Seanie Lee

Date of Birth: 1992.04.17, Nationality: Korean

126 Yangjae-dong, Seocho District, Seoul

☎ (+82) 10-4475-2273 | ✉ lsnfamily02@kaist.ac.kr | ⌂ seanie12.github.io | seanie12 | 🎓 Seanie Lee

Education

KAIST (Korea Advanced Institute of Science and Technology)

Daejeon, S.Korea

Ph.D in Artificial Intelligence *Mar. 2022 - present*

- Supervised by Sung Ju Hwang and Juho Lee
- Research interest: Safety Alignment.

KAIST (Korea Advanced Institute of Science and Technology)

Daejeon, S.Korea

M.S. in Artificial Intelligence *Mar. 2020 - Feb. 2022*

- Supervised by Sung Ju Hwang and Juho Lee
- Master Thesis: Data augmentation for natural language processing

Yonsei University

Seoul, S.Korea

B.A. in Library and Information Science *Mar. 2011 - Feb. 2018*

Experience

Apple

Seattle, US

Internship *October 2025 - May 2026*

- Machine Learning Research, hosted by Raviteja Vemulapalli.

Krafton

Seoul, Korea

Internship *July 2025 - Oct 2025*

- Research Internship

Mila

Montreal, Canada

Internship *January 2024 - June 2024*

- Research internship at Mila, advised by Yoshua Bengio.

Apple

Cambridge, UK

Internship *May 2023 - September 2023*

- Research internship at Siri team, hosted by Anders Johannsen.

Singapore National University

Singapore

Internship *July 2022 - September 2022*

- Remote internship at Deep Learning lab, supervised by Kenji Kawaguchi.

Korea Advanced Institute of Science and Technology

Daejeon, S.Korea

Teaching Assistant *Mar. 2020 - Dec. 2021*

- Deep Reinforcement Learning, AI611
- Mathematics for AI, AI503
- Deep Learning, AI502

42 Maru

Seoul, S.Korea

Internship *Feb. 2019 - Jan. 2020*

- Research on Question Answering, Semi-supervised Learning, Domain Generalization

Awards

2023	**Apple AI/ML PhD Fellowship**, Recipient of Apple Scholars in AI/ML	*Cupertino, US*
2022	**Google Travel Grant**, NeurIPS 2022	*US*
2019	**Silver Medal**, Named Entity Recognition in NAVER NLP Challenge	*Seoul, Korea*

Presentation

Seminar at Korea University. *Seoul, Korea*

PRESENTATION OF LARGE SCALE SET-ENCODING *May. 2025*

- Synthetic Data Generation for LLM Safeguards
- ICLR 2025, ACL Findings 2025

Seminar at Hanyang University. *Seoul, Korea*

PRESENTATION OF LARGE SCALE SET-ENCODING *April. 2025*

- Synthetic Data Generation for LLM Safeguards
- ICLR 2025, ACL Findings 2025

Tech. Talk, Nuremberg Institute of Technology Georg Simon Ohm. *Nürnberg, Germany*

PRESENTATION OF LARGE SCALE SET-ENCODING *Oct. 2023*

- Scalable Set Encoding with Universal Mini-Batch Consistency and Unbiased Full Set Gradient Approximation
- ICML 2023

Tech. talk, Samgsung SDS. *Seoul, South Korea*

PRESENTATION OF LARGE SCALE SET-ENCODING *22.May. 2023*

- Scalable Set Encoding with Universal Mini-Batch Consistency and Unbiased Full Set Gradient Approximation
- ICML 2023

Tech. talk, NAVER corp. *Online, South Korea*

PRESENTATION OF INFO-HCVAE *04.Dec. 2020*

- Generating Diverse and Consistent QA pairs from Contexts with Information-Maximizing Hierarchical Conditional VAEs
- ACL 2020 Long paper

Publication

(* indicates equal contribution)

PREPRINT

HoliSafe: Holistic Safety Benchmarking and Modeling with Safety Meta Token for Vision-Language Model *Arxiv*

