

# Hanseul Cho (조한슬)

Room 9410, Building #9; 85 Heogi-ro, Dongdaemun-gu, Seoul, Republic of Korea

[jhs4015@kaist.ac.kr](mailto:jhs4015@kaist.ac.kr)

[hanseuljo.github.io](https://hanseuljo.github.io)

[github.com/HanseulJo](https://github.com/HanseulJo)

[linkedin.com/in/hanseul-cho](https://www.linkedin.com/in/hanseul-cho)

[@hanseuljo](https://twitter.com/hanseuljo)

[Google Scholar \(Hanseul Cho\)](https://scholar.google.com/citations?user=HanseulCho)



## Personal Profile

I am a Ph.D. student at Kim Jaechul Graduate School of AI, Korea Advanced Institute of Science and Technology (KAIST AI), where I am fortunate to be advised by Prof. Chulhee “Charlie” Yun of Optimization & Machine Learning (OptiML) Laboratory, KAIST AI. Also, I’m a student researcher at Google NYC (05/05/2025–08/22/2025), working for Srdinadh Bhojanapalli. Previously, I completed my M.Sc. (in AI) and B.Sc. (in Math, minor in CS, Summa Cum Laude) at KAIST.<sup>1</sup>

My primary research interests lie in optimization, machine learning (ML), and deep learning (DL). During my journey to a Ph.D., my ultimate research goal is to **rigorously understand and practically overcome** the following **three critical challenges** in ML/DL:

### Generalizability 🏗️

**Out-of-distribution generalization of (large) language models.**

(e.g., length generalization and compositional generalization of Transformers)

### Adaptability 🔄

**Training adaptable models under an evolving environment.**

(e.g., continual learning, maintaining the plasticity of neural networks, sample-efficient reinforcement learning)

### Multifacetedness 🧩

**Learning with multiple (possibly conflicting and/or orthogonal) goals.**

(e.g., minimax optimization, bi-level optimization, fairness in ML)

## Publications

### Preprints

[C1] Chang, Hyeon<sup>\*</sup>, Jinho Park<sup>\*</sup>, Hanseul Cho, Sohee Yang, Miyoung Ko, Hyeonbin Hwang, Seungpil Won, Dohaeng Lee, Youbin Ahn, Minjoon Seo. “The Coverage Principle: A Framework for Understanding Compositional Generalization.” *arXiv preprint*. 2025. [\[arXiv\]](#)

### International Conferences

[C8] Cho, Hanseul<sup>\*</sup>, Jaeyoung Cha<sup>\*</sup>, Srinadh Bhojanapalli, Chulhee Yun. “Arithmetic Transformers Can Length-Generalize in Both Operand Length and Count.” **ICLR 2025**. [\[arXiv\]](#) [\[OpenReview\]](#)

[C7] Jung, Hyunji<sup>\*</sup>, Hanseul Cho<sup>\*</sup>, Chulhee Yun. “Convergence and Implicit Bias of Gradient Descent on Continual Linear Classification.” **ICLR 2025**. [\[arXiv\]](#) [\[OpenReview\]](#)

[C6] Cho, Hanseul<sup>\*</sup>, Jaeyoung Cha<sup>\*</sup>, Pranjal Awasthi, Srinadh Bhojanapalli, Anupam Gupta, Chulhee Yun. “Position Coupling: Improving Length Generalization of Arithmetic Transformers Using Task Structure.” **NeurIPS 2024** & Short version in ICML 2024 Workshop on Long-Context Foundation Models (LCFM). [\[arXiv\]](#) [\[OpenReview\]](#)

[C5] Shin, Baekrok<sup>\*</sup>, Junsoo Oh<sup>\*</sup>, Hanseul Cho, Chulhee Yun. “DASH: Warm-Starting Neural Network Training in Stationary Settings without Loss of Plasticity.” **NeurIPS 2024** & Short version in ICML 2024 Workshop on Advancing Neural Network Training (WANT): Computational Efficiency, Scalability, and Resource Optimization. [\[arXiv\]](#) [\[OpenReview\]](#)

