SANG-WOO LEE

PERSONAL INFORMATION

- Independent Researcher in London
- E-mail: sangwoolee.cs@gmail.com
- Webpage: https://sang-woo-lee.com
- Google Scholar: https://scholar.google.co.kr/citations?user=TMTTMuQAAAAJ

RESEARCH INTERESTS

Large Language Model, Natural Language Processing

EDUCATION

Ph.D. in Computer Science and Engineering Mar 2012 - Aug 2018
Seoul National University, Seoul, Republic of Korea

B.S. in Computer Science and Engineering Mar 2008 - Feb 2012
Seoul National University, Seoul, Republic of Korea

SCHOLARSHIPS

Presidential Science Scholarship, Korea Student Aid Foundation 2008 - 2012

WORK EXPERIENCE

Program Leader and Executive Director, NAVER Cloud Jan 2023 - Jan 2025

- Initiated and led modeling teams, overseeing 40 to 100 team members. Enhanced the company's Large Language Model, **HyperCLOVA X**, and developed Korean AI agents, including **CLOVA X**, based on the model.
- Directed key initiatives in alignment learning, visual multimodal learning, search-based LLM applications, data construction, and model evaluation.
- Supported AI models to drive revenue, including securing a 100 billion won contract and generating additional revenue streams.
- Promoted to Executive Director in April 2024, leading high-impact strategic initiatives.
- Initiated and served on NAVER's GPU resources board for training, overseeing GPU allocation and managing the operational needs of AI models across the company.

Adjunct Professor, KAIST AI Oct 2021 – Sep 2024

- Facilitated and promoted research collaboration between KAIST and NAVER, driving innovation and academic partnerships. Organized seminars for the AI599 class.

Technical Leader, NAVER May 2020 - Dec 2022

- Provided technical leadership for a 30-member modeling team, specializing in conversational AI systems.
- Implemented large-scale language models (**HyperCLOVA**) tailored for Korean and Japanese markets.
- Developed AI-driven phone-call-based systems, including **CareCall** (a senior-focused chatbot), **AiCall** (a reservation chatbot), and other conversational AI solutions.

Research Scientist and Manager, NAVER and NAVER Cloud July 2018 – Mar 2023

· Conducted research focused on dialog systems and large-scale language models.
· Led the NLP team at **NAVER AI Lab** from April 2021, driving foundational research in NLP.
· Published up to 20 research papers annually at the peak of productivity.
· Supported research and modeling efforts for NLP application teams across NAVER, enhancing core AI capabilities.

Project Initiator of AiCall, NAVER July 2018 – Apr 2020

· Launched and led the development of **AiCall**, a phone-call-based reservation chatbot, as the first team member of the project.
· Managed data engineering efforts, including constructing training datasets and creating evaluation systems, while directing system modeling.
· Successfully launched the system in Japan, generating peak annual revenue of approximately $10 million.
· Contributed to NAVER's strategy by facilitating the merger of **LINE** (NAVER's subsidiary) and **Yahoo Japan** through the success of the AiCall project.

PREPRINT

[1] Sohee Yang, **Sang-Woo Lee**, Nora Kassner, Daniela Gottesman, Sebastian Riedel, Mor Geva. *arXiv*. 2025.

TECHNICAL REPORT

[1] TEAM HyperCLOVA X. HyperCLOVA X Technical Report. *arXiv*. 2024.

CONFERENCE

[1] Jaehyung Kim, Jaehyun Nam, Sangwoo Mo, Jongjin Park, **Sang-Woo Lee**, Minjoon Seo, Jung-Woo Ha, Jinwoo Shin. SuRe: Improving Open-domain Question Answering of LLMs via Summarized Retrieval. *ICLR*. 2024.

[2] Hyuhng Joon Kim, Hyunsoo Cho, **Sang-Woo Lee**, Junyeob Kim, Choonghyun Park, Sang-goo Lee, Kang Min Yoo, Taeuk Kim. Universal Domain Adaptation for Robust Handling of Distributional Shifts in NLP. *Findings of EMNLP*. 2023.