YOUNGWAN LEE, KANGSAN KIM, KWANYONG PARK, ILCAHE JUNG, SOOJIN JANG, **SEANIE LEE**, YONG-JU LEE AND SUNG JU HWANG *2025*

- [paper][code]

CONFERENCES

FedSVD: Adaptive Orthogonalization for Private Federated Learning with LoRA *NeurIPS*

SEANIE LEE*, SANGWOO PARK*, DONG BOK LEE*, DOMINIK WAGNER, HAEBIN SEONG, TOBIAS BOCKLET, JUHO LEE, SUNG JU HWANG *2025*

- [paper][code]

Distilling LLM Agent into Small Models with Retrieval and Code Tools *NeurIPS Spotlight*

MINKI KANG, JONGWON JEONG, **SEANIE LEE**, JAEWOONG CHO AND SUNG JU HWANG *2025*

- [paper][code]

Reliable Decision-Making via Calibration-Oriented Retrieval-Augmented Generation *NeurIPS*

CHAEYUN JANG, DEUKHWAN CHO, **SEANIE LEE**, JUHO LEE AND HYUNGI LEE *2025*

- [paper]

Trajectory Balance with Asynchrony: Decoupling Exploration and Learning for Fast, Scalable LLM Post-Training *NeurIPS*

BRIAN R. BARTOLDSON, SIDDARTH VENKATRAMAN, JAMES DIFFENDERFER, MOKSH JAIN, TAL BEN-NUN, **SEANIE LEE**, MINSU KIM, JOHAN OBANDO-CERON, YOSHUA BENGIO AND BHAVYA KAILKHURA *2025*

- [paper][code]

SafeRoute: Adaptive Model Selection for Efficient and Accurate Safety Guardrails in Large Language Models *ACL Findings*

SEANIE LEE*, DONG BOK LEE*, DOMINIK WAGNER, MINKI KANG, HAEBIN SEONG, TOBIAS BOCKLET, JUHO LEE, SUNG JU HWANG *2025*

- [paper][code]

Personalized Fine-Tuning with Controllable Synthetic Speech from LLM-Generated Transcripts for Dysarthric Speech Recognition *Interspeech*

Dominik Wagner, Ilja Baumann, Natalie Engert, **Seanie Lee**, Elmar Nöth, Korbinian Riedhammer and Tobias Bocklet *2025*

- [paper]

HarmAug: Effective Data Augmentation for Knowledge Distillation of Safety Guard Models *ICLR*

Seanie Lee*, Haebin Seong*, Dong Bok Lee, Minki Kang, Xiaoyin Chen, Dominik Wagner, Yoshua Bengio, Juho Lee, Sung Ju Hwang *2025*

- [paper][code]

Learning Diverse Attacks on Large Language Models for Robust Red-teaming and Safety Tuning *ICLR*

Seanie Lee, Minsu Kim, Lynn Cherif, David Dobre, Juho Lee, Sung Ju Hwang, Kenji Kawaguchi, Gauthier Gidel, Yoshua Bengio, Nikolay Malkin, Moksh Jain *2025*

- [paper][code]

Optimized Speculative Sampling for GPU Hardware Accelerators *EMNLP*

Dominik Wagner, **Seanie Lee**, Ilja Baumann, Philipp Seeberger, Korbinian Riedhammer, Tobias Bocklet *2024*

- [paper][code]

Drug Discovery with Dynamic Goal-aware Fragment *ICML*

Seul Lee, **Seanie Lee**, Kenji Kawaguchi, Sung Ju Hwang *2024*

- [paper][code]

Effective and Efficient Conversation Retrieval for Dialogue State Tracking with Implicit Text Summaries *NAACL*

Seanie Lee, Jianpeng Cheng, Joris Driesen, Alexandru Coca, Anders Johannsen *2024*

- [paper]

Self-Supervised Dataset Distillation for Transfer Learning *ICLR*

Dong Bok Lee*, **Seanie Lee***, Joonho Ko, Kenji Kawaguchi, Juho Lee, Sung Ju Hwang *2024*

- [paper][code]

DiffusionNAG: Task-guided Neural Architecture Generation with Diffusion Models *ICLR*