[C4] Lee, Jaewook<sup>\*</sup>, Hanseul Cho<sup>\*</sup>, Chulhee Yun. “Fundamental Benefit of Alternating Updates in Minimax Optimization.” **ICML 2024** & Short version in ICLR 2024 Workshop on Bridging the Gap Between Practice and Theory in Deep Learning (BGPT). [\[arXiv\]](#) [\[OpenReview\]](#)  
• **Spotlight at ICML 2024. (Top 3.5%: (144+191) of 9,473 valid submissions)**

[C3] Lee, Junghyun<sup>\*</sup>, Hanseul Cho<sup>\*</sup>, Se-Young Yun, Chulhee Yun. “Fair Streaming Principal Component Analysis: Statistical and Algorithmic Viewpoint.” **NeurIPS 2023**. [\[arXiv\]](#) [\[OpenReview\]](#)

[C2] Lee, Hojoon<sup>\*</sup>, Hanseul Cho<sup>\*</sup>, Hyunseung Kim<sup>\*</sup>, Daehoon Gwak, Joonkee Kim, Jaegul Choo, Se-Young Yun, Chulhee Yun. “PLASTIC: Improving Input and Label Plasticity for Sample Efficient Reinforcement Learning.” **NeurIPS 2023**. [\[arXiv\]](#) [\[OpenReview\]](#)

[C1] Cho, Hanseul and Chulhee Yun. “SGDA with Shuffling: Faster Convergence for Nonconvex-PL Minimax Optimization.” **ICLR 2023**. [\[arXiv\]](#) [\[OpenReview\]](#)

### Notable Domestic Journals/Conferences

• Cho, Hanseul, Baekrok Shin, Chaewon Moon, Sang-Geun Hong, U-Ju Byeon, Jin-Yong Sung, Chulhee Yun. “Deep Model-Based Optimization of Jamming Effectiveness under Aircraft AESA Radar Operational Environment.” *The Journal of Korean Institute of Communications and Information Sciences (J-KICS)*, vol. 50, no. 11, 2025. pp. 00–00.

• Jung, Hyunji<sup>\*</sup>, Hanseul Cho<sup>\*</sup>, Chulhee Yun. “Convergence and Implicit Bias of Gradient Descent on Continual Linear Classification.” *Short version in the 11th Joint Conference of Korean Artificial Intelligence Association (JKAIA 2024)*.  
– **Best Paper Award (Top 3) & Oral presentation.**

• Cho, Hanseul and Chulhee Yun. “SGDA with Shuffling: Faster Convergence for Nonconvex-PL Minimax Optimization.” *Short version in the 7th Joint Conference of Korea Artificial Intelligence Association (JKAIA 2022)*.  
– **NAVER Outstanding Theory Paper Award (Top 3) & Oral presentation.**

<sup>\*</sup>: co-first authors.

<sup>1</sup>You can find the source code of this CV [here](#).

## Education

### Korea Advanced Institute of Science and Technology (KAIST)

Seoul, Republic of Korea

Ph.D. in Artificial Intelligence

Sept. 2023 – Current

- Advisor: Prof. [Chulhee Yun](#) (Optimization & Machine Learning (**OptiML**) Laboratory, Kim Jaechul Graduate School of AI (GSAI), KAIST)
- Anticipated Graduation Date: Aug. 2027

### KAIST

Seoul, Republic of Korea

M.Sc. in Artificial Intelligence

Mar. 2022 – Aug. 2023

- Advisor: Prof. [Chulhee Yun](#) (Optimization & Machine Learning (**OptiML**) Laboratory, Kim Jaechul Graduate School of AI (GSAI), KAIST)
- Thesis: “Improved Convergence Rate of SGDA by Shuffling: Focusing on the Nonconvex-PL Minimax Problems” (Approved by [Chulhee Yun](#), [Se-Young Yun](#), & [Donghwan Kim](#))
- GPA: 4.22/4.3

### KAIST

Daejeon, Republic of Korea

B.Sc. in Mathematical Sciences

Mar. 2017 – Feb. 2022

- Minor in Computing Sciences
- Summa Cum Laude (GPA: 4.05/4.3)

