YEONSUNG JUNG

Ph.D. student @ KAIST AI

ys.jung@kaist.ac.kr yeonsungjung.github.io Yeonsung Jung

RESEARCH INTERESTS

My recent research interests include, but are not limited to, **the generalization ability of self-improving agents** (closely working with Ninareh Mehrabi and Sina Shaham) and **visual reasoning in multimodal LLMs**.

Previously, I explored **reliable AI**, **jailbreaking in MLLMs**, and **visual autoregressive models**.

EDUCATION

Korea Advanced Institute of Science and Technology (KAIST)

- Ph.D. in Graduate School of AI (Sep. 2020 - Present)
- Advisor: Prof. Eunho Yang

Sungkyunkwan University (SKKU)

- M.S.E. in Electrical and Computer Engineering (Mar. 2018 - Feb. 2020)
- Advisor: Prof. Joyce Jiyoung Whang

Sungkyunkwan University (SKKU)

- B.Ec. in Statistics and B.E. in Computer Science and Engineering (Mar. 2014 - Feb. 2018)
- Advisor: Prof. Joyce Jiyoung Whang

CONFERENCE PUBLICATIONS

Playing the Fool: Jailbreaking LLMs and Multimodal LLMs with Out-of-Distribution Strategy
Joonhyun Jeong, Seyun Bae, **Yeonsung Jung**, Jaeryong Hwang, Eunho Yang
Computer Vision and Pattern Recognition (CVPR), 2025.

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

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

A Simple Remedy for Dataset Bias via Self-Influence: A Mislabeled Sample Perspective
Yeonsung Jung*, Jaeyun Song*, June Yong Yang, Jinhwa Kim, Sung-Yub Kim, Eunho Yang
Neural Information Processing Systems (NeurIPS), 2024.

PruNeRF: 3D-Aware Segment-Centric Dataset Pruning
Yeonsung Jung, Heecheol Yun, Joonhyung Park, Jinhwa Kim†, Eunho Yang†
International Conference on Machine Learning (ICML), 2024.

Fighting Fire with Fire: Contrastive Debiasing without Bias-free Data via Generative Bias-transformation
Yeonsung Jung, Hajin Shim, June Yong Yang, Eunho Yang
International Conference on Machine Learning (ICML), 2023.

Scalable Anti-TrustRank with Qualified Site-level Seeds for Link-based Web Spam Detection
Joyce Jiyoung Whang, **Yeonsung Jung**, Seonggoo Kang, Dongho Yoo and Inderjit S. Dhillon
Companion Proceedings of the Web Conference (WWW)
Workshop on CyberSafety: Computational Methods in Online Misbehavior, 2020.

Fast Asynchronous Anti-TrustRank for Web Spam Detection
Joyce Jiyoung Whang, **Yeonsung Jung**, Inderjit S. Dhillon, Seonggoo Kang, and Jungmin Lee
ACM International Conference on Web Search and Data Mining (WSDM)
Workshop on MIS2: Misinformation and Misbehavior Mining on the Web), 2018.

PREPRINTS

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
Under Review.

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

WORK EXPERIENCE

NAVER AI (External Collaborator) *Sep. 2023 - Feb. 2024*
Project: Robust Learning in Neural Radiance Fields
- The project outcome, *PruNeRF*, was presented at ICML 2024 (Vienna, Austria)

NAVER Search (Research Intern) *July 2019 - Aug 2019*
Project: Improve Search Performance through Click Graph Convolutional Networks
- The project outcome was applied to NAVER searching system.

PROJECTS

NAVER AI - Contents Generation and Modification via Deep learning
June 2021 - Present

SAMSUNG DS - Sub-task Generation based Adaptive Point/Regional Out-of-Distribution Detection
Oct. 2020 - Oct. 2022

MSIT (Ministry of Science and ICT) - A Machine Learning and Statistical Inference Framework for Explainable Artificial Intelligence
Sep. 2020 - Sep. 2022

KAIST - NT-AI Platform for Cancer Liquid Biopsy
July 2020 - June 2021

NAVER Search - Improving Search Engine Performance using Graph Structure Analysis
May 2017 - May 2019

SAMSUNG ELECTRONICS - TP Center Chip Damage Detection Modeling
Jan. 2018 - Aug. 2019

ACAMDEIC SERVICES

Conference Reviewer

- Neural Information Processing Systems (NeurIPS)
- International Conference on Machine Learning (ICML)
- International Conference on Learning Representations (ICLR)
- Computer Vision and Pattern Recognition (CVPR)
- Artificial Intelligence and Statistics (AISTATS)
- NACCL Workshop on TrustNLP

NACCL Workshop on TrustNLP