Seungjae Lim

sjlim@casys.kaist.ac.kr | LinkedIn

I'm Seungjae Lim, an M.S. student in the CASYS Lab at KAIST, working at the intersection of research and engineering in machine learning systems. My current focus is on system-level optimization for efficient LLM-based multi-agent serving. I seek opportunities to learn from diverse experiences and continuously evolve in both technical depth and practical impact.

Education

KAIST (Korea Advanced Institute of Science and Technology) Sep '24 – Now
M.S. Student in School of Computing
Research topic: Efficient Serving of Multi-Agent LLM Systems
Advisor: Prof. Youngjin Kwon
GPA: 4.1/4.3

KAIST (Korea Advanced Institute of Science and Technology) Mar '19 – Aug '24
B.S. in School of Computing
Minor in Electrical Engineering
GPA: 3.91/4.3 (major), 3.68/4.3 (cumulative)
Graduated as Top 10 Leadership & Volunteer Honoree

Employment

Research Intern, CASYS KAIST Jun '23 – Aug '24

- Research on effective LLM inference systems

ML Engineering Intern, NAVER CLOVA Aug '22 – Feb '23

- Improved delivery simulator with GNNs and embedding techniques
- Participated in AutoForecast (auto time-series forecasting) project

Publications

TTServing: Efficiently Serving Large Language Models with Test-time Search for Inference-time Scaling Under Review

Sukmin Cho, **Seungjae Lim**, Sangjin Choi, Hyunsu Ye, Wooseok Gwak, Youngjin Kwon

COSPEC: Efficient LLM Serving with Colocated Speculative Decoding Under Review

Sangjin Choi, Hyunjae Lee, **Seungjae Lim**, Sukmin Cho, Youngjin Kwon

RAG-based Unanswerable Question Detection in Clinical Text-to-SQL CIKM 2025

Donghee Han, **Seungjae Lim**, Mun Yi

Leveraging LLM-Generated Schema Descriptions for Unanswerable Question Detection in Clinical Data COLING 2025

Donghee Han*, **Seungjae Lim***, Daeyoung Roh, Sangryul Kim, Sehyun Kim, Mun Yi
*(*Equal contribution)*

Highlighted Projects

Efficient Multi-Agent LLM System Serving Ongoing

- Solving system-level challenges for fast Multi-Agent LLM serving

Mad Camp 2021 Winter

- Participated in a student-driven programming camp hosted by KAIST and Krafton
- Completed 4 full-stack web/app projects using Android Studio, React, Express, and MySQL

Awards

Excellence Award (2nd Place) @ Elice AI Hellothon 2024
- Awarded by Elice
- Caregiver cognitive-activity lesson guide creator & elderly interactive AI drawing diary service

Grand Prize (1st Place) @ SPARCS Science Hackathon 2024
- Awarded by Daejeon Mayor's
- Web-based XR labs for diverse experimental environments

Leadership & Volunteer Excellence Award @ KAIST 2019 - 2024
- Awarded by KAIST President
- Top 10 among all 24–25 graduates for exceptional leadership and volunteer work (175 hours)

Grand Prize (1st Place) @ Social Problem-Solving Volunteer Showcase 2024
- Awarded by Gyeonggi-do Governor
- Driving engagement via a generative reward system for social volunteering

Gold Prize (1st Place) @ ICISTS Hackafair 2024
- Awarded by the President of KAIST
- RAG-based profanity-filtering B2B platform with customizable policies

Best Prize (1st Place) @ Social Problem-Solving Idea Hackathon 2024
- Generative volunteer reward system for creating customizable characters

Best Prize (3rd) @ K-Data Science Hackathon 2023
- Awarded by the Chairman of the National Research Foundation of Korea
- Fine-tuned & served distortion-free K-LLM with RAG on Korean history

Top 10 @ Elice AI Edu Hackathon 2023
- Literacy-promotion app using generative AI with Chain-of-Thought

Silver Prize (2nd) @ KAIST-POSTECH-UNIST Data Science Contest 2023
- Awarded by ECMiner
- Business-owner prediction & pop-up / survey ad planning with Hana 1Q usage data

Best Prize (1st) @ Social Problem Solving Volunteer Hackathon 2023
- Real-estate listing & brokerage recommendation app for single female-headed households

Grand Prize (1st) @ Korea Donation for Education Idea Contest 2021
- Awarded by Minister of Science and ICT
- Dynamic coding-based fairy-tale platform for English education

Teaching Experience

Teaching Assistant, CS206 Data Structure, KAIST 2021 Fall

Skills

Languages: Korean, English

Programming: Python, C/C++, CUDA, vLLM