Seungone Kim

seungone@cmu.edu
https://seungonekim.github.io/

EDUCATION

Carnegie Mellon University, Language Technologies Institute (LTI)   September 2024 - Present
Ph.D. Student, School of Computer Science
Advisor: Prof. Graham Neubig and Prof. Sean Welleck

Korea Advanced Institute of Science and Technology (KAIST)   March 2023 - August 2024
M.S. in Artificial Intelligence
Advisor: Prof. Minjoon Seo

Yonsei University   March 2018 - February 2023
B.S. in Computer Science
Major GPA: 3.99 / 4.30; Cumulative GPA: 3.85 / 4.30

RESEARCH INTERESTS

LLM Evaluation, Systems that Improve with Feedback, Instruction Tuning, Chain-of-Thought Reasoning

PUBLICATIONS

[1] 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
ME-FoMo Workshop at ICLR 2024 & *ALVR Workshop* at ACL 2024

[2] LangBridge: Multilingual Reasoning Without Multilingual Supervision
Dongkeun Yoon, Joel Jang, Sungdong Kim, **Seungone Kim**, Sheikh Shafayat, Minjoon Seo
ACL 2024
ME-FoMo Workshop at ICLR 2024 & *KnowledgeNLP Workshop* at ACL 2024

[3] Multi-Task Inference: Can Large Language Models Follow Multiple Instructions at Once?
Guijin Son*, Sangwon Baek, Sangdae Nam, Ilgyun Jeong, **Seungone Kim***
ACL 2024
ME-FoMo Workshop at ICLR 2024 & *SpLU-RoboNLP Workshop* at ACL 2024 (Oral)

[4] 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
Instruction Workshop at NeurIPS 2023

[5] 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 & *LLM Agents Workshop* at
ICLR 2024 & *GenAI4DM Workshop* at ICLR 2024

[6] 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

[7] 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

[8] CoTEVer: Chain of Thought Prompting Annotation Toolkit for Explanation Verification
Seungone Kim, Se June Joo, Yul Jang, Hyungjoo Chae, Jinyoung Yeo
EACL 2023 (Demonstrations Track)

[9] Mind the Gap! Injecting Commonsense Knowledge for Abstractive Dialogue Summarization
Seungone Kim*, Se June Joo*, Hyungjoo Chae*, Chaehyeong Kim, Seung-won Hwang, Jinyoung Yeo
COLING 2022

PREPRINTS

[10] The BiGGen Bench: A Principled Benchmark for Fine-grained Evaluation of Language Models with Language Models
Seungone Kim, Juyoung Suk, JiYong Cho, Shayne Longpre, [23 Authors], Sean Welleck, Graham Neubig, Moontae Lee, Kyungjae Lee, Minjoon Seo
Under Review at NeurIPS 2024 (Datasets & Benchmark Track)

[11] 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
Under Review at EMNLP 2024

[12] Through the Looking-Glass: Tracing Shifts in AI Data Consent across the Web
Shayne Longpre, Robert Mahari, Ariel N. Lee, [31 Authors], **Seungone Kim**, [10 Authors], Daphne Ippolito, Sara Hooker, Jad Kabbara, Alex Pentland
Under Review at NeurIPS 2024 (Datasets & Benchmark Track)

[13] 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 NeurIPS 2024 (Datasets & Benchmark Track)

[14] Aligning to Thousands of Preferences via System Message Generalization
Seongyun Lee*, Sue Hyun Park*, **Seungone Kim**, Minjoon Seo
Under Review at NeurIPS 2024
ConvAI Workshop at ACL 2024 (Oral)

[15] 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
Under Review at EMNLP 2024
NLRSE Workshop at ACL 2024 (Oral)

[16] Language Models as Compilers: Simulating Pseudocode Execution Improves Algorithmic Reasoning in Language Models
Hyungjoo Chae, Yeonghyeon Kim, **Seungone Kim**, Kai Tzu-iunn Ong, Beong-woo Kwak, Moohyeon Kim, Seonghwan Kim, Taeyoon Kwon, Jiwan Chung, Youngjae Yu, Jinyoung Yeo
Under Review at EMNLP 2024
NLRSE Workshop at ACL 2024

[17] 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, Hannaneh Hajishirzi, Yejin Choi, Prithviraj Ammanabrolu
arXiv Preprint

RESEARCH EXPERIENCE		
LG AI Research, Advanced Machine Learning Lab	January 2024 - June 2024; Seoul, Korea	
Research Intern		
Host: Kyungjae Lee, Moontae Lee		
NAVER AI Lab	March 2023 - December 2023; Seongnam, Korea	
Research Intern		
Host: Jamin Shin		

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)

Conference Workshop Reviewer
Instruction Workshop at NeurIPS 2023

INVITED TALKS

Model-based Evaluation Seminar (SKT)   July 2024; Seoul, Korea
Host: TBD

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
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

Innovation Session Series (SAP)   June 2023; Frankfurt, Germany
Host: Laux Micahel

AWARDS

Honorable Mention Award at NeurIPS 2023 Instruction Workshop   2023
DEI Scholarship at NeurIPS 2023   2023

MENTORING

Minyeong Kim: B.S. Student at SNU   August 2024 - Present
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

| LEADERSHIP EXPERIENCE | **Korean Augmentation to the US Army (KATUSA)** Sergeant at Eighth Army (United States) | January 2019 - August 2020 |