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.
PRESENTATION OF LARGE SCALE SET-ENCODING
- Synthetic Data Generation for LLM Safeguards
- ICLR 2025, ACL Findings 2025

Seoul, Korea
May. 2025

Seminar at Hanyang University.
PRESENTATION OF LARGE SCALE SET-ENCODING
- Synthetic Data Generation for LLM Safeguards
- ICLR 2025, ACL Findings 2025

Seoul, Korea
April. 2025

Tech. Talk, Nuremberg Institute of Technology Georg Simon Ohm.
PRESENTATION OF LARGE SCALE SET-ENCODING
- Scalable Set Encoding with Universal Mini-Batch Consistency and Unbiased Full Set Gradient Approximation
- ICML 2023

Nürnberg, Germany
Oct. 2023

Tech. talk, Samgsung SDS.
PRESENTATION OF LARGE SCALE SET-ENCODING
- Scalable Set Encoding with Universal Mini-Batch Consistency and Unbiased Full Set Gradient Approximation
- ICML 2023

Seoul, South Korea
22.May. 2023

Tech. talk, NAVER corp.
PRESENTATION OF INFO-HCVAE
- Generating Diverse and Consistent QA pairs from Contexts with Information-Maximizing Hierarchical Conditional VAEs
- ACL 2020 Long paper

Online, South Korea
04.Dec. 2020

Publication

(* indicates equal contribution)

PREPRINT

Rethinking Reward Models for Multi-Domain Test-Time Scaling
DONG BOK LEE*, **SEANIE LEE***, SANGWOO PARK, MINKI KANG, JINHEON BAEK, DONGKI KIM, DOMINIK WAGNER, JIONGDAO JIN, HEEJUN LEE, TOBIAS BOCKLET, JINYU WANG, JINGJING FU, SUNG JU HWANG, JIANG BIAN AND LEI SONG
- [paper][code]

Arxiv
2025

HoliSafe: Holistic Safety Benchmarking and Modeling with Safety Meta Token for Vision-Language Model
YOUNGWAN LEE, KANGSAN KIM, KWANYONG PARK, ILCAHE JUNG, SOOJIN JANG, **SEANIE LEE**, YONG-JU LEE AND SUNG JU HWANG
- [paper][code]

Arxiv
2025

CONFERENCES

FedSVD: Adaptive Orthogonalization for Private Federated Learning with LoRA
SEANIE LEE*, SANGWOO PARK*, DONG BOK LEE*, DOMINIK WAGNER, HAEBIN SEONG, TOBIAS BOCKLET, JUHO LEE, SUNG JU HWANG
- [paper][code]

NeurIPS
2025

Distilling LLM Agent into Small Models with Retrieval and Code Tools
MINKI KANG, JONGWON JEONG, **SEANIE LEE**, JAEWOONG CHO AND SUNG JU HWANG
- [paper][code]

NeurIPS Spotlight
2025

Reliable Decision-Making via Calibration-Oriented Retrieval-Augmented Generation
CHAEYUN JANG, DEUKHWAN CHO, **SEANIE LEE**, JUHO LEE AND HYUNGI LEE
- [paper]

NeurIPS
2025

Trajectory Balance with Asynchrony: Decoupling Exploration and Learning for Fast, Scalable LLM Post-Training
BRIAN R. BARTOLDSON, SIDDARTH VENKATRAMAN, JAMES DIFFENDERFER, MOKSH JAIN, TAL BEN-NUN, **SEANIE LEE**, MINSU KIM, JOHAN OBANDO-CERON, YOSHUA BENGIO AND BHAVYA KAILKHURA
- [paper][code]

NeurIPS
2025

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

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

ACL Findings

2025

• [paper][code]

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

DOMINIK WAGNER, ILJA BAUMANN, NATALIE ENGERT, SEANIE LEE, ELMAR NÖTH, KORBINIAN RIEDHAMMER AND TOBIAS BOCKLET

Interspeech

2025

• [paper]

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

SEANIE LEE*, HAEBIN SEONG*, DONG BOK LEE, MINKI KANG, XIAOYIN CHEN, DOMINIK WAGNER, YOSHUA BENGIO, JUHO LEE, SUNG JU HWANG

ICLR

2025

• [paper][code]

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

SEANIE LEE, MINSU KIM, LYNN CHERIF, DAVID DOBRE, JUHO LEE, SUNG JU HWANG, KENJI KAWAGUCHI, GAUTHIER GIDEL, YOSHUA BENGIO, NIKOLAY MALKIN, MOKSH JAIN

ICLR

2025

• [paper][code]

Optimized Speculative Sampling for GPU Hardware Accelerators

DOMINIK WAGNER, SEANIE LEE, ILJA BAUMANN, PHILIPP SEEBERGER, KORBINIAN RIEDHAMMER, TOBIAS BOCKLET

EMNLP

2024

• [paper][code]

Drug Discovery with Dynamic Goal-aware Fragment

SEUL LEE, SEANIE LEE, KENJI KAWAGUCHI, SUNG JU HWANG

ICML

2024

• [paper][code]

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

SEANIE LEE, JIANPENG CHENG, JORIS DRIESEN, ALEXANDRU COCA, ANDERS JOHANNSEN

NAACL

2024

• [paper]

Self-Supervised Dataset Distillation for Transfer Learning

DONG BOK LEE*, SEANIE LEE*, JOONHO KO, KENJI KAWAGUCHI, JUHO LEE, SUNG JU HWANG

ICLR

2024

• [paper][code]

DiffusionNAG: Task-guided Neural Architecture Generation with Diffusion Models

SOHYUN AHN*, HAYEON LEE*, JAEHYEONG JO, SEANIE LEE, SUNG JU HWANG

ICLR

2024

• [paper][code]

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

JEFFREY WILLETTE*, SEANIE LEE*, BRUNO ANDREIS, KENJI KAWAGUCHI, JUHO LEE, SUNG JU HWANG

ICML

2023

• [paper][code]

Margin-based Neural Network Watermarking

BYUNGJOO KIM, SUYOUNG LEE, SEANIE LEE, SOOEL SON, SUNG JU HWANG

ICML

2023

• [paper]

Self-Supervised Set Representation Learning for Unsupervised Meta-Learning

DONG BOK LEE*, SEANIE LEE*, KENJI KAWAGUCHI, YUNJI KIM, JIHWAN BANG, JUNG-WOO HA, SUNG JU HWANG

ICLR

2023

• [paper]

Self-Distillation for Further Pre-training of Transformers

SEANIE LEE, MINKI KANG, JUHO LEE, SUNG JU HWANG, KENJI KAWAGUCHI

ICLR

2023

• [paper][code]

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

SEANIE LEE*, BRUNO ANDREIS*, KENJI KAWAGUCHI, SUNG JU HWANG

NeurIPS

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