

Hyeonbin Hwang
Ph.D Student at KAIST AI

Contact

E-mail hbin0701@kaist.ac.kr
Homepage <https://hbin0701.github.io>

Interest

My long-term goal is to build models that can learn beyond what human data can teach. I'm currently working on an agentic system for post-training, leveraging autonomous research agents and iterative harness optimization.

Experience

Research Intern

SuperIntelligence Lab, LG AI Research
Seoul, Republic of Korea

Sep 2026 –

Research Intern

AI Companion Team, KRAFTON AI
Seoul, Republic of Korea

May 2025 – Nov 2025

- [Game-Play Agent for PUBG \(w. NVIDIA\)](#)

Research Intern

LK Lab, KAIST
Seoul, Republic of Korea

Dec 2022 – Feb 2024

- In-Context Learning
- Fine-grained LM Evaluation

Research Intern

Naver CLOVA, Healthcare AI
Seongnam, Republic of Korea

Mar 2022 – Aug 2022

- Medical Text Summarization & Embedding

Education

KAIST

Ph.D., Graduate School of AI
Advisor : Seongjoon Oh

Aug 2026 – Aug 2030 (Exp.)

KAIST

M.S., Graduate School of AI
Advisor : Minjoon Seo

Feb 2024 – Aug 2026

KAIST

B.S., School of Computing
Advisor : Tae-Kyun Kim

Sep 2019 – Feb 2024

Languages

- | | |
|-----------|--------------------|
| - Korean | Native |
| - English | Fluent (TOEFL 113) |
| - Spanish | Fluent (DELE C1) |

Honors & Awards

- Young-Han Kim Global Leadership Scholarship Awardee (2022)
- Qualcomm-KAIST Innovation Award Hackathon Winner (2023)

Reviewer Experience

- ICLR (2026), ICML (2026, **Gold Reviewer** (Top 25%)), NeurIPS (2026), TMLR (2026)

Selected Publications

[1] **Hwang, H.**, & Park, Y. (2026). *Intrinsic Task Symmetry Drives Generalization in Algorithmic Tasks* (ICML 2026)

[2] **Hwang, H.**, Jeon, B., Kim, S., Kim, J., Chang, H., Yang, S., Won, S., Lee, D., Ahn, Y., ... (2025). *Let's Predict Sentence by Sentence*. arXiv preprint arXiv:2505.22202. (COLM 2026 RAM 2 Workshop **Oral**)

[3] Chang, H., Park, J., Cho, H., Yang, S., Ko, M., **Hwang, H.**, Won, S., Lee, D., Ahn, Y., ... (2025). *The Coverage Principle: A Framework for Understanding Compositional Generalization*. arXiv preprint arXiv:2505.20278. (ICLR 2026)

[4] Lee, S., Kim, S., Seo, M., Jo, Y., Go, D., **Hwang, H.**, Park, J., Yue, X., Welleck, S., ... (2025). *The CoT Encyclopedia: Analyzing, Predicting, and Controlling how a Reasoning Model will Think*. arXiv preprint arXiv:2505.10185. (ICLR 2026)

[5] Kim, J., Lee, H., Cho, H., Jang, J., **Hwang, H.**, Won, S., Ahn, Y., Lee, D., & Seo, M. (2024). *Knowledge Entropy Decay during Language Model Pretraining Hinders New Knowledge Acquisition*. (ICLR 2025 **Oral** & AAAI Workshop KnowFM 2025 **Best Paper**)

[6] Kim, S., Suk, J., Cho, J. Y., Longpre, S., Kim, C., Yoon, D., Son, G., Cho, Y., & **Hwang, H.** (2024). *The BiGGen Bench: A Principled Benchmark for Fine-grained Evaluation of Language Models with Language Models* (NAACL 2025 **Best Paper**)

[7] **Hwang, H.**, Kim, D., Kim, S., Ye, S., & Seo, M. (2024). *Self-Explore to Avoid the Pit: Improving the Reasoning Capabilities of Language Models with Fine-grained Rewards*. (EMNLP Findings 2024 & ACL Workshop NLRSE 2024 **Oral**)

[8] Ye, S., **Hwang, H.**, Yang, S., Yun, H., Kim, Y., & Seo, M. (2023). *Investigating the Effectiveness of Task-Agnostic Prefix Prompt for Instruction Following*. (AAAI 2024)

[9] Ye, S., Kim, D., Kim, S., **Hwang, H.**, Kim, S., Jo, Y., Thorne, J., Kim, J., & Seo, M. (2023). *FLASK: Fine-grained language model evaluation based on alignment skill sets*. (ICLR 2024 **Spotlight**)

[10] **Hwang, H.**, Kim, S., Park, W. J., Seo, J., Ko, K., & Yeo, H. (2022). *Vision transformer equipped with neural resizer on facial expression recognition task*. (ICASSP 2022)