

Seungone Kim

seungone@cmu.edu

<https://seungonekim.github.io/>

EDUCATION	Carnegie Mellon University, Language Technologies Institute (LTI) Ph.D. Student, School of Computer Science Advisor: Prof. Graham Neubig and Prof. Sean Welleck	August 2024 - Present
	Korea Advanced Institute of Science and Technology (KAIST) M.S. in Artificial Intelligence Advisor: Prof. Minjoon Seo	March 2023 - August 2024
	Yonsei University B.S. in Computer Science Major GPA: 3.99 / 4.30; Cumulative GPA: 3.85 / 4.30	March 2018 - February 2023
RESEARCH INTERESTS	LLM Evaluation, Systems that Improve with Feedback, Instruction Tuning, Chain-of-Thought Reasoning	
CONFERENCE PUBLICATIONS	<ul style="list-style-type: none">[1] Prometheus 2: An Open Source Language Model Specialized in Evaluating Other Language Models Seungone Kim*, Juyoung Suk*, Shayne Longpre, Bill Yuchen Lin, Jamin Shin, Sean Welleck, Graham Neubig, Moontae Lee, Kyungjae Lee, Minjoon Seo EMNLP 2024[2] Language Models as Compilers: Simulating Pseudocode Execution Improves Algorithmic Reasoning in Language Models Hyunjoo Chae, Yeonghyeon Kim, Seungone Kim, Kai Tzu-iunn Ong, Beong-woo Kwak, Moohyeon Kim, Seonghwan Kim, Taeyoon Kwon, Jiwan Chung, Youngjae Yu, Jinyoung Yeo EMNLP 2024[3] Self-Explore to Avoid the Pit: Improving the Reasoning Capabilities of Language Models with Fine-grained Rewards Hyeonbin Hwang, Doyoung Kim, Seungone Kim, Seonghyeon Ye, Minjoon Seo EMNLP 2024 Findings[4] Consent in Crisis: The Rapid Decline of the AI Data Commons Shayne Longpre, Robert Mahari, Ariel Lee, Campbell Lund, Hamidah Oderinwale, William Brannon, Nayan Saxena, Naana Obeng-Marnu, Tobin South, Cole Hunter, Kevin Klyman, Christopher Klammer, Hailey Schoelkopf, Nikhil Singh, Manuel Cherep, Ahmad Anis, An Dinh, Caroline Chitongo, Da Yin, Damien Sileo, Deividas Mataciunas, Diganta Misra, Emad Alghamdi, Enrico Shipole, Jianguo Zhang, Joanna Materzynska, Kun Qian, Kush Tiwary, Lester James Validad Miranda, Manan Dey, Minnie Liang, Mohammed Hamdy, Niklas Muennighoff, Seonghyeon Ye, Seungone Kim, Shrestha Mohanty, Vipul Gupta, Vivek Sharma, Vu Minh Chien, Xuhui Zhou, Yizhi Li, Caiming Xiong, Luis Villa, Stella Biderman, Hanlin Li, Daphne Ippolito, Sara Hooker, Jad Kabbara, Sandy Pentland NeurIPS 2024 (Datasets & Benchmark Track)[5] Aligning to Thousands of Preferences via System Message Generalization Seongyun Lee*, Sue Hyun Park*, Seungone Kim, Minjoon Seo NeurIPS 2024[6] Prometheus-Vision: Vision-Language Model as a Judge for Fine-Grained Evaluation Seongyun Lee*, Seungone Kim*, Sue Hyun Park, Geewook Kim, Minjoon Seo ACL 2024 Findings	

