- Github LINK

Research on Meta-Learning Improvement 2022.02 - 2022.12

- Studying/implementing meta-learning in general
- Considered how to improve Model-Agnostic Meta-Learning (MAML) under the guidance of a professor
- Conducted research on applying MAML to natural language processing by combining it with pruning and knowledge distillation
- Github LINK

Research on Improving Information Security Behavior 2020.10 - 2022.04

- Developing a framework to guide better information security compliance behavior in the enterprise
- Teach principal investigators programming fundamentals
- Teach principal investigators on machine learning and deep learning concepts
- Github LINK

Awards & Honors

Awards

IEIE Summer Conference, Encouragement Award	2024.06
SW-focused University Collaborative AI Competition, Encouragement Award	2022.09

Honors

Merit Scholarships	2022.09
Merit Scholarships	2022.03
S&T Foundation Scholarships	2022.02

Technologies

Languages: Python, C/C++, C#, Java, SQL, Scala

Credentials: ADsP