

# SOHEE YANG

Ph.D. Student/Research Scientist Intern @ UCL/DeepMind ◊ She/Her/Hers

London, England, United Kingdom ◊ Last Update: 24 May 2023

[soheeyang.cs@gmail.com](mailto:soheeyang.cs@gmail.com) ◊ [github.com/soheeyang](https://github.com/soheeyang) ◊ [soheeyang.github.io](https://soheeyang.github.io) ◊ [@soheeyang\\_](https://twitter.com/soheeyang_)

## EDUCATION

---

- Ph.D. in Computer Science** Mar 2023 - Present  
@ **UCL NLP**, Department of Computer Science, University College London  
& Language Team, **DeepMind** (50/50 Time Split)  
Advisors: **Pontus Stenetorp** and **Sebastian Riedel**
- M.S. in Artificial Intelligence** Mar 2021 - Feb 2023  
@ **Language & Knowledge Lab**, Kim Jaechul Graduate School of AI, KAIST  
Advisor: **Minjoon Seo**
- B.S. in Computer Science and Engineering (*Summa Cum Laude*)** Mar 2014 - Feb 2018  
@ School of CSEE, Handong Global University  
GPA: 4.45 / 4.5 (1st rank in CSEE)

## EMPLOYMENT

---

- Language Team, DeepMind** Mar 2023 - Present  
*Part-Time Research Scientist Intern during Ph.D. (Hosts: Sebastian Riedel and Elena Gribovskaya)*
- Interpretability and Controllability of LLMs
- UCL NLP, University College London** Mar 2023 - Present  
*Research Assistant (Supervisors: Pontus Stenetorp and Sebastian Riedel)*
- Interpretability and Controllability of LLMs
- Language & Knowledge Lab, KAIST** Mar 2021 - Feb 2023  
*Research Assistant (Supervisor: Minjoon Seo)*
- Prompt-based Learning of LLMs, Continual Learning of LLMs, Information Retrieval
- Clova AI, NAVER Corp.** Sep 2018 - Feb 2021  
*Research Engineer*
- Text Parser of OCR @ Parsing Team [[Demo](#)] Oct 2019 - Feb 2021
  - NLU of AiCall (Dialogue System) @ DUET TF [[Demo](#)] [[JP Video](#)] [[KR Video](#)] Sep 2018 - Oct 2019
- AI Team, Kakao Corp.** Jan 2018 - Jun 2018  
*Research Intern*
- NLU and Synthetic Data Generator for Dialogue System @ Context Part
- Deep Learning Lab, Handong Global University** Jun 2016 - Dec 2017  
*Undergraduate Researcher (Supervisor: In-Jung Kim)*
- Deep Learning-based Dialogue State Tracker and Intent Classifier for Dialogue Systems
  - Smart T-Commerce Service Using Big Data and Deep Learning [[Paper](#)] [[Patent](#)]

## PUBLICATIONS

---

\* indicates equal contribution.

### Preprints

- [1] **Sohee Yang**, Jonghyeon Kim, Joel Jang, Seonghyeon Ye, Hyunji Lee, Minjoon Seo. **Improving Probability-based Prompt Selection Through Unified Evaluation and Analysis**. *arXiv preprint*, May 2023. [[Paper](#)] [[Code](#)] [[Tweet](#)]
- [2] Seonghyeon Ye, Hyeonbin Hwang, **Sohee Yang**, Hyeongu Yun, Yireun Kim, Minjoon Seo. **In-Context Instruction Learning**. *arXiv preprint*, Feb 2023. [[Paper](#)] [[Code](#)] [[Tweet](#)]

