Giwon **Hong**

PhD student in Informatics (ILCC) at the University of Edinburgh

giwon.hong@ed.ac.uk Google scholar Homepage LinkedIn

RESEARCH INTERESTS

Natural Language Processing

In-context Learning, Knowledge Conflicts, Hallucinations, Question Answering, Information Retrieval, Graph QA, Data Scarcity, Interpretability & Explainability

EDUCATION

The University of Edinburgh — Edinburgh, UK
PhD student in ILCC program — Sep. 2023 - Present

- Supervisor: Pasquale Minervini (Principal), Edoardo Ponti

Korea Advanced Institute of Science and Technology (KAIST) — Daejeon, Korea
M.S. in School of Computing — Feb. 2018 - Feb. 2020

- Thesis committee: Sung-Hyong Myaeng, Alice Oh, Meeyoung Cha
- GPA: 3.98 / 4.30 (96.44%)

Sungkyunkwan University (SKKU) — Suwon, Korea
B.S. in Computer Science and Engineering — Mar. 2014 - Feb. 2018

- GPA: 4.00 / 4.50 (94.3%)
- Major GPA : 4.31 / 4.5 (97.72%)

PUBLICATIONS

* indicates equal contribution.

[1] **Theorem Prover as a Judge for Synthetic Data Generation** — arXiv Preprint 2025
Joshua Ong Jun Leang, **Giwon Hong**, Wenda Li, and Shay B Cohen [pdf]

[2] **Mixtures of In-Context Learners** — arXiv Preprint 2024
Giwon Hong, Emile van Krieken, Edoardo Ponti, Nikolay Malkin, Pasquale Minervini [pdf]

[3] **Steering Knowledge Selection Behaviours in LLMs via SAE-Based Representation Engineering** — NAACL 2025
Yu Zhao, Alessio Devoto, **Giwon Hong**, and 5 more authors [pdf]

[4] **Are We Done with MMLU?** — NAACL 2025
Aryo Pradipta Gema, Joshua Ong Jun Leang, **Giwon Hong**, and 13 more authors [pdf]

[5] **The Hallucinations Leaderboard – An Open Effort to Measure Hallucinations in Large Language Models** — arXiv Preprint 2024
Giwon Hong*, Aryo Pradipta Gema*, Rohit Saxena*, and 8 more authors [pdf]

[6] **Edinburgh Clinical NLP at SemEval-2024 Task 2: Fine-tune your model unless you have access to GPT-4** — SemEval-2024
Aryo Pradipta Gema*, **Giwon Hong***, Pasquale Minervini, and Luke Daines, Beatrice Alex [pdf]

[7] **Why So Gullible? Enhancing the Robustness of Retrieval-Augmented Models against Counterfactual Noise** Findings of NAACL 2024
Giwon Hong*, Jeonghwan Kim*, Junmo Kang*, and Sung-Hyon Myaeng, Joyce Jiyoung Whang [pdf]

[8] **FinePrompt: Unveiling the Role of Finetuned Inductive Bias on Compositional Reasoning in GPT-4** Findings of EMNLP 2023
Jeonghwan Kim*, **Giwon Hong***, Sung-Hyon Myaeng, and Joyce Jiyoung Whang [pdf]

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

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

[11] **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]

[12] **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]

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

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

[15] Book chapter **”Finding Datasets in Publications: The KAIST Approach”** Sage London, 2020
In Rich Search and Discovery for Research Datasets
Haritz Puerto-San-Roman, **Giwon Hong**, Minh-Son Cao, Sung-Hyon Myaeng [Link]

[16] **Aligning Open IE Relations and KB Relations using a Siamese Network Based on Word Embedding** IWCS, 2019
Rifki Afina Putri, **Giwon Hong**, Sung-Hyon Myaeng [pdf]

EXPERIENCES

KAIST IR&NLP Lab July 2020 - July 2023

Technical Research Personnel

- Alternative to mandatory military service (∼2023.07.08).
- Working on Question Answering (with Data scarcity, Numbers, and Graphs), Neural IR.
- Person in charge of the Exobrain project, detailed task 1 (KAIST).