[3] Dongryeol Lee, Segwang Kim, Minwoo Lee, Hwanhee Lee, Joonsuk Park, **Sang-Woo Lee**, Kyomin Jung. Asking Clarification Questions to Handle Ambiguity in Open-Domain QA. *Findings of EMNLP*. 2023.

[4] Deokjae Lee, JunYeong Lee, Jung-Woo Ha, Jin-Hwa Kim, **Sang-Woo Lee**, Hwaran Lee, Hyun Oh Song. Query-Efficient Black-Box Red Teaming via Bayesian Optimization. In *ACL*. 2023.

[5] Kyuyong Shin, Hanock Kwak, Wonjae Kim, Jisu Jeong, Seungjae Jung, Kyung-Min Kim, Jung-Woo Ha, **Sang-Woo Lee**. Pivotal Role of Language Modeling in Recommender Systems: Enriching Task-specific and Task-agnostic Representation Learning. In *ACL*. 2023.

[6] Hyunsoo Cho, Hyuhng Joon Kim, Junyeob Kim, **Sang-Woo Lee**, Sang-goo Lee, Kang Min Yoo, Taeuk Kim. Prompt-augmented Linear Probing: Scaling beyond the Limit of Few-shot In-context Learners. In *AAAI*. 2023.

[7] Jin-Hwa Kim, Yunji Kim, Jiyoung Lee, Kang Min Yoo, **Sang-Woo Lee**. Mutual Information Divergence: A Unified Metric for Multimodal Generative Models. In *NeurIPS*. 2022.

[8] Junyeob Kim*, Kang Min Yoo*†, Hyuhng Joon Kim, Hyunsoo Cho, Hwiyeol Jo, **Sang-Woo Lee**, Sang-goo Lee, Taeuk Kim†. Ground-Truth Labels Matter: A Deeper Look into Input-Label Demonstrations. In *EMNLP*. 2022.

[9] Sanghwan Bae, Donghyun Kwak, Soyoung Kang, Min Young Lee, Sungdong Kim, Yuin Jeong, Hyeri Kim, **Sang-Woo Lee**, Woomyoung Park, Nako Sung. Keep Me Updated! Memory Management in Long-term Conversations. In *Findings of EMNLP*. 2022.

[10] Sanghwan Bae, Donghyun Kwak, Sungdong Kim, Donghoon Ham, Soyoung Kang, **Sang-Woo Lee**, Woomyoung Park. Building a Role Specified Open-Domain Dialogue System Leveraging Large-Scale Language Models. In *NAACL*. 2022.

[11] Seongjin Shin*, **Sang-Woo Lee***, Hwijeen Ahn, Sungdong Kim, HyoungSeok Kim, Boseop Kim, Kyunghyun Cho, Gichang Lee, Woomyoung Park, Jung-Woo Ha, Nako Sung. On the Effect of Pretraining Corpora on In-context Learning by a Large-scale Language Model. In *NAACL*. 2022.

[12] Kyungjae Lee, Wookje Han, Seung-Won Hwang, Hwaran Lee, Joonsuk Park, **Sang-Woo Lee**. Continual Plug-and-Adapt for CuQA. In *Findings of ACL*. 2022.

[13] Yeon Seonwoo, Juhee Son, Jiho Jin, **Sang-Woo Lee**, Ji-Hoon Kim, Jung-Woo Ha, Alice Oh. Two-Step Question Retrieval for Open-Domain QA. In *Findings of ACL*. 2022.

[14] Boseop Kim*, HyoungSeok Kim*, **Sang-Woo Lee***, Gichang Lee, Donghyun Kwak, Dong Hyeon Jeon, Sunghyun Park, Sungju Kim, Seonhoon Kim, Dongpil Seo, Heungsub Lee, Minyoung Jeong, Sungjae Lee, Minsub Kim, Suk Hyun Ko, Seokhun Kim, Taeyong Park, Jinuk Kim, Soyoung Kang, Na-Hyeon Ryu, Kang Min Yoo, Minsuk Chang, Soobin Suh, Sookyo In, Jinseong Park, Kyungduk Kim, Hiun Kim, Jisu Jeong, Yong Goo Yeo, Donghoon Ham, Dongju Park, Min Young Lee, Jaewook Kang, Inho Kang, Jung-Woo Ha, Woomyoung Park, Nako Sung. **What Changes Can Large-scale Language Models Bring? Intensive Study on HyperCLOVA: Billions-scale Korean Generative Pretrained Transformers**. In *EMNLP*. 2021.

