

Doyoung Kim

✉ doyoungkim@kaist.ac.kr 🌐 <https://doyoungkim-ml.github.io/> 🐦 @doyoungkim_ml

Masters student developing generalist decision-making models by modeling human cognitive processes

Education

Graduate School of AI (GSAI), KAIST

Feb 2023 – Aug 2025

Master's degree in AI

Korea Advanced Institute of Science and Technology (KAIST) *Feb 2017 – Aug 2022*

Bachelor of Mathematics, Double Major in Computer Science

Gyeonggi Science High School

Feb 2014 – Dec 2016

Mathematics

Publications

Google Scholar: <https://scholar.google.com/citations?user=PJR9ogMAAAAJ>

Semantic Scholar: <https://www.semanticscholar.org/author/Doyoung-Kim/2180527259>

* *denotes equal contributions*

Peer-Reviewed Conference/Journal Papers:

- [C8] Hyeonbin Hwang, **Doyoung Kim**, Seungwon Kim, Seonghyeon Ye, Minjoon Seo. “Self-Explore to Avoid the Pit: Improving the Reasoning Capabilities of Language Models with Fine-grained Rewards.” EMNLP Findings, 2024.
- [C7] Hyunji Lee*, **Doyoung Kim***, Jihoon Jun, Sejune Joo, Joel Jang, Kyoung-Woon On, Minjoon Seo. “Semiparametric Token-Sequence Co-Supervision.” ACL, 2024.
- [C6] Hyunji Lee, Sejune Joo, Chaeun Kim, Joel Jang, **Doyoung Kim**, Kyoung-Woon On, Minjoon Seo. “How Well Do Large Language Models Truly Ground?” NAACL, 2024.
- [C5] Seonghyeon Ye*, **Doyoung Kim***, Sungdong Kim, Hyeonbin Hwang, Seungwon Kim, Yongrae Jo, James Thorne, Juho Kim, Minjoon Seo. “FLASK: Fine-grained Language Model Evaluation based on Alignment Skill Sets.” *ICLR Spotlight, 2024*.
- [C4] Seungwon Kim, Sejune Joo, **Doyoung Kim**, Joel Jang, Seonghyeon Ye, Jaemin Shin, Minjoon Seo. “The CoT Collection: Improving Zero-shot and Few-shot Learning of Language Models via Chain-of-Thought Fine-Tuning.” EMNLP, 2024.
- [C3] Joel Jang, Seungwon Kim, Seonghyeon Ye, **Doyoung Kim**, Lajanugen Logeswaran, Moon-tae Lee, Kyungjae Lee, Minjoon Seo. “Exploring the benefits of training expert language models over instruction tuning.” ICML, 2023.
- [C2] Seonghyeon Ye, **Doyoung Kim**, Joel Jang, Joongbo Shin, Minjoon Seo. “Guess the instruction! making language models stronger zero-shot learners.” ICLR, 2023.
- [C1] Seonghyeon Ye, Joel Jang, **Doyoung Kim**, Yongrae Jo, Minjoon Seo. “Efficiently Enhancing Zero-Shot Performance of Instruction Following Model via Retrieval of Soft Prompt.” EMNLP Findings, 2023.

Workshop:

- [W1] **Doyoung Kim**, Jongwon Lee, Jinho Park, Minjoon Seo. “How language models extrapolate outside the training data: A case study in Textualized Gridworld.” NeurIPS Compositional Learning Workshop 2024.

Projects

** denotes equal contributions*

[Pr1] Seonghyeon Ye*, Yongrae Jo*, **Doyoung Kim***, Sungdong Kim, Hyeonbin Hwang, Minjoon Seo. “SelFee: Iterative Self-Revising LLM Empowered by Self-Feedback Generation” 2023.

Invited Talks

Fine-grained Evaluation for GenAI, National Information Society Agency (NIA)	Jul 2024
Fine-grained Evaluation for GenAI, Adobe Research	Aug 2023

Selected Honors and Awards

Silver medal , College Students Mathematics Competition, Math major Div.	Dec 2020
Bronze medal , College Students Mathematics Competition, Freshmen Div.	Dec 2017

Professional Services

Volunteer Activity:

- TA for ICPC, World Friends ICT in Ethiopia (ASTU) Jul 2017 - Aug 2017

Teaching Assistant:

- Upstage Boostcamp AI Tech, Machine Reading & Comprehension Mar 2024 - Oct 2024

Student Mentoring:

- Jaehyuk Doo, Now MS at KAIST AI Jul 2024 - Dec 2024

Reviewer:

- **ARR rolling reviews** (2024 Feb, April, June)
- **NeurIPS 2024 Workshop on Compositional Learning** (2024)