

JAEYOON JUNG

🏠 <https://lastdefiance20.github.io/> ✉ lastdefiance20@gmail.com 📄 [jaeyoon-jung-935300268](https://jaeyoon-jung-935300268.github.io/)

Summary

Hi, I'm an undergraduate student with a broad interest in artificial intelligence. My current research focuses on several key areas: multimodal models capable of handling diverse inputs and outputs (including video, image, text, audio, and action), fact-checking using large language and vision-language models, virtual try-on systems, and adapting instruction-following capabilities of English-based large language models to Korean and other languages.

Education

Soongsil University

Mar. 2020 – Feb 2026

Bachelor of Science in Artificial Intelligence

Seoul, South Korea

- includes a 2-year mandatory alternative military service at MAUM.AI
- GPA 4.19 / 4.5, currently 2nd/100

Research Experience

MAUM.AI

Dec 2024 – Current

Senior AI Research Scientist

Pangyo, South Korea

- Developing a Korean Vision-Language Model (VLM) and building surveillance VLMs for real-time video anomaly detection on edge devices (e.g., Jetson NX, Hailo-10H).
- Working on WoRV team [LINK] to build a End-to-End Vision-Language Action model to drive or act like human. Collaborating with SNUPI Lab, led by Prof. Youngjae Yu.

AI Research Scientist

Apr 2023 – Nov 2024

- Improved the quality of RAG by enhancing the retriever. By training it with a high-quality dataset and selecting the best models, the retriever achieved a maximum 68% score improvement(Recall@K) compared to its previous performance.
- Optimized STF(Speech-To-Face) model by using ONNX and TensorRT, reducing the 1-second video generation time from 1.6 seconds to 0.7 seconds. This enhancement enables real-time streaming in real-world applications.
- Worked on transferring instruction-following capabilities from English to Korean in open-source large language models. This work aims to facilitate the development of high-quality Korean language models at a low cost.
- Lead the creation of a Korean VLM and a comprehensive benchmarking framework for evaluating Korean VLMs. We open-source the dataset and model, in addition to developing a proprietary version for on-premises deployment.
- Contributed to the development of vision-language action models under team WoRV, focused on enabling commonsense reasoning-based VLM navigation using only visual perception.
- Experienced with training large language models(up to 70B), using deepspeed for multi-node training. While each node comprises 8 NVIDIA H100 80GB GPUs, eight DGX H100 systems interconnected with NVSwitch are used for training.

HUMANE Lab

June 2024 – Current

Undergraduate research student advised by Kunwoo Park

Soongsil University, South Korea

- Work for building automated fact-checking for both using only text and multi-modal

Replica AI

March 2025 – April 2025

ML Engineer, independent contractor

Remote, United State

- Work for virtual-try-on(VITON) fashiontech

MISys Lab

June 2021 – August 2022

Undergraduate research student advised by Kanghee Kim

Soongsil University, South Korea

- Studying C++ programming, developing motor control systems using EtherCAT, and building an electric go-kart.

Publications

CostNav: A Navigation Benchmark for Cost-Aware Evaluation of Embodied Agents

Nth author

Haebin Seong, Sungmin Kim, Minchan Kim, Yongjun Cho, Myunchul Joe, Suhwan Choi, Jaeyoon Jung, Jiyong Youn, Yoonshik Kim, Samwoo Seong, Yubeen Park, Youngjae Yu, Yunsung Lee

Under Review

Is a Picture Worth Thousands of Words? Adaptive Agentic Multimodal Fact-Checking with Visual Evidence Necessity