Sohyun Ahn*, Hayeon Lee*, Jaehyeong Jo, **Seanie Lee**, Sung Ju Hwang *2024*

- [paper][code]

Scalable Set Encoding with Universal Mini-Batch Consistency and Unbiased Full Set Gradient Approximation *ICML*

Jeffrey Willette*, **Seanie Lee***, Bruno Andreis, Kenji Kawaguchi, Juho Lee, Sung Ju Hwang *2023*

- [paper][code]

Margin-based Neural Network Watermarking *ICML*

Byungjoo Kim, Suyoung Lee, **Seanie Lee**, Sooel Son, Sung Ju Hwang *2023*

- [paper]

Self-Supervised Set Representation Learning for Unsupervised Meta-Learning *ICLR*

Dong Bok Lee*, **Seanie Lee***, Kenji Kawaguchi, Yunji Kim, Jihwan Bang, Jung-Woo Ha, Sung Ju Hwang *2023*

- [paper]

Self-Distillation for Further Pre-training of Transformers *ICLR*

Seanie Lee, Minki Kang, Juho Lee, Sung Ju Hwang, Kenji Kawaguchi *2023*

- [paper][code]

Set-based Meta-Interpolation for Few-Task Meta-Learning *NeurIPS*

Seanie Lee*, Bruno Andreis*, Kenji Kawaguchi, Sung Ju Hwang *2022*

- [paper] [code]

On Divergence Measures for Bayesian Pseudocoresets *NeurIPS*

Balhae Kim, Jungwon Choi, **Seanie Lee**, Yoonho Lee, Jung-Woo Ha, Juho Lee *2022*

- [paper]

Set Based Stochastic Subsampling

ICML

Bruno Andreis, **Seanie Lee**, A. Tuan Nguyen, Juho Lee, Eunho Yang, Sung Ju Hwang

2022

- [paper]

Sequential Reptile: Inter-Task Gradient Alignment for Multilingual Learning

ICLR

Seanie Lee*, Hae Beom Lee*, Juho Lee, Sung Ju Hwang

2022

- [paper]

Learning to Perturb Word Embeddings for Out-of-distribution QA

ACL

Seanie Lee*, Minki Kang*, Juho Lee, Sung Ju Hwang

2021

- [paper][code]

Contrastive Learning with Adversarial Perturbations for Conditional Text Generation

ICLR

Seanie Lee*, Dong Bok Lee*, Sung Ju Hwang

2021

- [paper][code]

Meta-GMVAE: Mixture of Gaussian VAE for Unsupervised Meta-Learning

ICLR

Dong Bok Lee, Dongchan Min, **Seanie Lee**, Sung Ju Hwang

2021

- [paper][code]

Generating Diverse and Consistent QA pairs from Contexts with Information-Maximizing Hierarchical Conditional VAEs

ACL

Dong Bok Lee*, **Seanie Lee***, WooTae Jeong, Donghwan Kim, Sung Ju Hwang

2020

- [paper] [code][video]

g2pM: A Neural Grapheme-to-Phoneme Conversion Package for Mandarin Chinese Based on a New Open Benchmark Dataset

INTERSPEECH

Kyubyong Park*, **Seanie Lee***

2020

- [paper][code]

References

Sung Ju Hwang

Advisor

Associate Professor in KAIST.

2020-2025

e-mail: sjhwang82@kaist.ac.kr.

Juho Lee

Advisor

Associate Professor in KAIST.

2020-2025

e-mail: juholee@kaist.ac.kr.

Yoshua Bengio

Collaborator

Full Professor at Université de Montréal and Scientific Director of Mila – Quebec AI Institute.

2024-2025

e-mail: yoshua.bengio@mila.quebec.

Kenji Kawaguchi

Collaborator

Presidential Young Professor in the Department of Computer Science at NUS.

2022-present

e-mail: kenji@comp.nus.edu.sg

Nikolay Malkin

Collaborator

Chancellor's Fellow at University of Edinburgh, School of Informatics

2024

e-mail: nmalkin@ed.ac.uk