- [7] **LangBridge: Multilingual Reasoning Without Multilingual Supervision**
Dongkeun Yoon, Joel Jang, Sungdong Kim, **Seungone Kim**, Sheikh Shafayat, Minjoon Seo
ACL 2024
- [8] **Multi-Task Inference: Can Large Language Models Follow Multiple Instructions at Once?**
Guijin Son*, Sangwon Baek, Sangdae Nam, Ilgyun Jeong, **Seungone Kim***
ACL 2024
- [9] **Prometheus: Inducing Fine-grained Evaluation Capability in Language Models**
Seungone Kim*, Jamin Shin*, Yejin Cho*, Joel Jang, Shayne Longpre, Hwaran Lee, Sangdoo Yun, Seongjin Shin, Sungdong Kim, James Thorne, Minjoon Seo
ICLR 2024
- [10] **FLASK: Fine-grained Language Model Evaluation based on Alignment Skill Sets**
Seonghyeon Ye, Doyoung Kim, Sungdong Kim, Hyeonbin Hwang, **Seungone Kim**, Yongrae Jo, James Thorne, Juho Kim, Minjoon Seo
ICLR 2024 (Spotlight)
Best Paper Honorable Mention, *Instruction Workshop* at NeurIPS 2023
- [11] **The CoT Collection: Improving Zero-shot and Few-shot Learning of Language Models via Chain-of-Thought Fine-Tuning**
Seungone Kim*, Se June Joo*, Doyoung Kim, Joel Jang, Seonghyeon Ye, Jamin Shin, Minjoon Seo
EMNLP 2023
- [12] **Exploring the Benefits of Training Expert Language Models over Instruction Tuning**
Joel Jang, **Seungone Kim**, Seonghyeon Ye, Doyoung Kim, Lajanugen Logeswaran, Moontae Lee, Kyungjae Lee, Minjoon Seo
ICML 2023
- [13] **CoTEVer: Chain of Thought Prompting Annotation Toolkit for Explanation Verification**
Seungone Kim, Se June Joo, Yul Jang, Hyunjoo Chae, Jinyoung Yeo
EACL 2023 (Demonstrations Track)
- [14] **Mind the Gap! Injecting Commonsense Knowledge for Abstractive Dialogue Summarization**
Seungone Kim*, Se June Joo*, Hyunjoo Chae*, Chaehyeon Kim, Seung-won Hwang, Jinyoung Yeo
COLING 2022

WORKSHOP
PAPERS

- [15] **Personalized Soups: Personalized Large Language Model Alignment via Post-hoc Parameter Merging**
Joel Jang, **Seungone Kim**, Bill Yuchen Lin, Yizhong Wang, Jack Hessel, Luke Zettlemoyer, Han-naneh Hajishirzi, Yejin Choi, Prithviraj Ammanabrolu
Adaptive Foundation Models Workshop at NeurIPS 2024

PREPRINTS

- [16] **Evaluating Language Models as Synthetic Data Generators**
Seungone Kim, Juyoung Suk, Xiang Yue, Vijay Viswanathan, Seongyun Lee, Yizhong Wang, Kiril Gashteovski, Carolin Lawrence, Sean Welleck, Graham Neubig
Under Review at ACL 2025

- [17] **The BiGGen Bench: A Principled Benchmark for Fine-grained Evaluation of Language Models with Language Models**
Seungone Kim, Juyoung Suk, JiYong Cho, Shayne Longpre, Chaeun Kim, Dongkeun Yoon, Guijin Son, Yejin Cho, Sheikh Shafayat, Jinheon Baek, Sue Hyun Park, Hyeonbin Hwang, Jinkyung Jo, Hyowon Cho, Haebin Shin, Seongyun Lee, Hanseok Oh, Noah Lee, Namgyu Ho, Se June Joo, Miyoung Ko, Yoonjoo Lee, Hyungjoo Chae, Jamin Shin, Joel Jang, Seonghyeon Ye, Bill Yuchen Lin, Sean Welleck, Graham Neubig, Moontae Lee, Kyungjae Lee, Minjoon Seo
Under Review at NAACL 2025
- [18] **MM-Eval: A Multilingual Meta-Evaluation Benchmark for LLM-as-a-Judge and Reward Models**
 Guijin Son, Dongkeun Yoon, Juyoung Suk, Javier Aula-Blasco, Mano Aslan, Vu Trong Kim, Shayekh Bin Islam, Jaume Prats-Cristia, Luc'ia Tormo-Banuelos, **Seungone Kim**
Under Review at NAACL 2025
- [19] **KMMLU: Measuring Massive Multitask Language Understanding in Korean**
 Guijin Son, Hanwool Lee, Sungdong Kim, **Seungone Kim**, Niklas Muennighoff, Taekyoon Choi, Cheonbok Park, Kang Min Yu, Stella Biderman
Under Review at NAACL 2025
- [20] **Pangea: A Fully Open Multilingual Multimodal LLM for 39 Languages**
 Xiang Yue, Yueqi Song, Akari Asai, **Seungone Kim**, Jean de Dieu Nyandwi, Simran Khanuja, Anjali Kantharuban, Lintang Sutawika, Sathyanarayanan Ramamoorthy, Graham Neubig
Under Review at ICLR 2025
- [21] **Better Instruction-Following Through Minimum Bayes Risk**
 Ian Wu, Patrick Fernandes, Amanda Bertsch, **Seungone Kim**, Sina Pakazad, Graham Neubig
Under Review at ICLR 2025
- [22] **Bridging the Data Provenance Gap Across Text, Speech, and Video**
 Shayne Longpre, Nikhil Singh, Manuel Cherep, Kushagra Tiwary, Joanna Materzynska, William Brannon, Robert Mahari, Manan Dey, Mohammed Hamdy, Nayan Saxena, Ahmad Mustafa Anis, Emad A. Alghamdi, Vu Minh Chien, Naana Obeng-Marnu, Da Yin, Kun Qian, Yizhi LI, Minnie Liang, An Dinh, Shrestha Mohanty, Deividas Mataciunas, Tobin South, Jianguo Zhang, Ariel N. Lee, Campbell S. Lund, Christopher Klamm, Damien Sileo, Diganta Misra, Enrico Shippole, Kevin Klyman, Lester James Validad Miranda, Niklas Muennighoff, Seonghyeon Ye, **Seungone Kim**, Vipul Gupta, Vivek Sharma, Xuhui Zhou, Caiming Xiong, Luis Villa, Stella Biderman, Alex Pentland, Sara Hooker, Jad Kabbara
Under Review at ICLR 2025

