Yeonsung Jung

PH.D. CANDIDATE · KAIST AI

ys.jung@kaist.ac.kr | yeonsungjung.github.io | LinkedIn

Research Interests

My current research interests lie in building **Self-improving agents** (closely working with Ninareh Mehrabi and Sina Shaham from Meta SuperIntelligence Labs) and developing **Reliable Visual reasoning** for vision–language models. Beyond these, I have explored a broad range of topics in **Reliable AI**, including model robustness spanning both 2D and 3D vision as well as domain-agnostic methods, **Jailbreaking in MLLMs**, **Visual Autoregressive Models**, and **Image editing**.

Education

Ph.D. in Graduate School of AI — *Seoul, Korea*

Korea Advanced Institute of Science and Technology (KAIST) — *Sep. 2020 – Present*

- Advisor: Prof. Eunho Yang

M.S.E. in Electrical and Computer Engineering — *Suwon, Korea*

Sungkyunkwan University (SKKU) — *Mar. 2018 – Feb. 2020*

- Advisor: Prof. Joyce Jiyoung Whang

B.Ec. in Statistics & B.E. in Computer Science and Engineering — *Suwon, Korea*

Sungkyunkwan University (SKKU) — *Mar. 2014 – Feb. 2018*

- Advisor: Prof. Joyce Jiyoung Whang

Preprints

Co-Evolving Agents: Learning from Failures as Hard Negatives
Yeonsung Jung, Trilok Padhi, Sina Shaham, Dipika Khullar, Joonhyun Jeong, Ninareh Mehrabi†, Eunho Yang†

Web Agents Are Still Greedy: Progress-Aware Action Generation and Selection via Meta-Plan
Joonhyun Jeong, Gilhyun Nam, **Yeonsung Jung**, Eunho Yang

MeZO-A3dam: Memory-efficient Zeroth-order Adam with Adaptivity Adjustments for Fine-tuning LLMs
Sihwan Park*, Jihun Yun*, Sung-Yub Kim, June Yong Yang, **Yeonsung Jung**, Souvik Kundu, Kyungsu Kim, Eunho Yang

3D Scene Decomposition under Occlusion via Multi-view-aware Inpainting
Heecheol Yun, **Yeonsung Jung**, Eunho Yang

Conference Publications

Early Timestep Zero-Shot Candidate Selection for Instruction-Guided Image Editing — *ICCV 2025*
Joowon Kim, Ziseok Lee, Donghyeon Cho, Sanghyun Jo, **Yeonsung Jung**, Kyungsu Kim, Eunho Yang

Playing the Fool: Jailbreaking LLMs and Multimodal LLMs with Out-of-Distribution Strategy — *CVPR 2025*
Joonhyun Jeong, Seyun Bae, **Yeonsung Jung**, Jaeryong Hwang, Eunho Yang

Preserve or Modify? Context-Aware Evaluation for Balancing Preservation and Modification in Text-Guided Image Editing — *CVPR 2025*
Yoonjeon Kim*, Soohyun Ryu*, **Yeonsung Jung**, Hyunkoo Lee, Joowon Kim, June Yong Yang, Jaeryong Hwang, Eunho Yang

LANTERN: Accelerating Visual Autoregressive Models with Relaxed Speculative Decoding — *ICLR 2025*
Doohyuk Jang*, Sihwan Park*, June Yong Yang, **Yeonsung Jung**, Jihun Yun, Souvik Kundu, Sung-Yub Kim†, Eunho Yang†

A Simple Remedy for Dataset Bias via Self-Influence: A Mislabeled Sample Perspective — *NeurIPS 2024*
Yeonsung Jung*, Jaeyun Song*, June Yong Yang, Jinhwa Kim, Sung-Yub Kim, Eunho Yang

PruNeRF: 3D-Aware Segment-Centric Dataset Pruning — *ICML 2024*
Yeonsung Jung, Heecheol Yun, Joonhyung Park, Jinhwa Kim†, Eunho Yang†

Fighting Fire with Fire: Contrastive Debiasing without Bias-free Data via Generative Bias-transformation *ICML 2023*
Yeonsung Jung, Hajin Shim, June Yong Yang, Eunho Yang

Scalable Anti-TrustRank with Qualified Site-level Seeds for Link-based Web Spam Detection *WWW Workshop 2020*
Joyce Jiyoung Whang, **Yeonsung Jung**, Seonggoo Kang, Dongho Yoo, Inderjit S. Dhillon

Fast Asynchronous Anti-TrustRank for Web Spam Detection *WSDM Workshop 2018*
Joyce Jiyoung Whang, **Yeonsung Jung**, Inderjit S. Dhillon, Seonggoo Kang, Jungmin Lee

Work Experience

External Collaborator, NAVER AI *Seongnam, Korea*
Project: Robust Learning in Neural Radiance Fields (*PruNeRF*, presented at ICML 2024, Vienna) *Sep. 2023 – Feb. 2024*

Research Intern, NAVER Search *Seongnam, Korea*
Project: Improve Search Performance through Click Graph Convolutional Networks (applied to production) *Jul. 2019 – Aug. 2019*

Academic Services

NeurIPS	Reviewer
ICML	Reviewer
ICLR	Reviewer
CVPR	Reviewer
ICCV	Reviewer
AAAI	Reviewer
AISTATS	Reviewer
NAACL	Reviewer

Teaching Experience

Instructor, LG Innotek *2022–2025*
Machine Learning & Deep Learning Basics (PBL)

Instructor, HD Hyundai *2023*
AI-Driven Anomaly Detection and Predictive Maintenance for Construction Machinery Operating Oil
AI-Based Prediction of Remaining Useful Life (RUL) for Construction Machinery Bearings
Image Segmentation for Environmental Recognition at Construction Sites

Teaching Assistant, Daewoo Shipbuilding & Marine Engineering (DSME) *2022–2023*
Augmentation, Self-supervised Learning, and NeRF
Recent Advances in GANs
Image to Image Translation & Spurious Correlation

Teaching Assistant, SAMSUNG DS *2020–2022*
Bayesian Learning Basics
Deep Generative Models
Machine Learning Basics
Data Analysis, Machine Learning, and AI

Instructor, LG Electronics *2022*
Kaggle Data Analysis with Python

Instructor, LG Chem. *2022*
AI Projects for Manufacturing Processes (PBL)