### University of Twente

Enschede, Netherlands

Exchange Student Program

Feb. 2020 – Jul. 2020

- Major in Applied Mathematics

### Incheon Science High School (ISHS)

Incheon, Republic of Korea

High School

Mar. 2015 – Feb. 2017

- Early graduation by one year (i.e., two-year course)

## Awards

2025	<b>Top Reviewer (Top 1.88%: 206 of 10,943 reviewers)</b> , ICML 2025	Vancouver, Canada
2024	<b>Best Paper Award (Top 3)</b> , JKAIA 2024	Republic of Korea
2024	<b>Top Reviewer (Top 8.60%: 1,304 of 15,160 reviewers)</b> , NeurIPS 2024	Vancouver, Canada
2022	<b>NAVER Outstanding Theory Paper Award (Top 3)</b> , JKAIA 2022	Republic of Korea
2022	<b>Summa Cum Laude</b> , Bachelor's, KAIST	Republic of Korea
2017 – 2020	<b>The National Scholarship for Science and Engineering</b> , Korea Student Aid Foundation	Republic of Korea
2017 Fall	<b>Dean's List</b> , The School of Freshman, KAIST	Republic of Korea

## Experiences

### Google

New York, United States

Intern: Student Researcher Program (On-Site), Engineering

May 5th 2025 – Aug. 22nd 2025

- Host: [Srinadh Bhojanapalli](#) at Google NYC
- Research Topic: Advanced architectural component in Transformers

### SNU-KAIST AI/ML Theory Workshop

Gangneung, Republic of Korea

Organizer

Aug. 12th–14th, 2024

- Homepage: [nick-jhlee.github.io/snu-kaist-workshop](https://nick-jhlee.github.io/snu-kaist-workshop)
- Jointly organized by three research groups of Prof. Ernest K. Ryu, Prof. Min-hwan Oh, and Prof. Chulhee Yun.

### Machine/Deep Learning Theory + Physics (MDLTP) Seminar

Seoul, Republic of Korea

Organizer

Jul. 2022 – Feb. 2023

- Homepage: [sites.google.com/view/mdlt-p](https://sites.google.com/view/mdlt-p)
- Jointly organized by OSI Lab, OptiML, and CSSPL
- Topics: Learning theory, loss landscape, trajectory analysis, (stochastic) optimization, high-dimensional statistics, statistical/mathematical physics, scientific machine learning, and more.

### Geometric Deep Learning Seminar

Seoul, Republic of Korea

Seminar Participant

2022

- A seminar organized by OptiML and OSI Lab
- Resources: [[Homepage](#)] [[Lecture Videos](#)] [[Book](#)]

## KAIST 2021 Post-AI Research Project

Undergraduate Researcher

Daejeon, Republic of Korea

May 2021 – Dec. 2021

- Jointly advised by Prof. Sangyoon Yi (DS Lab, GSFS, KAIST) & Prof. Jinkyoo Park (Sys. Int. Lab, ISysE, KAIST)
- Project: Research on 'AI-augmented Organizations' of Collaborative Decision Making and Learning. Below, I list my contribution:
  1. *Algorithm Design*: Devised a model-based randomized algorithm for a single-player finite-horizon NK landscape optimization game
  2. *Experiment Assistance*: Conducted experiments on human-AI cooperation based on the algorithm that I devised

## Individual Study: Optimization for Deep Learning

Undergraduate Student @ KAIST

Daejeon, Republic of Korea

Mar. 2021 – Jun. 2021

- Advised by Prof. Jinwoo Shin (ALIN Lab, GSAI, KAIST)
- (1) gradient-based optimizers for large-batch setting (e.g., LARS & LAMB); (2) theoretical analysis on gradient clipping (paper reading)

## Individual Study: Deep Learning in Computer Vision

Undergraduate Student @ KAIST

Daejeon, Republic of Korea

Sep. 2020 – Feb. 2021

- Advised by Prof. Jong-chul Ye (BISPL, BBE, KAIST)
- Assignment: Semantic segmentation of kidney tumor with U-Net (with KiTS19 challenge dataset)
- Self-taught PyTorch coding on Linux Ubuntu

