Saehun Chun

Integrated M.S./Ph.D. Student
Department of Software, Sungkyunkwan University
Suwon, Republic of Korea

thousae@skku.edu · Homepage · Google Scholar · GitHub · LinkedIn

RESEARCH INTERESTS

- Efficient AI Systems and Foundation-Model Serving
- Memory and Computation Reuse for Interactive Multimodal AI
- Embodied AI and Robot Foundation Models

EDUCATION

2025–Present **Sungkyunkwan University** Suwon, Republic of Korea
Integrated M.S./Ph.D. Program in Software
Department of Software | Lab: CSI Lab | Advisor: Honguk Woo
GPA: 4.14 / 4.5

2019–2024 **Sungkyunkwan University** Suwon, Republic of Korea
B.S. in Software
Department of Software | GPA: 3.41 / 4.5

RESEARCH EXPERIENCE

2024–Present **CSI Lab, Sungkyunkwan University** Suwon, Republic of Korea
Graduate Researcher

- Study efficient VLA and multimodal inference, focusing on temporal reuse, asynchronous execution, and stale-state mitigation.
- Develop structured memory reuse for embodied agents (HeMeR, NeurIPS 2026 Oral).
- Develop function-level reuse for robot policies (FCGraft, ICML 2026 Poster).

2023–2024 **VLDB Lab, Sungkyunkwan University** Suwon, Republic of Korea
Undergraduate Research Assistant

- Investigated undo-page buffer pollution in HTAP workloads and evaluated memory-management mechanisms in MySQL/InnoDB using C++.

PUBLICATIONS 5 PAPERS

International Conferences (3)

[1] HeMeR: Heterogeneous Memory Reconciliation for Embodied Agents via Structured KV Reuse.
Saehun Chun, Wonje Choi, Jinwoo Jang, TaeYoon Kwack, and Honguk Woo*.
Conference on Neural Information Processing Systems (NeurIPS), 2026. **Oral.**

- Reconciles heterogeneous agent memory through selective KV reuse. In real-world tests versus MemRL, success rises from 64.28% to 83.20%, while inference time falls from 2.94 to 1.87 s/step.

[2] Functional Cache Grafting: Robust and Rapid Code-Policy Synthesis for Embodied Agents.
Saehun Chun, Wonje Choi, Sera Choi, Sanghyun Ahn, and Honguk Woo*.
International Conference on Machine Learning (ICML), 2026. **Poster.**

- Composes validated robot-policy functions via cache-stitching and repairs faulty spans via cache-patching. Policy synthesis is 2.3× faster than RAGCache on average across evaluated tasks.

[3] Efficient Skill Grounding via Code Refactoring with Small Language Models.
Sera Choi, Wonje Choi, **Saehun Chun**, Daehee Lee, Chaeun Lee, Jooyoung Kim, and Honguk Woo*.
International Conference on Machine Learning (ICML), 2026. **Poster.**

- Grounds reusable skills through ontology-guided, localized code refactoring while preserving skill semantics. RECENT achieves 73.0% and 82.5% task success under kinematic and end-effector variations, respectively.

Domestic Conferences (2)

[1] Undo Read Buffer: Preventing Buffer Pollution due to Undo Blocks in HTAP Workloads.
Saehun Chun, Jonghyeok Park, Sang-Won Lee*, and Young Ik Eom.
Korea Computer Congress (KCC), 2024.

- Separates undo/data buffers in MySQL/InnoDB, cutting data-page reads and writes by about 30% and 25%.

[2] Analysis of Buffer Pollution Caused by Undo Pages in HTAP Workloads.
Saehun Chun, Bo-Hyun Lee, Kyong-Shik Lee, Sang-Won Lee*, and Young Ik Eom.
Korean DataBase Conference (KDBC), 2023.

- Quantifies undo-page buffer pollution in MySQL/InnoDB, showing 21–27% lower throughput in HTAP tests.

RESEARCH PROJECTS

2024–2029 **HiPER: High-Performance Exabyte Storage Systems**
Project Manager / Graduate Researcher
Samsung Electronics

- Study storage-scale memory for foundation-model embodied agents, storing and reusing KV caches for efficient, reliable action selection.

2025–2029 **Accelerated Insight Reasoning via Continual Learning**
Project Researcher
Funded by MSIT through IITP

- Develop reusable robot-policy functions and structured neural memory to retain validated experience and reduce repeated computation across embodied tasks.

2022–2026 **Adaptive Personality for Intelligent Agents**
Project Researcher
Funded by MSIT through IITP

- Develop embodied agents that adapt reusable skills to changing tasks and robot configurations through code refactoring and feedback-driven policy repair.

Jan. 2024 **Global Capstone Design**
Project Manager
Sungkyunkwan University & Victoria University of Wellington

- Prototyped ML-based MySQL buffer management; presented the work and collaborated with students at Victoria University of Wellington.

2020 **Industry-Academic Collaboration Project**
Project Manager
SK M&N Service

- Led a Python/TensorFlow project for speech-based video summarization. Repository.

AWARDS AND HONORS

2025–2026 **Graduate Research Fellowship**
National Research Foundation of Korea

TEACHING EXPERIENCE

2026–Present **GenAI PowerUser Program Level 4 (Multi Agent framework)**
Hands-on Training Instructor
Samsung Electronics University

Fall 2026 **Artificial Intelligence Project**
Hands-on Training Instructor
Sungkyunkwan University

Spring 2026 **System Programming**
Teaching Assistant
Sungkyunkwan University

Spring 2025 **Mobile Application Programming**
Teaching Assistant
Sungkyunkwan University

RESEARCH EXPERTISE AND TECHNICAL SKILLS

Embodied AI and Skill Grounding: Foundation-model agents, executable robot policies, ontology-guided skill adaptation, and feedback-driven policy repair.

Efficient Inference: VLA/multimodal inference, temporal reuse, asynchronous execution, and stale-state mitigation.

Agent Memory: Typed agent memory, LoRA-aware KV reuse, and selective state reconciliation.

ML Systems: Memory- and storage-efficient model serving, cache management, and HTAP buffer management.

Benchmarks and Platforms: ALFWorld, Habitat, ALFRED, TEACh, RLBench; Franka FR3 and UR7e.