(Updated July 5, 2024)

SUKMIN CHO

Republic of Korea
nelllpic@kaist.ac.kr

RESEARCH STATEMENT

My research focuses on the interaction and the communication with **Large Language Models** (LLMs) based on the **Prompting**, such as understanding how different input prompts affect LLMs and finding the optimal prompts eliciting their potential, usually in **Question-Answering** (QA), **Information Retrieval** (IR) and **Retrieval-Augmented** (RAG) tasks. Additionally, I am passionate about applying NLP techniques to **real-world challenges**, such as combating trolls on social media, supporting mental illness diagnosis, and translating sign language to connect the deaf and non-deaf communities.

EDUCATION

Korea Advanced Institute of Technology and Science (KAIST) Daejeon, Korea
Ph.D. in School of Computing *September 2022 - Present*

Korea Advanced Institute of Technology and Science (KAIST) Daejeon, Korea
M.S. in School of Computing *March 2021 - August 2022*
Thesis: Template-based document labeling for dense retrieval

Korea Advanced Institute of Technology and Science (KAIST) Daejeon, Korea
B.S. in School of Computing *March 2016 - February 2021*
Minor in Mathematics

EMPLOYMENT

Graduate student, KAIST *March 2021 - Present*
Advisor: Prof. Jong C. Park

- Conducting research on the acceleration of Large Language Model inference
- Servicing as project manager of the national project, "Prediction and augmentation of the credibility distribution via linguistic analysis and automated evidence document collection", starting in 2022.
- Conducted research on the robustness of Retrieval-Augmented Generation System
- Conducted research on adaptation of LLMs on open-domain question-answering tasks

NAND Quality Assessment Intern, SK Hynix *January - February 2019*

- Developed visualization system for NAND testing.

INTERNATIONAL PUBLICATIONS

C16 Database-Augmented Query Representation for Information Retrieval
Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang, and Jong C. Park
Under review

C15 Selective Knowledge Distillation: Enhancing the Quality of Explanatory Rationale on Mental Health Analysis
Hoyun Song, Huije Lee, Jisu Shin, Sukmin Cho, Jong C. Park
Under review

C14 Typos that Broke the RAG's Back: Genetic Attack on RAG Pipeline by Simulating Documents in the Wild via Low-level Perturbations
Sukmin Cho, Soyeong Jeong, Jeongyeon Seo, Taeho Hwang, and Jong C. Park
Under review

C13 Towards Effective Counter-Responses: Aligning Human Preferences with Strategies to Combat Online Trolling
Huije Lee, Hoyun Song, Jisu Shin, Sukmin Cho, SeungYoon Han, and Jong C. Park
Under review

C12 Universal Gloss-level Representation for Gloss-free Sign Language Translation and Production
Eui Jun Hwang, Sukmin Cho, Huije Lee, Youngwoo Yoon, and Jong C Park
Under review

C11 DSLR: Document Refinement with Sentence-Level Re-ranking and Reconstruction to Enhance Retrieval-Augmented Generation
Taeho Hwang, Soyeong Jeong, Sukmin Cho, SeungYoon Han, and Jong C. Park
Knowledge Augmented Methods for NLP Workshop **(KnowledgeNLP@ACL)**, 2024.

C10 Preprocessing Mediapipe Keypoints with Keypoint Reconstruction and Anchors for Isolated Sign Language Recognition
Kyunggen Roh, Huije Lee, Eui Jun Hwang, Sukmin Cho, and Jong C. Park
Representation and Processing of Sign Languages: Evaluation of Sign Language Resources **(RPSL@LREC-COLING)**, 2024.

C9 Adaptive-RAG: Learning to Adapt Retrieval-Augmented Large Language Models through Question Complexity
Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang, and Jong C. Park
North American Chapter of the Association for Computational Linguistics **(NAACL)**, 2024.

C8 Self-Knowledge Distillation for Learning Ambiguity
Hancheol Park, Soyeong Jeong, Sukmin Cho, and Jong C. Park
arXiv preprint, 2024.

C7 Improving Zero-shot Reader by Reducing Distractions from Irrelevant Documents in Open-Domain Question Answering
Sukmin Cho, Jeongyeon Seo, Soyeong Jeong and Jong C. Park
Findings of Empirical Methods in Natural Language Processing **(Findings of EMNLP)**, 2023.

C6 Test-Time Self-Adaptive Small Language Models for Question Answering
Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang and Jong C. Park
Findings of Empirical Methods in Natural Language Processing **(Findings of EMNLP)**, 2023.

C5 Discrete Prompt Optimization via Constrained Generation for Zero-shot Re-ranker
Sukmin Cho, Soyeong Jeong, Jeongyeon Seo and Jong C. Park
Findings of Association for Computational Linguistics **(Findings of ACL)**, 2023..

C4 Sign language production with avatar layering: A critical use case over rare words
Jung-Ho Kim, Eui Jun Hwang, Sukmin Cho, Du Hui Lee and Jong C. Park
International Conference on Language Resources and Evaluation **(LREC)**, 2022.

C3 Query generation with external knowledge for dense retrieval
Sukmin Cho, Soyeong Jeong, Wonsuk Yang and Jong C. Park
Deep Learning Inside Out **(DeeLIO@ACL)**, 2022.

C2 Augmenting Document Representations for Dense Retrieval with Interpolation and Perturbation
Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang and Jong C. Park
Annual Meeting of the Association for Computational Linguistics **(ACL)**, 2022. **(Oral)**

C1 Non-Autoregressive Sign Language Production via Knowledge Distillation
Eui Jun Hwang, Jung-Ho Kim, Sukmin Cho, and Jong C. Park
Arxiv Preprint, 2022.

ACADEMIC SERVICE

Reviewer of **ACL ARR 2024** February, April, June Reviewer 2024
Reviewer of **ACL ARR 2023** December Reviewer 2023
Reviewer of **IEEE Access** 2023

AWARD

Best Paper Award at Korea Computer Congress (KCC) 2024
Best Paper Award at Annual Conference on Human & Cognitive Language Technology (HCLT) 2023
Best Presentation Award at Korea Computer Congress (KCC) 2022

SKILLS

Language: Korean (mother tongue), English (fluent)
Programming: Python, C, Java