RESEARCH EXPERIENCE	Meta, FAIR	May 2025 - August 2025; Seattle, United States
	Research Intern Host: Ilia Kulikov	
	LG AI Research, Advanced Machine Learning Lab Research Intern Host: Kyungjae Lee, Moontae Lee	January 2024 - June 2024; Seoul, Korea
	NAVER AI Lab Research Intern Host: Jamin Shin	March 2023 - December 2023; Seongnam, Korea
SERVICES	Conference Reviewer EACL 2023, ACL 2023, EMNLP 2023, NAACL 2024 (ARR Oct & Dec 2023), ACL 2024 (ARR Feb 2024), NeurIPS 2024, EMNLP 2024 (ARR June 2024), ICLR 2025, NAACL 2025 (ARR Oct 2024), ACL 2024 (ARR Dec 2024)	
	Conference Workshop Reviewer <i>Instruction Workshop at NeurIPS 2023</i>	

INVITED TALKS	Model-based Evaluation Seminar (SKT)	July 2024; Seoul, Korea
	Host: SKT LLM Team	
	NLP Reading Group (Microsoft Research India)	June 2024; Bengaluru, India
	Host: Sanchit Ahuja	
	Multimodal Weekly Webinar (Twelve Labs)	May 2024; San Francisco, USA
	Host: James Le	
	RAM Talk (AI at Meta)	May 2024; New York, USA
	Host: Ilia Kulikov	
	NAVER Tech Meetup (NAVER Cloud)	February 2024; Seoul, Korea
AWARDS	Host: Inho Kang	
	Samsung AI Forum (Samsung Research Global AI Center)	November 2023; Seoul, Korea
	Host: Daehyun Kim	
	AI Seminar Series (Objective Inc)	November 2023; San Francisco, USA
	Host: Pablo Mendes	
MENTORING	Innovation Session Series (SAP)	June 2023; Frankfurt, Germany
	Host: Laux Micahel	
	NEC Student Research Fellowship	2024
	Honorable Mention Award at NeurIPS 2023 Instruction Workshop	2023
LEADERSHIP EXPERIENCE	DEI Scholarship at NeurIPS 2023	2023
	Minyeong Kim: B.S. Student at SNU	August 2024 - Present
	Srijith Radhakrishnan: M.S. Student at CMU	September 2024 - December 2024
	Jessica Kim: B.S. Student at MIT	June 2024 - August 2024
	Juyoung Suk: B.S. Student at KAIST → M.S. Student at KAIST AI	July 2023 - June 2024
	Guijin Son: B.S. Student at Yonsei University	July 2023 - June 2024
	Korean Augmentation to the US Army (KATUSA)	January 2019 - August 2020
	Sergeant at Eighth Army (United States)	