### Peer-Reviewed Conference/Journal Papers

- [1] Joel Jang, Dongkeun Yoon, **Sohee Yang**, Sungmin Cha, Moontae Lee, Lajanugen Logeswaran, Minjoon Seo. **Knowledge Unlearning for Mitigating Privacy Risks in Language Models**. In *ACL*, 2023. [[Paper](#)] [[Code](#)] [[Tweet](#)]
- [2] Hyunji Lee, Jaeyoung Kim, Hoyeon Chang, Hanseok Oh, **Sohee Yang**, Vlad Karpukhin, Yi Lu, Minjoon Seo. **Contextualized Generative Retrieval**. In *Findings of ACL*, 2023. [[Paper](#)] [[Code](#)] [[Tweet](#)]
- [3] Hyunji Lee, **Sohee Yang**, Hanseok Oh, Minjoon Seo. **Generative Multi-hop Retrieval**. In *EMNLP*, 2022. [[Paper](#)] [[Code](#)] [[Tweet](#)]
- [4] Joel Jang\*, Seonghyeon Ye\*, Changho Lee, **Sohee Yang**, Joongbo Shin, Janghoon Han, Gyeonghun Kim, Minjoon Seo. **TemporalWiki: A Lifelong Benchmark for Training and Evaluating Ever-Evolving Language Models**. In *EMNLP*, 2022. [[Paper](#)] [[Code](#)] [[Tweet](#)]
- [5] Joel Jang, Seonghyeon Ye, **Sohee Yang**, Joongbo Shin, Janghoon Han, Gyeonghun Kim, Stanley Jungkyu Choi, Minjoon Seo. **Towards Continual Knowledge Learning of Language Models**. In *ICLR*, 2022. [[Paper](#)] [[Code](#)] [[Tweet](#)]
- [6] Wonseok Hwang, Jinyeong Yim, Seunghyun Park, **Sohee Yang**, Minjoon Seo. **Spatial Dependency Parsing for Semi-Structured Document Information Extraction**. In *Findings of ACL*, 2021. [[Paper](#)]
- [7] Sewon Min et al. **NeurIPS 2020 EfficientQA Competition: Systems, Analyses and Lessons Learned**. In *PMLR*, 2021. [[Paper](#)] [[Project](#)]
- [8] **Sohee Yang** and Minjoon Seo. **Designing a Minimal Retrieve-and-Read System for Open-Domain Question Answering**. In *NAACL*, 2021. [[Paper](#)] [[Code](#)] [[Poster](#)] [[Slides](#)] [[Video](#)]
- [9] **Sohee Yang** and Minjoon Seo. **Is Retriever Merely an Approximator of Reader?**. In *Spa-NLP Workshop at ACL 2022*, 2020. [[Paper](#)] [[Poster](#)] [[Slides](#)] [[Video](#)] [[Model](#)]
- [10] Jung-Woo Ha\*, Kihyun Nam\*, Jingu Kang, Sang-Woo Lee, **Sohee Yang**, Hyunhoon Jung, Eunmi Kim, Hyeji Kim, Soojin Kim, Hyun Ah Kim, Kyoungtae Doh, Chan Kyu Lee, Nako Sung, Sunghun Kim. **ClovaCall: Korean Goal-Oriented Dialog Speech Corpus for Automatic Speech Recognition of Contact Centers**. In *INTERSPEECH*, 2020. [[Paper](#)] [[Code](#)]
- [11] Sungdong Kim, **Sohee Yang**, Gyuwan Kim, Sang-Woo Lee. **Efficient Dialogue State Tracking by Selectively Overwriting Memory**. In *ACL*, 2020. [[Paper](#)] [[Code](#)]
- [12] Sang-Woo Lee, Tong Gao, **Sohee Yang**, Jaejun Yoo, Jung-Woo Ha. **Large-Scale Answerer in Questioner’s Mind for Visual Dialog Question Generation**. In *ICLR*, 2019. [[Paper](#)] [[Code](#)] [[Poster](#)] [[Slides](#)]

## HONORS, AWARDS & SCHOLARSHIPS

---

### Honors & Awards

- 1st Place, **EfficientQA Competition** 500MB Track Human Eval, NeurIPS 2020 Dec 2020
- Gyeongsangbuk-do Province Governor's Commendation, HGU Feb 2018
- *Summa Cum Laude*, HGU Feb 2018
- 1st Place (Grand Prize), Senior Capstone Design Competition, HGU Dec 2017

### Scholarships

- Full tuition funded by DeepMind at UCL Mar 2023 - Present
- 92% of tuition covered by the school at KAIST Mar 2021 - Feb 2023
- Full tuition covered by the school and Korean Government at HGU (merit-based) Mar 2014 - Feb 2018
  - National Science and Engineering Scholarship (Government of South Korea)
  - Merit Scholarship for Academic Excellence (top 1% for 4 times)
  - Entrance Scholarship (top 5%)

## TALKS

---

### **EfficientQA & Designing a Minimal Retrieve-and-Read System for ODQA** Dec 2020

- Tech Talk, Upstage AI, Online

### **DUET (AiCall) and Potential Needs in Dialog** May 2019

- NAVER LABS Europe × Clova AI Workshop, NAVER LABS Europe, Grenoble, France

### **Large-Scale Answerer in Questioner's Mind for Visual Dialog Question Generation** May 2019

- Naver Expo Lunch Talk, ICLR 2019, New Orleans, United States

### **Deep Learning-based Dialogue System** Aug 2018

- Summer School, Deep Learning Society: Deepest, Seoul, South Korea

## SERVICES

---

### **Reviewer**

- ICLR 2023, ACL Workshop 2022

### **Secondary Reviewer**

- ICML 2019, ICLR 2019

## STUDENT ASSOCIATIONS

---

### **Deep Learning Society: Deepest** Jul 2018 - Feb 2021 *Seoul National University*

- Director Feb 2019 - Jun 2019

### **S/W Development Club: Seulgijjagi** Feb 2014 - Feb 2018 *Handong Global University*

- President Feb 2016 - Jun 2016
- President Feb 2015 - Jun 2015

## TEACHING EXPERIENCE

---

### Teaching Assistant

*Handong Global University*

- C Programming Aug 2017 - Dec 2017
- Operating Systems Feb 2017 - Jun 2017
- Computer Architecture and Organization Aug 2016 - Dec 2016
- Logic Design Feb 2016 - Jun 2016
- C Programming Dec 2015 - Jan 2016

## LANGUAGE PROFICIENCY

---

- Korean - Native Level Fluency
- English - High Level Fluency
  - IELTS Academic Overall 8.0 (taken on 31 Dec 2022)
  - TOEFL iBT 109 (taken on 02 Sep 2018)
  - GRE Writing 4.5 (taken on 19 Nov 2017)