1st author

Jaeyoon Jung, Yejun Yoon and Kunwoo Park

Workshop on Human-Centric AI at CIKM, 2025

Under Review

D2E: Scaling Vision-Action Pretraining on Desktop Data for Transfer to Embodied AI Suhwan Choi*, Jaeyoon Jung* , Haebin Seong*, Minchan Kim, Minyeong Kim, Yongjun Cho, Yoonshik Kim, Park Yu Been, Youngjae Yu, and Yunsung Lee <i>Under Review</i>	<i>Co-1st author</i>
Exploring Fine-Tuning of Large Audio Language Models for Spoken Language Understanding under Limited Speech data Youngwon Choi, Jaeyoon Jung , Hyeonyu Kim, Huu-Kim Nguyen, Hwayeon Kim <i>Under Review</i>	<i>2nd author</i>
Team HUMANE at AVeriTeC 2025: HerO 2 for Efficient Fact Verification Yejun Yoon, Jaeyoon Jung , Seunghyun Yoon and Kunwoo Park <i>Eighth Workshop on Fact Extraction and VERification (FEVER) at ACL, 2025</i>	<i>2nd author</i>
Hypothetical Documents or Knowledge Leakage? Rethinking LLM-based Query Expansion Yejun Yoon, Jaeyoon Jung , Seunghyun Yoon, Kunwoo Park <i>Findings of the Association for Computational Linguistics (ACL Findings), 2025</i>	<i>2nd author</i>
KOFFVQA: An Objectively Evaluated Free-form VQA Benchmark for Large Vision-Language Models in the Korean Language Yoonshik Kim* and Jaeyoon Jung* <i>IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2025</i>	<i>Co-1st author</i>
CANVAS: Commonsense-Aware Navigation System for Intuitive Human-Robot Interaction Suhwan Choi*, Yongjun Cho*, Minchan Kim*, Jaeyoon Jung* , Myunchul Joe, Yubeen Park, Minseo Kim, Sungwoong Kim, Sungjae Lee, Hwiseong Park, Jiwan Chung, Youngjae Yu <i>International Conference on Robotics and Automation (ICRA), 2025</i> <i>Workshop on Open-World Agents at NeurIPS (Oral, Outstanding Paper Awards), 2024</i>	<i>Co-1st author</i>
The Herd of Open LLMs for Verifying Real-World Claims Yejun Yoon*, Jaeyoon Jung* , Seunghyun Yoon, Kunwoo Park <i>Seventh Workshop on Fact Extraction and VERification (FEVER) at EMNLP (Oral), 2024</i>	<i>Co-1st author</i>
EnCLAP++: Analyzing the EnCLAP Framework for Optimizing Automated Audio Captioning Performance Jaeyeon Kim, Minjeon Jeon, Jaeyoon Jung , Sang Hoon Woo, Jinjoo Lee <i>Detection and Classification of Acoustic Scenes and Events (DCASE) Workshop, 2024</i>	<i>3rd author</i>
EnCLAP: Combining Neural Audio Codec and Audio-Text Joint Embedding for Automated Audio Captioning Jaeyeon Kim, Jaeyoon Jung , Jinjoo Lee, Sang Hoon Woo <i>IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2024</i>	<i>2nd author</i>

Challenge

AVerImaTeC shared task hosted by the ninth FEVER workshop at EACL 2026 1st Place • Built multimodal fact-checking pipeline with agentic multimodal models	December 2025
PokéAgent Challenge at NeurIPS 2025 5th Place, Judge’s Choice Award • Participated in Track 2: speedrunning, long-horizon RPG gameplay	November 2025
AVeriTeC shared task hosted by the eight FEVER workshop at ACL 2025 2nd Place • We enhance our previous year’s fact-checking pipeline using new retrieval method and quantization	May 2025
AVeriTeC shared task hosted by the seventh FEVER workshop at EMNLP 2024 2nd Place • We build our fact-checking pipeline with open LLMs, specifically llama in AVeriTeC fact-checking dataset. • Since our pipeline was developed with open LLMs, our system was selected as a baseline for the eight workshop.	August 2024
Detection and Classification of Acoustic Scenes and Events 2024 Challenge 2nd Place • Participated in Task 6: Automated Audio Captioning, and Task 8: Language-Based Audio Retrieval.	June 2024

Awards

5th Uni-DTHON Challenge 1st Place	November 2025
NeurIPS 2024 - Open World Agent Workshop Outstanding Paper Award	December 2024
LG Display Product Quality Classification (LG Aimers 2) 1st Place • Implemented a robust regressor and fitted classifier, then ensembled them using custom hard voting to create a generalized model. Because our model focused on avoiding overfitting, we achieved a public score of 10, but secured first place in the final ranking, while the scores of other public top 10 models fluctuated.	March 2023
Samsung AI Challenge (3D Metrology) 1st Place • Experimented the diverse method to transfer simulated SEM images to real ones for predict depth maps. Since only simulated SEM images were paired with depth maps, we utilized CycleGAN to perform unpaired domain transfer from simulated SEM images to real SEM images.	October 2022

- improved the quality of predicting depth maps by employing cosine similarity-based KNN to identify relevant depth maps for reference during the generation stage.

LG Innotek Radar Performance Prediction (LG Aimers 1) | 2nd Place **September 2022**

- Conducted an analysis of tabular data from a radar building factory, using both domain knowledge and SHAP values. Throughout this process, derived variables were generated using domain knowledge, and irrelevant variables were removed based on SHAP values, resulting in a more informative dataset.
- Trained the boosting ensemble model based on refined dataset with stratified kfold method since the data distribution is imbalanced.

2023 SWUniv AI Challenge 4th Prize	November 2023
2022 Soongsil University AI contest 2nd Prize	November 2022
2022 SWUniv AI Challenge 7th Place	November 2022
2022 Naver AI RUSH Reach the final	July – August 2022
2022 Dankook University AI Challenge 7th Place	July 2022
2021 Soongsil University AI contest participation Prize	November 2021

Activities

Deepest: SNU Deep Learning Society **September 2025 - Current**

Pseudo Lab: Season 11 Member **March 2025 - June 2025**

Honors

Soongsil University Academic Scholarship	2022
Soongsil University Academic Scholarship	2022
Soongsil University Academic Scholarship	2021
Soongsil University Academic Scholarship	2021
Soongsil University Academic Scholarship	2020

Academic Services

Journal reviewer: IEEE Signal Processing Letters (SPL)