Seungyoo Lee

KAIST Kim Jaechul AI Graduate School
punctuate@kaist.ac.kr
luj0407@gmail.com
https://huggingface.co/DopeorNope
https://github.com/DopeorNope-Lee
https://kr.linkedin.com/in/seungyoo-lee-62b098281

PERSONAL DATA

- Nationality: Korean
- Skills: Python (Torch, TensorFlow), C++, R

EDUCATION

- **Doctor of Philosophy (Ph.D.) in Artificial Intelligence** Sep 2026 – Present
  KAIST, Kim Jaechul Graduate School of AI, Seoul, Korea
  Research Focus: Large Language Models
  Advisor: Prof. Juho Lee, SIML Lab

- **Master of Science in Artificial Intelligence** Sep 2024 – Aug 2026
  KAIST, Kim Jaechul Graduate School of AI, Seoul, Korea
  Research Focus: Large Language Models
  Advisor: Prof. Juho Lee, SIML Lab

- **Bachelor of Science in ICT Convergence and Business**
  Handong Global University, Pohang, Korea
  Primary Major: ICT Convergence, Secondary Major: Business
  GPA (Major): 4.37/4.5 | GPA (Total): 4.24/4.5
  Advisor: Prof. Xiaopeng Yang

RESEARCH INTEREST

- ✓ Large Language Model (LLM)
- ✓ Synthetic Data Generation
- ✓ Multi-Modal

PUBLICATIONS (INTERNATIONAL)

Conference

1. Lee, H., **Lee, S.**, & Lee, J. (2025). Variational Bayesian Pseudo-Coreset. In *International Conference on Learning Representations (ICLR) 2025.*
2. **Lee, S.**, Nam, G., Choi, M., Lee, H†., & Lee, J†. (2025). PANGEA: Projection-Based Augmentation with Non-Relevant General Data for Enhanced Domain Adaptation in LLMs. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems (NeurIPS) 2025*.
3. **Lee, S.**, Nam, G., Lee, H†., & Lee, J†. (2026). OCNR: Stabilizing Self-Play by Mitigating Iteration-Collapse With One-Class Novelty Rewards. In *Forty-third International Conference on Machine Learning (ICML) 2026*.
4. Kim, S. E., **Lee, S.**, Fong, E., Lee, H†., & Lee, J†. (2026). From Drift to Coherence: Stabilizing Beliefs in LLMs. In *Forty-third International Conference on Machine Learning (ICML) 2026*.
5. Jang, C., Choi, M., Kim, Y., **Lee, S.**, Lee, J†., & Lee, H†. (2026). Confidence is Not Universal: Task-Dependent Calibration and Emergent Behavior in LLMs. In *Forty-third International Conference on Machine Learning*

(ICML) 2026.

6. **Lee, S.**, Nam, G., Lee, S., Lee, H., & Lee, J. (2026). SAVE: Sparsity-Aware Influence Estimation for Vocabulary-Expanded LLMs. In *The Fortieth Annual Conference on Neural Information Processing Systems (NeurIPS) 2026*.
7. Kim, S. H*., Jang, C*., **Lee, S.**, Ham, J., Bak, Y., Kim, B., & Lee, J. (2026). Mixture-Trained Merging for Unified Multi-Objective Models. In *The Fortieth Annual Conference on Neural Information Processing Systems (NeurIPS) 2026*.

Journals (SCIE)

1. **Lee, S*.**, Han, K*., Park, S., & Yang, X. (2022). Vehicle Distance Estimation from a Monocular Camera for Advanced Driver Assistance Systems. *Symmetry*, *14*(12), 2657.
2. Yang, X., Park, S., **Lee, S.**, Han, K., Lee, M. R., Song, J. S., ... & Do Yang, J. (2023). Estimation of right lobe graft weight for living donor liver transplantation using deep learning-based fully automatic computed tomographic volumetry. *Scientific Reports*, *13*(1), 17746.
3. **Lee, S**., Han, K., Shin, H., Park, H., Kim, S., Kim, J., Yang, X†., Yang, J. D†., Yu, H. C., & You, H. (2025). G-UNETR++: A Gradient-Enhanced Network for Accurate and Robust Liver Segmentation from Computed Tomography Images. *Applied Sciences*, 15(2), 837.
4. **Lee, S**., Han, K., Shin, H., Kim, S., Park, H., Kim, J., ... & Yu, H. C. (2026). Automatic Liver Couinaud Segmentation from Computed Tomography Scans with a Gradient Enhanced Hierarchical Cascade Deep Learning Network. *Current Problems in Surgery*, 75, 101957.
5. Shin, H*., Han, K*., **Lee, S***., Park, H., Kim, S., Kim, J., ... & You, H. (2026). Deep Learning-Based Liver Tumor Segmentation from Computed Tomography Scans with a Gradient-Enhanced Network. Diagnostics, 16(3), 429.

