Junmo Kang

Ph.D. Student at Georgia Tech

junmo.kang@gatech.edu | jm-kang.github.io

RESEARCH INTERESTS

Natural Language Processing (NLP)

Large Language Models, Instruction-Driven Learning, Retrieval-Augmentation, Efficient NLP, Robustness

My research focuses on developing NLP models that are **efficient** and **robust**, with the goal of ensuring their practicality in real-world scenarios. In particular, I am interested in *i) how to induce NLP systems to be more scalable and cheaper in terms of data, compute, or parameters*, and *ii) how to design NLP models that remain robust when confronted with unseen cases in the wild*.

EDUCATION

Georgia Institute of Technology — *Atlanta, GA, USA*

Ph.D. in Computer Science — Aug. 2022 - Present

- Research Assistant at NLP Lab (Advisor: Alan Ritter, Wei Xu)
- GPA: 4.0 / 4.0

Korea Advanced Institute of Science and Technology (KAIST) — *Daejeon, Republic of Korea*

M.S. in Computer Science — Feb. 2021

- Research Assistant at IR&NLP Lab (Advisor: Sung-Hyon Myaeng)
- Thesis committee: Sung-Hyong Myaeng, Hojin Choi, Alice Oh
- GPA: 4.03 / 4.30

Chungnam National University — *Daejeon, Republic of Korea*

B.E. in Computer Science & Engineering — Feb. 2019

- Summa Cum Laude
- GPA: 4.30 / 4.50 (Rank: 1/125 in CSE, Major GPA: 4.41)

PUBLICATIONS

* indicates equal contribution

[1] Self-Specialization: Uncovering Latent Expertise within Large Language Models — Preprint
Junmo Kang, Hongyin Luo, Yada Zhu, James Glass, David Cox, Alan Ritter, Rogerio Feris, Leonid Karlinsky [pdf]

[2] Schema-Driven Information Extraction from Heterogeneous Tables — Preprint
Fan Bai, **Junmo Kang**, Gabriel Stanovsky, Dayne Freitag, Alan Ritter [pdf]

[3] Discern and Answer: Mitigating the Impact of Misinformation in Retrieval-Augmented Models with Discriminators — Preprint
Giwon Hong*, Jeonghwan Kim*, **Junmo Kang***, Sung-Hyon Myaeng, Joyce Jiyoung Whang [pdf]

[4] Distill or Annotate? Cost-Efficient Fine-Tuning of Compact Models — ACL 2023
Junmo Kang, Wei Xu, Alan Ritter [pdf]

[5] Graph-Induced Transformers for Efficient Multi-Hop Question Answering — EMNLP 2022
Giwon Hong, Jeonghwan Kim, **Junmo Kang**, Sung-Hyon Myaeng [pdf]

[6] Exploiting Numerical-Contextual Knowledge to Improve Numerical Reasoning in Question Answering Findings of NAACL 2022
Jeonghwan Kim, **Junmo Kang**, Giwon Hong, Kyung-min Kim, Sung-Hyon Myaeng [pdf]

[7] Ultra-High Dimensional Sparse Representations with Binarization for Efficient Text Retrieval EMNLP 2021
Kyoung-Rok Jang, **Junmo Kang**, Giwon Hong, Sung-Hyon Myaeng, Joohee Park, Taewon Yoon, Heecheol Seo [pdf]

[8] Leveraging Order-Free Tag Relations for Context-Aware Recommendation EMNLP 2021
Junmo Kang, Jeonghwan Kim, Suwon Shin, Sung-Hyon Myaeng [pdf]

[9] Have You Seen That Number? Investigating Extrapolation in Question Answering Models EMNLP 2021
Jeonghwan Kim, Giwon Hong, Kyung-min Kim, **Junmo Kang**, Sung-Hyon Myaeng [pdf]

[10] Can You Distinguish Truthful from Fake Reviews? User Analysis and Assistance Tool for Fake Review Detection HCI+NLP@EACL 2021
Jeonghwan Kim*, **Junmo Kang***, Suwon Shin*, Sung-Hyon Myaeng [pdf]

[11] Regularization of Distinct Strategies for Unsupervised Question Generation Findings of EMNLP 2020
Junmo Kang*, Giwon Hong*, Haritz Puerto San Roman*, Sung-Hyon Myaeng [pdf]

[12] Handling Anomalies of Synthetic Questions in Unsupervised Question Answering COLING 2020
Giwon Hong*, **Junmo Kang***, Doyeon Lim*, Sung-Hyon Myaeng [pdf]

[13] Let Me Know What to Ask: Interrogative-Word-Aware Question Generation MRQA@EMNLP 2019
Junmo Kang*, Haritz Puerto San Roman*, Sung-Hyon Myaeng [pdf]

EXPERIENCE

MIT-IBM Watson AI Lab *Cambridge, MA, USA*
Research Intern (Host: Leonid Karlinsky, Rogerio Feris) May. 2023 - Aug. 2023
- Self-alignment of large language models for specialization [1].

Georgia Tech *Atlanta, GA, USA*
Graduate Research Assistant (Advisor: Alan Ritter, Wei Xu) Aug. 2022 - Present
- Instructing LLMs with schema for efficient and robust table extraction [2].
- Cost-efficiency of annotation and distillation [4].

KAIST IR&NLP Lab *Daejeon, Republic of Korea*
Research Associate (Host: Sung-Hyon Myaeng) Mar. 2021 - Jul. 2022
- Instruction-based learning for robust retrieval-augmented LLMs [3].
- Efficient methods for multi-hop QA [5].

KAIST *Daejeon, Republic of Korea*
Graduate Research Assistant (Advisor: Sung-Hyon Myaeng) Feb. 2019 - Feb. 2021
- Question generation and unsupervised question answering for data-efficiency [11,12,13].
- Sample-efficient and robust representations for numerical reasoning [9,6].
- Generative model that leverages inter-dependency of tags while alleviating the order sensitivity [8].
- Sparse representations for passage retrieval that can take advantage of an efficient inverted index and symbolic IR techniques [7].

HONORS & AWARDS

Microsoft Accelerate Foundation Models Research [link]	2023
Graduated with Highest Honor in CSE, Chungnam National University	2019
Grand Prize, Business ICT Competition	2018
Excellence Award, Startup Competition	2018
NAVER AI Hackathon Finalist	2018
Grand Prize, Daejeon Startup School	2017
Best Excellence Award, Startup Picnic	2016
Microsoft Imagine Cup Korea Finalist	2016