

# Seungone Kim

✉ seungone@kaist.ac.kr • 🌐 <https://github.com/SeungoneKim>

My main research goal is to understand the inherent limitations of large language models (LLMs) and to enhance their capabilities for more effective deployment in real-world scenarios.

Specifically, I am interested in (1) improving the **reasoning & planning** abilities of LLM agents [C2, C4, O1]; (2) ensuring LLMs to truly follow human instructions via **alignment methods** (e.g., Instruction Tuning and RLHF) [C3, P3]; (3) understanding the properties and behaviors of LLMs through **fine-grained evaluation** [P1, P2]; (4) grounding LMs to share **commonsense knowledge** as humans do [C1].

## Education

---

- **Korea Advanced Institute of Science and Technology (KAIST)** **Seoul, Korea**  
*M.S. in Artificial Intelligence* *March 2023 – Present*  
Advisor : Minjoon Seo
- **Yonsei University** **Seoul, Korea**  
*B.S. in Computer Science* *March 2018 – February 2023*

## Publications

---

### Peer-Reviewed Conference Papers.....

- **[C4]** *The CoT Collection: Improving Zero-shot and Few-shot Learning of Language Models via Chain-of-Thought Fine-Tuning*  
The 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)  
**Seungone Kim\***, SeJune Joo\*, Doyoung Kim, Joel Jang, Seonghyeon Ye, Jamin Shin, Minjoon Seo  
[Paper] [Code]
- **[C3]** *Exploring the Benefits of Training Expert Language Models over Instruction Tuning*  
The 40th International Conference on Machine Learning (ICML 2023)  
Joel Jang, **Seungone Kim**, Seonghyeon Ye, Doyoung Kim, Lajanugen Logeswaran, Moontae Lee, Kyungjae Lee, Minjoon Seo  
[Paper] [Code]
- **[C2]** *CoTEVer: Chain of Thought Prompting Annotation Toolkit for Explanation Verification*  
The 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2023)  
**Seungone Kim**, Se June Joo, Yul Jang, Hyungjoo Chae, Jinyoung Yeo  
[Paper] [Code]
- **[C1]** *Mind the Gap! Injecting Commonsense Knowledge for Abstractive Dialogue Summarization*  
Proceedings of the 29th International Conference on Computational Linguistics (COLING 2022)  
**Seungone Kim\***, Se June Joo\*, Hyungjoo Chae\*, Chaehyeon Kim\*, Seungwon Hwang, Jinyoung Yeo  
[Paper] [Code]

## Preprints

---

- **[P1] Prometheus: Inducing Fine-grained Evaluation Capability in Language Models**  
Under Review at ICLR 2024 & NeurIPS 2023 Instruction Following Workshop  
**Seungone Kim\***, Jamin Shin\*, Yejin Cho\*, Joel Jang, Shayne Longpre, Hwaran Lee, Sangdoo Yun, Seongjin Shin, Sungdong Kim, James Thorne, Minjoon Seo  
[Paper] [Code]
- **[P2] FLASK: Fine-grained Language Model Evaluation based on Alignment Skill Sets**  
Under Review at ICLR 2024 & NeurIPS 2023 Instruction Following Workshop  
Seonghyeon Ye, Doyoung Kim, Sungdong Kim, Hyeonbin Hwang, **Seungone Kim**, Yongrae Jo, James Thorne, Juho Kim, Minjoon Seo  
[Paper] [Code]
- **[P3] Multi-objective Reinforcement Learning from Personalized Human Feedback**  
Planning for Submission at ICML 2024  
Joel Jang, **Seungone Kim**, Bill Yuchen Lin, Yizhong Wang, Jack Hessel, Luke Zettlemoyer, Hannaneh Hajishirzi, Yejin Choi, Prithviraj Ammanabrolu  
[Paper] [Code]

## Research Experience

---

- **NAVER AI Lab** **Seongnam, Korea**  
*Research Internship at Language Research Team* *Mar 2023 – Present*  
I have lead a research project, exploring whether open-sourced LMs could replace GPT-4 evaluation [P1].
- **Korea Advanced Institute of Science and Technology (KAIST)** **Seoul, Korea**  
*Research Internship at Language and Knowledge Laboratory* *Jul 2022 – Feb 2023*  
I participated in a research project building Expert language models, overcoming the weakness of a single Instruction-tuned LM [C3].
- **Seoul National University** **Seoul, Korea**  
*Research Internship at Language and Data Intelligence Laboratory* *Sep 2021 – Jun 2022*  
I have participated in a research project building a multilingual language model for low-resource languages via continual cross-lingual pre-training.

## Invited Talks

---

- **Commonsense Knowledge Consortium 1st Workshop** **Seoul, Korea**  
*Presented paper accepted at COLING 2022[C1]* *July 2022*  
Host: Prof. Il Chul Moon(KAIST), Chanyoung Park(KAIST), Jinyoung Yeo(Yonsei), Seungwon Hwang(SNU), Joseph Lim(KAIST), Jonghyeon Choi (Yonsei)
- **Yonsei University AI Workshop** **Seoul, Korea**  
*Presented paper accepted at COLING 2022[C1]* *October 2022*  
Host: Prof. Jonghyeon Choi (Yonsei)
- **Innovation Session Series** **Frankfurt, Germany**  
*Presented paper accepted at EACL 2023[C2]* *June 2023*  
Host: Laux Michael (SAP)

## Services

---

- **Conference Reviewer:** EACL 2023, EMNLP 2023
- **Conference Workshop Reviewer:** NeurIPS 2023 (Instruction Following WS)
- **Secondary Reviewer:** ACL 2023

## Mentoring

---

- **Juyoung Suk:** B.S. Student, KAIST (July 2023 - Present)

## Teaching Experience & Extracurricular Experience

---

- **Korea Army Augmented to US Army (KATUSA)**      **United States Army Garrison Humphreys**  
*Sergeant*      *January 2019 – August 2020*  
I have worked as a translator and a Senior Katusa for Chief Warrant Officers flying Chinook Helicopters across the Peninsula, serving for General Robert B. Abrams, the 25th Commander of U.S. Forces Korea.
- **Samsung Dream Class**      **Samsung Welfare**  
*Mentor*      *March 2022 – February 2023*  
I have participated in the Dream Class program as a Mentor, organized by Samsung in Korea. I have taught Python programming to middle-school students including object-oriented programming, machine learning with Numpy and Pytorch, and data visualization with Seaborn.