PUBLICATIONS (KCI)

Conference

1. Han, K*., **Lee, S***., Shin, J., Jang, S., Seon, H., & Yang, X. (2023). Human Activity Recognition using Sensor Series Transformer with High-frequency Positional Encoding. *Korea Computer Congress (KCC) 2023, KIISE*, 2052-2054.

RESEARCH EXPERIENCES

- **Research Intern** at *Kakao* Mar. 2026 ~ Aug. 2026
  - ✓ Encoder-Free Multimodal Models
  - ✓ Domain-Specific Large Language Models

- **Student Researcher** at *AI Medical Vision Lab, Handong Global University* ***(Head)*** Mar. 2022 ~ Mar.2024
  (Prof. Xiaopeng Yang)
  - ✓ Vehicle Distance Estimation Using a Monocular Camera
  - ✓ Liver segmentation with deep-learning model
  - ✓ Visual Data-Based Driver Fatigue Detection Using Deep Learning Models

- **Research Assistance** at *Korea Real Estate Agency, Handong Global University* Aug. 2021 ~ Mar.2024
  (Prof. Jung Eui Hong)
  - ✓ Analysis of Local Housing Market Using Machine Learning Algorithm
  - ✓ Development of Risk Prediction Model for Fraudulent Subscription Using Machine Learning Algorithm
  - ✓ Improvement of Machine Learning Algorithm for the Advancement of Housing Market Analysis

- ✓ Detection of Unqualified Real Estate Subscriptions Using Machine Learning (In-depth Study on Fraudulent Subscriptions)

AWARDS AND HONORS

2023 Month's LLM: ***NIA Director's Prize***, National Information Society Agency, Korea

2022 Healthcare AI Hackathon Competition Using Walking Data: ***Grand Prize***, LINC 3.0 Project Group, Korea

2022 Problem-Solving Idea Contest Using Software: ***Best Prize***, Handong Global University, Korea

EXTRA-CURRICULAR ACTIVITIES

- **Instructor Experience**

✓ ***Quantization for Fine-Tuning LLM Models*** at FastCampus **(online).** Aug. 2024 ~Nov.2024
- Taught optimization techniques for large language models using quantization
- Covered the latest trends and strategies for optimizing LLMs

✓ ***AI Data Engineering Advanced Course*** at LG U+ **(offline)** Oct. 2024 ~Nov.2024
- Delivered lectures on LLM and Quantization
- Guided practical sessions on fine-tuning and quantization of LLM models

✓ ***GPU Optimization for Fine-Tuning LLM Models*** at FastCampus **(online)** Oct. 2024 ~ Jan.2025
- Delivered lectures on efficient LLM training using GPU optimization
- Shared real-world strategies for GPU optimization in practical projects
- Synthetic data generation

✓ ***Developing High-Performance AI Models with Open Source LLMs*** at FastCampus **(online)** March. 2025
- Covered the reasoning mechanisms behind state-of-the-art LLMs (OpenAI o1, DeepSeek-R1, Kimi k1.5)
- Taught inference-time (test-time scaling, CoT, budget forcing) and training-time (RLHF, GRPO, DPO, RFT) methods
- Guided projects on training reasoning models with open-source LLMs under limited compute

✓ ***Advanced Software Course (TSP Division)*** at Samsung Electronics **(offline)** May. 2025
- Conducted an intensive 5-day training on advanced software development and optimization strategies
- Delivered practical sessions focused on enhancing software engineering capabilities for the TSP division

✓ ***Master Class: Building Custom LLMs (5-Week Intensive)*** at FastCampus **(offline)** June 2025 ~ Present
- Led a 5-week intensive offline bootcamp aimed at developing custom Large Language Models
- Mentored industry professionals from top tech companies on end-to-end LLM development
- Guided hands-on projects focusing on fine-tuning, evaluation, and deployment of domain-specific models
- Delivered across multiple cohorts (2025 – 2026, ongoing)

✓ ***Generative AI & Business Innovation*** at Hyundai HDS **(offline)** June. 2025
- Conducted sessions on the application of Generative AI in the insurance and finance sector
- Focused on leveraging AI tools for business process improvement

