Changyeon Kim

(+82) 10-9020-3925 | changyeon.kim@kaist.ac.kr | changyeon.page | csmile-1006 | @cykim1006 | Changyeon Kim

Research Interest

My research centers on training artificial agents to align with human intents, even when explicit reward signals are limited or unavailable. I focus on developing algorithms that derive suitable reward functions from real human preferences or foundational vision-language models, which encapsulate extensive human knowledge. Beyond this primary theme, I also explore broader decision-making challenges, including (M)LLM-based agents, offline reinforcement learning, and generalization of RL solutions across diverse environments.

Education

Korea Advanced Institute of Science and Technology — *Daejeon, S.Korea*

PhD in Artificial Intelligence — *Feb. 2022 - Present*

Advisor: Kimin Lee and Jinwoo Shin

Korea Advanced Institute of Science and Technology — *Daejeon, S.Korea*

B.Sc. in Computer Science and Mathematics (minor) — *Mar. 2016 - Feb. 2021*

Work Experience

Visiting PhD Student @ University of Texas at Austin — *Austin, USA*

with **Yuke Zhu** — *Jul. 2024 - Jan. 2025 (Expected)*

- Ongoing research for reusable reward design encouraging human-aligned behaviors without human reward engineering.

External Collaborator

with **Joseph J. Lim (KAIST)** — *Mar. 2024 - June. 2024*

- Developed a visual reward learning algorithm [P1] for solving complex long-horizon robotic manipulations.

External Collaborator

with **Lisa Lee (Google DeepMind)** — *Apr. 2023 - Aug. 2023*

- Developed an imitation learning algorithm [C3] using multimodal representations for improving generalization ability in unseen variations.

External Collaborator

with **Honglak Lee (University of Michigan)** — *Mar. 2022 - June. 2024*

- Developed an imitation learning algorithm [C3] using multimodal representations for improving generalization ability in unseen variations.
- Developed a reinforcement learning algorithm [W1] for improving generalization ability in varying dynamics.
- Developed a preference-based reinforcement learning algorithm [C2] for modeling non-Markovian human preferences.
- Developed a visual reward learning algorithm [P1] for solving complex long-horizon robotic manipulations.

Machine Learning Engineer — *Seongnam, S.Korea*

Kakao, Recommendation Team — *Dec. 2020 - Feb. 2022*

- Developed ML platform for recommendation system.
- Implemented data pipeline from user feedback to refined user-item interaction matrix data.

Publications

C: Conference, W: Workshop, P: Preprint, *: Equal contribution

[W3] Subtask-Aware Visual Reward Learning from Segmented Demonstrations — *Munich, Germany*

Changyeon Kim, Minho Heo, Doohyun Lee, Jinwoo Shin, Honglak Lee, Kimin Lee, Joseph J Lim — *Nov, 2024.*

- CoRL 2024 Workshop on Mastering Robot Manipulation in a World of Abundant Data (**MRM-D**)

[W2] B-MoCA: Benchmarking Mobile Device Control Agents across Diverse Configurations — *Vienna, Austria*

Juyong Lee, Taywon Min, Minyong Ahn, Dongyoon Hahm, Haeone Lee, **Changyeon Kim**, Kimin Lee — *May, 2024.*

- ICLR 2024 Workshop on Generative Models for Decision Making (**GenAI4DM**), **Spotlight**

[C3] Guide Your Agent with Adaptive Multimodal Rewards — *New Orleans, USA*

Changyeon Kim, Younggyo Seo, Hao Liu, Lisa Lee, Jinwoo Shin, Honglak Lee, Kimin Lee — *Dec, 2023.*

- Conference on Neural Information Processing Systems (**NeurIPS**), 2023.
- **Finalist**, Qualcomm Innovation Fellowship Korea 2024

[C2] Preference Transformer: Modeling Human Preferences using Transformers for RL *Kigali, Rwanda*

CHANGYEON KIM*, JONGJIN PARK*, JINWOO SHIN, HONGLAK LEE, PIETER ABBEEL, KIMIN LEE *May, 2023.*

- International Conference on Learning Representations (**ICLR**), 2023.

[W1] Dynamics-Augmented Decision Transformer for Offline Dynamics Generalization *New Orleans, USA*

CHANGYEON KIM*, JUNSU KIM*, YOUNGGYO SEO, KIMIN LEE, HONGLAK LEE, JINWOO SHIN *Nov, 2022.*

- NeurIPS 2022 Workshop on Offline Reinforcement Learning (NeurIPSW)

[C1] Collecting the Public Perception of AI and Robot Rights *Online*

GABRIEL LIMA, CHANGYEON KIM, SEUNGHO RYU, CHIHYOUNG JEON, MEEYOUNG CHA *Oct, 2020.*

- Conference on Computer-Supported Cooperative Work and Social Computing (**CSCW**), 2020.

[P1] MOI-Mixer: Improving MLP-Mixer with Multi Order Interactions in Sequential Recommendation

HOJOON LEE, DONGYOON HWANG, SUNGHWAN HONG, CHANGYEON KIM, SEUNGRYONG KIM, JAEGUL CHOO

- ArXiv Preprint.

Honors & Awards

2024	**Finalist**, Qualcomm Innovation Fellowship Korea	*Seoul, S.Korea*
2023	**Travel Award**, Conference on Neural Information Processing Systems (NeurIPS)	*New Orleans, USA*
2023	**Scholarship**, KAIST-Google Partnership Program	*Daejeon, S.Korea*
2023	**East Asia Student Travel Grant**, Google	*New Orleans, USA*
2023	**Travel Award**, International Conference on Learning Representations (ICLR)	*Kigali, Rwanda*
2019	**Dean's List (Fall Semester)**, Department of Engineering, KAIST	*Daejeon, S.Korea*
2019	**Line Scholarship (Fall Semester)**, School of Computing, KAIST	*Daejeon, S.Korea*
2017 - 19	**National Science and Engineering Scholarship**, Korea Ministry of Science and ICT	*Daejeon, S.Korea*
2017	**Kwanjeong Scholarship (Spring Semester)**, KAIST	*Daejeon, S.Korea*

Invited Talks

Guide Your Agent with Adaptive Multimodal Rewards *NeurIPS 2023*

IMITATION LEARNING FRAMEWORK WITH VLM REWARDS FOR BETTER GENERALIZATION

- LG AI Research, New Orleans, USA (2023)

Academic Services

Conference Reviewer ICML (2024, 2025), NeurIPS (2024), ICLR (2025)
Workshop Reviewer ICML Frontiers4LCD 2023, CoRL MRM-D 2024

Skills

ML/DL Pytorch, Pytorch-lightning, JAX/Flax
Programming Python, C++
Big Data Kafka, SQL, MongoDB, Hadoop, Trino(Presto)
DevOps Git, Docker, Kubernetes
Languages Korean (Native), English (Fluent), Japanese (Advanced)