## Individual Study: Statistical Learning Theory

Undergraduate Student @ KAIST

Daejeon, Republic of Korea

Jun. 2020 – Aug. 2021

- Advised by Prof. Yeonseung Chung (MAS, KAIST)
- Resource: Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. "An Introduction to Statistical Learning: with Applications in R." Springer, 2013. [\[link\]](#)

# Ongoing Projects

## Length Generalization in Language Models (e.g., Transformers)

KAIST AI + Google Research

Seoul, Republic of Korea

Nov. 2023 – Current

- See: [C6], [C8].
- Transformer (Vaswani et al., 2017) models have exhibited tremendous success in various areas of artificial intelligence such as natural language processing (NLP), computer vision, and reinforcement learning.
- However, Transformers often struggle to achieve *length generalization*: they usually fail to extrapolate their performance onto longer sequences than those at the training time.
- We aim to come up with a unifying framework to solve the length generalization problem on several synthetic tasks and arithmetic tasks, such as addition, multiplication, copying/reversing under the presence of duplicate tokens, and the parity task. The exact structures for those tasks are known, so we might be able to inject the structure into our Transformer model directly.
- Furthermore, we would like to find a method that automatically discovers the hidden structure of the tasks to apply our methodology for length generalization to general NLP tasks.

## Fair (Streaming) Principal Component Analysis - Complexity Lower Bound

KAIST AI

Seoul, Republic of Korea

Mar. 2023 – Current

- In Lee et al. (2023) [C3], we have studied the memory-efficient & group-wise fair principal component analysis (PCA) methodology called 'Fair Streaming PCA'.
- We proved the upper bound of sample complexity of our main algorithm 'Fair Noisy Power Method (FNPM)' mainly exploiting matrix Bernstein inequality. However, the tightness of the bound has not been checked yet.

## Overcoming Loss of Plasticity of Neural Network

KAIST AI

Seoul, Republic of Korea

Nov. 2023 – Current

- See: [C2], [C5].
- Neural networks often struggle with the problem called 'loss of plasticity,' which refers to the phenomenon that a model loses the ability to adapt to the shifts in data distribution, especially when new data points appear.
- Interestingly, re-initializing the model partially or entirely, so-called "re-learning," greatly helps the model to recover its adaptability in practice.
- Why is it beneficial? Furthermore, can we develop a "smart" way of re-learning instead of re-initializing the whole model?

## Services

---

### Top-tier ML Conference/Journal Reviewer

From time to time

- TMLR. (3 papers)
- ICML 2025. (6 papers) — awarded Top Reviewer (Top 1.88%: 206 of 10,943 reviewers)
- ICLR 2025. (3 papers)
- NeurIPS 2024. (6 papers) — awarded Top Reviewer (Top 8.60%: 1,304 of 15,160 reviewers) + Free Registration
- ICLR 2024. (2 papers)
- NeurIPS 2023. (2 papers)

### 1st GPU server manager of OptiML lab

June 2022 – Feb 2024

- Being involved in installing OptiML lab's very first 5 GPU servers and a storage server
- Allocating GPU nodes to lab members
- Managing errors occurred in the servers

## Languages

---

- English** Professional Proficiency (i.e., sufficient for academic activities)
- Korean** Native proficiency
- Others** Had some introductory courses on French, German, Classical Latin, & Chinese.

## Skills

---

- Programming** Familiar: Python 🐍 (PyTorch, NumPy, Scikit-learn, Jupyter, Pandas, JAX, etc.), MATLAB.  
Novice: C, C++, R, HTML/CSS, Scala
- Computer Misc.** Familiar:  $\LaTeX$  (Overleaf/VSCoDe/MacTex), Git 🔌, Microsoft Office, Keynote  
Novice: Adobe (Lightroom, Premiere Pro, After Effects, Photoshop)
- Music & Hobby** Playing the drums 🥁 and percussions. Begun to learn in 2009. Joined to music bands listed below:
- ISHS (*Cha-rang*, 2015–2016)
  - KAIST (*Muse KAIST*, 2017–2019; *Carpe Diem*, 2019)
  - Club “Music Space” (Team *Woodstone*, 2024)
- I’m a huge music fan; mostly, jazz (fusion, acid, hard bop, cool, bossa nova, etc.), funk, K-indie, Rock, Latin, and many more.