✓ ***Generative AI Fundamentals*** at Kocom **(offline)** July. 2025
- Provided comprehensive training on Generative AI principles and prompt engineering
- Guided attendees through practical use cases of AI in industry

✓ ***Generative AI & LLM Optimizations*** at Hyundai Motor Company **(offline)** Sep. 2025

- Delivered advanced lectures on Generative AI trends and LLM optimization strategies
- Covered practical techniques for model efficiency and deployment in corporate environments

✓ ***Signature: Large Language Model (LLM) Development*** at FastCampus **(online)** Oct. 2025 ~ Mar. 2026
- Comprehensive course on the full lifecycle of Foundation LLMs: Pre-training, SFT, and DPO
- Covered advanced reasoning model architectures and optimization strategies for 1B+ parameter models
- Delivered in-depth curriculum on data engineering, tokenization, and GPU utilization for LLM training
- Covered RLHF/GRPO, test-time scaling, and quantization (PTQ/QAT) with multi-GPU distributed training

✓ ***Advanced Course: Multimodal LLMs & Quantization*** at Samsung Electronics **(offline)** 2026 ~ Present
- Leading an advanced course on multimodal large language models and quantization
- Covering model compression and efficient deployment of multimodal LLMs

✓ ***Generative AI & LLM Training*** at Hyundai Mobis, Samsung SDS **(offline)** 2026 ~ Present
- Delivering corporate lectures on LLMs and generative AI for industry professionals

✓ ***LLM-based AI Agents*** at Financial Security Institute **(online)** 2026 ~ Present
- Delivering an online course on LLM-based AI agents for financial-sector professionals

✓ ***Hybrid Agentic Coding with Claude Code & Open-Weight LLMs*** at FastCampus **(online)** Oct. 2026 ~ Present
- Designed hybrid workflows routing tasks between Claude Code and open-weight models (DeepSeek, Qwen, GLM) via OpenRouter
- Covered LLM fundamentals, agents, workflows, and multi-agent systems, culminating in a capstone project
- Taught task-specific model selection and cost optimization while preserving development quality

✓ ***Invited Talk***
- Skelter Labs
- YouTube Teddy Note
- Fast Campus
- Hyundai Motors
- Samsung Electronics

- **Partner Researcher** at *Markr.AI research Team* Aug.2024~May 2026

✓ ***AutoRAG***
- Developing Domain Specific LLM in sparse domain data pool
- LLM fine-tuning engineering
- LLM Quantization
- LLM Guidance Using Sequential Monte Carlo Sampling

✓ ***Developing Korean LLM***
- LLM fine-tuning engineering
- Synthetic data generation

- **Researcher** at *Markr with Media Group Human & Forest* Dec. 2022 ~Aug.2024

✓ ***Robustness Large Language Model (LLM) and Korean LLM Development***
- Developed the LLM that secured the first place on the 'Open LLM Leaderboard' as of January 14, 2024 **Model Name**: *DopeorNope/SOLARC-MOE-10.7Bx6* (https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard).
- Developed the LLM that secured the first place on the 'Open Ko LLM Leaderboard' as of November 11, 2023 **Model Name**: *HumanF-MarkrAI/COKAL-DPO-13b-v2*

(https://huggingface.co/spaces/upstage/open-ko-llm-leaderboard).

- The SFT model was trained with the LoRA methodology and the DPO methodology, a reinforcement learning tuning methodology, was applied.
- In data cleansing tasks, the Near Dedup algorithm was applied using Min Hash and Min Hash LSH to stabilize the distribution of text data by reducing the Jaccard similarity to 0.8 or higher.
- Applied MOE methodology to enhance model's performance and make it more robustness.
- Developed the implementing the IA3 fine-tuning methodology, which dramatically reduces the number of parameters while achieving great performance (https://github.com/Marker-Inc-Korea/K-G-OAT).

✓ ***NeRF Project***

- Developing D- NeRF Model (Neural Radiance Fields for Dynamic Scenes)
- Developing GUI Prototype for 3D Novel View Synthesis
- Developing Human 3D Pose Estimation Model

- **Researcher** at *Markr with Media Group Human & Forest* Sep.2022~Aug.2024

✓ ***Medical Vision***

- Liver Segmentation
- Couinaud Segmentation
- Tumor Segmentation

✓ ***Computer Vision***

- Autonomous Driving
- Human Pose Estimation