

Chanhee Lee

4328 Scorpius St, Orlando, FL 32816, United States

chanhee26.lee@gmail.com | (+1) 602-451-5792 | [LinkedIn](#) | [Portfolio](#) | [GitHub](#)

RESEARCH INTERESTS

- CXL-based memory systems, AI/ML systems and software platforms, on-device and edge LLMs, and LLM workloads analysis/visualization.

SUMMARY

- AI systems and software platform engineer with 9+ years at Samsung Research and recent postdoctoral work on CXL-based memory systems and on-device LLMs
- Designed production AI inference platforms for Android/TVs and distributed software frameworks for IoT
- Interested in LLM optimization, AI system HW/SW codesign, and web-based AI workload visualization

PROFESSIONAL EXPERIENCE

- **University of Central Florida**, Orlando, FL, United States
Postdoctoral Researcher, Department of Computer Science Dec. 2024 - Present
 • Designed memory-efficient LLM inference for CXL memory systems (Gem5 simulation).
 • Extended Linux kernel for fault-tolerant non-volatile memory (NVM) in real-time edge systems [[linux_pmo](#)].
 • Lead 7 students to analyze on-device mlc-llm/llama.cpp performance and develop a personalized web AI agent.
- **Arizona State University**, Tempe, AZ, United States
Visiting Scholar (Voluntary), School of Computing and Augmented Intelligence Jan. 2024 - Dec. 2024
 • Led on-device LLM personalization (Android + knowledge graphs) [[EMSOFT WIP 2024](#)] [[WWW Short 2025](#)].
 • Built Rust reactive engine in Lingua Franca (CAP-theorem guarantees) [[lf-rust-rti](#)].
- **Samsung Electronics Inc.**, Seoul, Republic of Korea
Staff Engineer/Best Reviewer, Platform Team, Samsung Research Aug. 2014 - Dec. 2024
 • Led platform team demo at Samsung Research annual showcase (**Top 3/10 major divisions**, ~2K researchers), GPU/NPU inference + SmartTrainer TV app
 • Led SmartFS filesystem development (fixed power-cut crashes for Samsung home appliance production boards).
 • Ported Oxigraph graph DB REST APIs to Android for edge AI workloads.
 • Led OCF IoTivity D2D framework for ARTIK production boards (1st author, [Springer LNCS ICIoT 2018](#)) [[RT-OCF](#)].
 • Led various overseas research teams to develop GUI-based performance analysis and visualization tools.

EDUCATION

- **Ph.D.** Electrical Engineering & Computer Science, Seoul National University 2014
- **M.S.** Computer Science, KAIST 2009 | **B.S.** Computer Science, KAIST 2007

SKILLS

- **Languages:** Rust, C/C++, Python, Java, React, NodeJS
- **AI/ML:** llama-cpp/MLC-LLM, PyTorch/TensorFlow, LLM workload analysis/visualization, on-device ML optimization
- **Platforms:** CXL/Gem5, Linux kernel, Docker, CI/CD, embedded systems, visualization framework, reactive systems

PUBLICATIONS

- 8 conference papers (EMSOFT, WWW) + 2 journal papers (JSPS, TECS) | 250+ citations [[Google Scholar](#)]

AWARDS & LEADERSHIP

- UCF Preeminent Postdoc Fellow (2025-27), Samsung Ph.D. Scholarship (2013)
- Invited Talk: "Remote inference with IoT" (Hanyang Univ., 2023)
- Reviewer: ISCA/ASPLOS/HPCA/PPoPP/LCTES (2025) and ICS (2026), ACM Member