KAIST IR&NLP Lab Mar. 2020 - June 2020

Research Associate

Samsung SDS Senior Data Scientist Course Feb. 2020 - June 2020

Teaching Assistant

- Class for data processing, analysis, and machine learning (ML) related applications.
- Advising course projects about data analysis and ML techniques.

Korea Advanced Institute of Science and Technology (KAIST) Mar. 2019 - Dec. 2019

Teaching Assistant

- Teaching assistant for the Text Mining course from probabilistic (e.g., CRF, LDA)to neural-based (e.g., CNN, RNN, LSTM) approaches (2019 1st semester)
- Teaching assistant for the Information Retrieval course (e.g., BM25, PRF, L2R) (2019 2nd semester)

PROJECTS

Development of AI Technology to Support Expert Decision-making that can Explain the Reasons/Grounds for Judgement Results Based on Expert Knowledge Apr. 2022 - July 2023

Funded by Korean Government (Ministry of Science and ICT)
Hosted by Electronics and Telecommunications Research Institute (ETRI)

- Working on a neuro-symbolic (semi-parametric, KB-based) dynamic learning technology that can effectively model an environment in which knowledge continuously changes.

Exobrain Mar. 2018 - Mar. 2023

Funded by Korean Government (Ministry of Science and ICT)
Hosted by Electronics and Telecommunications Research Institute (ETRI)

- The purpose of the research is to provide an **expert-level question answering** service in an environment of the knowledge industry such as law, patents, etc.
- **Participant of Detailed task 3** (2018.03-2019.06)
- **Project manager of Detailed task 3** (2019.06-2019.12)
- **Project manager of Detailed task 1 (KAIST)** (2020.01-Present)
- Researched on extracting KB relations constituting triples for a graph-based QA model [16].
- Lead researcher for an ensemble model that combines the graph-based QA model and reading comprehension QA model (1st rank in the leaderboard of TriviaQA Wikipedia at the date of 08/10/19).
- Worked on solving the anomalies of synthetic questions through inverse BLEU-based paraphrasing and confidence score-based filtering [13].
- Presented a sample-efficient and robust number representation in extrapolation for numerical question answering [10, 11].
- Suggested a method for injecting structural information into the Transformer architecture[9].

Deep Matching for Efficient Search Mar. 2020 - June 2020

Funded by NAVER Corp.

- **Participant**
- Proposed a novel, efficient and explainable passage retrieval system based on binarized sparse representations that can utilize an inverted index and symbolic techniques [12].

Machine learning for context association and smart interaction suggestion June 2018 - May 2019

Funded by Korean Government (the Ministry of Science and ICT)

- **Participant**
- Proposed a framework to improve unsupervised question answering by combining different strategies of question generation[14].

HONORS & AWARDS

Rich Context Competition Feb. 15, 2019

Honorable mention (2nd Place)

- By the Coleridge Initiative at New York University.
- The Rich Context Competition was run by the Coleridge Initiative at New York University and aimed to extract dataset mentions from science publications.
- Finalist (Top 4) in phase 1
- 2nd place in phase 2 ($2,000)
- Proposed a system to retrieve datasets from papers based on a RCQA model and a question generation. [15].

Scholarship (SKKU) 2014 - 2018

- Jang Young-sil Scholarship (2014 - 2017)
- Academic excellence A (2017 - 2018)

SKILLS

Programming Languages

- Python, C/C++, Java, Javascript

Frameworks & Tools

- PyTorch, PyTorch Lightning, Huggingface, Docker, Codalab, Tensorflow, DGL (Deep Graph Library), NLP Toolkit (SpaCy, NLTK), KBs (Freebase, Wikidata)

English

- TOEFL (iBT): Total 107| Reading 30| Listening 30| Speaking 23| Writing 24

SERVICES

Review Committee

- 2022: **EMNLP**
- 2023: **ACL, EMNLP, ARR (Oct.), ARR (Dec.)**
- 2024: **COLM**
- 2025: **ICLR, COLM**