Hyeonbin Hwang

Ph.D. Student, Graduate School of AI, KAIST
hbin0701@kaist.ac.kr | hbin0701.github.io

Research Interests

My research focuses on language model reasoning and agentic AI. I develop general post-training methods driven by self-training and reinforcement learning, with the long-term goal of extracting richer learning signals from data.

Education

KAIST Aug 2026 – Present
Ph.D. in Artificial Intelligence, Graduate School of AI
Advisor: Seongjoon Oh

KAIST Feb 2024 – Aug 2026
M.S. in Artificial Intelligence, Graduate School of AI
Advisor: Minjoon Seo

KAIST Sep 2019 – Feb 2024
B.S. in Computer Science, School of Computing
Advisor: Tae-Kyun Kim

Experience

LG AI Research Sep 2026 – Present
Research Intern, SuperIntelligence Lab *Seoul, South Korea*
Working on credit assignment in long-horizon RL.

KRAFTON AI May 2025 – Nov 2025
Research Intern, AI Companion Team *Seoul, South Korea*
Worked on a gameplay agent for PUBG in collaboration with NVIDIA.

NAVER CLOVA Mar 2022 – Aug 2022
Research Intern, Healthcare AI *Seongnam, South Korea*
Researched medical text summarization and domain-specific embeddings.

Honors & Awards

Hackathon Winner, Qualcomm–KAIST Innovation Award 2023

Young-Han Kim Global Leadership Scholarship 2022

Academic Service

Conference reviewer: ICLR 2026, ICML 2026 (**Gold Reviewer**), NeurIPS 2026.

Journal reviewer: Transactions on Machine Learning Research (TMLR), 2026.

Languages

Korean (native); English (fluent; TOEFL 113); Spanish (fluent; DELE C1).

Publications

[1] **PUBG Ally: A Conversational Embodied Agent as an AI Teammate**
PUBG Ally Team: I. Chen, Y. Cho, S. Chung, J. Hong, **H. Hwang**, …, K. Yoo.
Preprint, 2026.

[2] **Intrinsic Task Symmetry Drives Generalization in Algorithmic Tasks**
H. Hwang, Y. Park.
ICML 2026.

[3] **Characterizing Pattern Matching and Its Limits on Compositional Task Structures**
H. Chang, J. Park, H. Cho, S. Yang, M. Ko, **H. Hwang**, S. Won, D. Lee, Y. Ahn, M. Seo.
ICLR 2026.

[4] **The CoT Encyclopedia: Analyzing, Predicting, and Controlling how a Reasoning Model will Think**
S. Lee, S. Kim, M. Seo, Y. Jo, D. Go, **H. Hwang**, J. Park, X. Yue, S. Welleck, G. Neubig, M. Lee, M. Seo.
ICLR 2026.

[5] **Latent Reasoning via Sentence Embedding Prediction**
H. Hwang, B. Jeon, S. Kim, J. Kim, H. Chang, S. Yang, S. Won, D. Lee, Y. Ahn, M. Seo.
COLM 2025 RAM 2 Workshop, ***Oral****.*

[6] **The BiGGen Bench: A Principled Benchmark for Fine-grained Evaluation of Language Models with Language Models**
S. Kim, J. Suk, …, **H. Hwang**, …, M. Seo.
NAACL 2025, ***Best Paper****.*

[7] **Knowledge Entropy Decay during Language Model Pretraining Hinders New Knowledge Acquisition**
J. Kim, H. Lee, H. Cho, J. Jang, **H. Hwang**, S. Won, Y. Ahn, D. Lee, M. Seo.
ICLR 2025, ***Oral****; AAAI 2025 KnowFM Workshop,* ***Best Paper****.*

[8] **Self-Explore: Enhancing Mathematical Reasoning in Language Models with Fine-grained Rewards**
H. Hwang, D. Kim, S. Kim, S. Ye, M. Seo.
Findings of EMNLP 2024; ACL 2024 NLRSE Workshop, ***Oral****.*

[9] **FLASK: Fine-grained Language Model Evaluation based on Alignment Skill Sets**
S. Ye, D. Kim, S. Kim, **H. Hwang**, S. Kim, Y. Jo, J. Thorne, J. Kim, M. Seo.
ICLR 2024, ***Spotlight****.*

[10] **Investigating the Effectiveness of Task-Agnostic Prefix Prompt for Instruction Following**
S. Ye, **H. Hwang**, S. Yang, H. Yun, Y. Kim, M. Seo.
AAAI 2024.

[11] **MED-SE: Medical Entity Definition-based Sentence Embedding**
H. Hwang, H. Yoo, Y. Choi.
Preprint, 2022.

[12] **Vision Transformer Equipped with Neural Resizer on Facial Expression Recognition Task**
H. Hwang, S. Kim, W.-J. Park, J. Seo, K. Ko, H. Yeo.
ICASSP 2022.