

Keunyoung Kim

keunyoung.kim@snu.ac.kr | [linkedin](#)

RESEARCH INTERESTS

My research focuses on **efficient large language model training**, with an emphasis on reducing the computational and memory costs of pre-training. I am particularly interested in low-rank optimization, gradient subspace methods, and optimizer design, including how optimization capacity should be allocated across model components and how gradient structure and training dynamics affect convergence. I am also increasingly interested in broader approaches to memory-, compute-, and communication-efficient foundation model training.

EDUCATION

Seoul National University <i>Ph.D. in Artificial Intelligence (Advisor: Prof. Nojun Kwak)</i>	Seoul, Korea 2026 – Present
Seoul National University <i>M.S. in Artificial Intelligence (Advisor: Prof. Nojun Kwak)</i>	Seoul, Korea 2026
Republic of Korea Air Force Academy <i>B.S. in Systems Engineering & Bachelor of Military Science (Dual Degree)</i>	Cheongju, Korea 2016

RESEARCH EXPERIENCE

- Developed **MoARa**, a module-aware low-rank optimization method for efficient LLM pre-training that allocates projection rank according to module-specific sensitivity.
- Designed a structure-preserving decomposition method to retain local gradient magnitude information during low-rank gradient projection.
- Reduced steps-to-target from **150K to 94K** in 7B-model pre-training, achieving a **34.7% reduction in wall-clock training time** relative to GaLore.
- Demonstrated transfer of module-wise sensitivity profiles across model scales, reusing profiling results from a 350M model for 1B and 7B pre-training.

PUBLICATIONS

Keunyoung Kim and Nojun Kwak. “MoARa: Module-Aware Rank Allocation and Structure-Preserving Decomposition for Low-Rank LLM Pre-training.” *EMNLP 2026 Main Conference*. To appear.

PROFESSIONAL EXPERIENCE

Captain (O-3), Information and Communications Officer <i>Republic of Korea Air Force (ROKAF)</i>	2016 – 2025 Korea
• Led planning, coordination, and execution of information and communications training on the ROKAF side during large-scale ROK–U.S. combined exercises, coordinating with U.S. Air Force counterparts (A-6). • Planned, deployed, and maintained information and communications systems, ensuring reliable operation and lifecycle support. • Managed information and communications projects and budgets, coordinating requirements, schedules, and execution. • Operated and maintained network, server, and communications infrastructure supporting ROKAF operations.	

PROJECTS

Multi-Node GPU Research Cluster <i>Infrastructure Architecture & Deployment Planning</i>	2026 Seoul National University
• Designed the overall architecture and operating model for a multi-user research cluster with 6 compute nodes and 36 NVIDIA GPUs, including centralized job scheduling, isolated user environments, shared storage, and high-speed interconnects. • Defined GPU allocation, container isolation, storage, and operational policies for Slurm-based scheduling, and coordinated system deployment and validation.	

FELLOWSHIPS AND AWARDS

AI Fellowship (Fully Funded) <i>Seoul National University</i>	2026 – present
Academic Excellence Scholarship <i>Seoul National University</i>	2026