[15] Kang Min Yoo, Dongju Park, Jaewook Kang, **Sang-Woo Lee**, Woomyeong Park. **GPT3Mix: Leveraging Large-scale Language Models for Text Augmentation**. In *Findings of EMNLP*. 2021.

[16] Sungdong Kim, Minsuk Chang, **Sang-Woo Lee**. **NeuralWOZ: Learning to Collect Task-Oriented Dialogue via Model-based Simulation**. In *ACL*. 2021.

[17] Yeon Seonwoo, **Sang-Woo Lee**, Ji-Hoon Kim, Jung-Woo Ha, Alice Oh. **Weakly Supervised Pre-Training for Multi-hop Retriever**. In *Findings of ACL*. 2021.

[18] Minjeong Kim, Gyuwan Kim, **Sang-Woo Lee**, Jung-Woo Ha. **ST-BERT: Cross-modal Language Model Pre-training For End-to-end Spoken Language Understanding**. In *ICCASP*. 2021.

[19] Jung-Woo Ha*, Kihyun Nam*, Jin Gu Kang, **Sang-Woo Lee**, Sohee Yang, Hyunhoon Jung, Eunmi Kim, Hyeji Kim, Soojin Kim, Hyun Ah Kim, Kyoungtae Doh, Chan Kyu Lee, Sunghun Kim. **ClovaCall: Korean Goal-oriented Dialog Speech Corpus for Automatic Speech Recognition of Contact Centers**. In *Interspeech*. 2020.

[20] Sungdong Kim, Sohee Yang, Gyuwan Kim, **Sang-Woo Lee**. **Efficient Dialogue State Tracking by Selecitvely Overwriting Memory**. In *ACL*. 2020.

[21] **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.

[22] **Sang-Woo Lee**, Yujung Heo, Byoung-Tak Zhang. **Answerer in Questioner's Mind: Information Theoretic Approach to Goal-oriented Visual Dialog**. In *NeurIPS*. 2018. (Spotlight)

[23] **Sang-Woo Lee**, Jin-Hwa Kim, Jaehyun Jun, Jung-Woo Ha, Byoung-Tak Zhang. **Overcoming Catastrophic Forgetting by Incremental Moment Matching**. In *NIPS*. 2017. (Spotlight)

[24] Jin-Hwa Kim, **Sang-Woo Lee**, Dong-Hyun Kwak, Min-Oh Heo, Jeonghee Kim, Jung-Woo Ha, Byoung-Tak Zhang. **Multimodal Residual Learning for Visual QA**. In *NIPS*. 2016.

[25] **Sang-Woo Lee**, Chung-Yeon Lee, Dong-Hyun Kwak, Jiwon Kim, Jeonghee Kim, and Byoung-Tak Zhang. **Dual Memory Architectures for Lifelong Learning of Everyday Human Behaviors**. In *IJCAI*. 2016.

JOURNAL

[1] Christina Baek, **Sang-Woo Lee**, Beom-Jin Lee, Dong-Hyun Kwak, Byoung-Tak Zhang. **Enzymatic Weight Update Algorithm for DNA-based Molecular Learning**. *Molecules*, 24(7):1409. 2019.

[2] **Sang-Woo Lee**, Chung-Yeon Lee, Dong-Hyun Kwak, Jung-Woo Ha, Jeonghee Kim, and Byoung-Tak Zhang. **Dual-memory Neural Networks for Modeling Cognitive Activities of Humans via Wearable Sensors**. *Neural Networks*, 92:17-28. 2017.

Last Updated on Jun 14, 2025