

CHANYOUNG JUNG

FuriosaAI ◇ Seoul, Republic of Korea

cy.jung@furiosa.ai ◇ linkedin.com/in/chanyoungjung ◇ github.com/lovelyzzkei

EDUCATION

Yonsei University - Seoul, Republic of Korea Mar 2024 - Feb 2026
M.S in Computer Science and Engineering

Yonsei University - Seoul, Republic of Korea Mar 2018 - Feb 2024
B.S in Computer Science and Engineering
GPA: 4.24/4.50, Rank: 6/106, Graduated with High Honors

EXPERIENCE

FuriosaAI, Seoul, Republic of Korea Mar 2026 - Present
AI Software Engineer

Mobile Embedded Systems Lab., Seoul, Republic of Korea Dec 2022 - Feb 2026
Research Assistant

- Architected **on-device, real-time, and end-to-end vision AI systems for mobile/edge SoCs** (Qualcomm Snapdragon, NVIDIA Jetson), optimizing system performance (latency, energy) and model accuracy
- Authored 4 top-tier mobile systems papers (1 first-author at MobiSys '25, 3 co-authored at EuroSys '26, MobiCom '24/'25) and awarded the **Best Paper Award** at MobiSys 2025.
- Awarded a government fellowship (\approx \$9,000) and served as a teaching assistant for Operating Systems and System Programming courses.

PROJECTS

viNPU @ ACM EuroSys 2026 || Qualcomm AI Engine Direct SDK (QNN), ONNX, Android 2025

- Designed an inference optimization framework for the efficient execution of Vision Foundation Models (VFMs) on mobile NPUs.
- Authored a **graph-level dataflow optimization** to enhance the on-chip data locality of the attention mechanism.
- Achieved an average 11.5x speedup and up to 44.9x energy savings over baselines while maintaining negligible accuracy degradation compared to floating-point execution on Qualcomm SoCs (Galaxy S22, S25).

ARIA @ ACM MobiSys 2025 (Best Paper Award) || QNN, LiteRT, Android 2024

- Designed a **heterogeneous computing pipeline** that distributes VFM inference workloads across GPU and NPU based on processors' performance characteristics.
- Developed an inference engine for Snapdragon SoCs that schedules VFM inference based on scene dynamics, device motion, and GPU performance variations due to thermal throttling.
- Achieved sustained performance of 30 FPS with up to 72% improvement of prediction accuracy compared to baselines.

AWARDS & HONORS

Best Paper Award. ACM MobiSys 2025 (Anaheim, CA).

Grand Prize. Software Capstone Design, Yonsei University (2023).

Academic Honors (Yonsei University)

- Graduated with High Honors (Feb. 2024).
- Recipient of Semester Honors: Highest Honors (2018), High Honors (2019), and Honors (2019, 2023).

PUBLICATIONS (*CO-FIRST AUTHOR)

viNPU: Optimizing Vision Transformer Inference on Mobile NPUs

Jeho Lee, Gunjoong Kim, Chanyoung Jung, Jaehee Kim, Seonghoon Park, Hojung Cha

The 21st ACM SIGOPS European Conference on Computer Systems (EuroSys '26)

ARIA: Optimizing Vision Foundation Model Inference on Heterogeneous Mobile Processors for Augmented Reality

Chanyoung Jung*, Jeho Lee*, Gunjoong Kim, Jiwon Kim, Seonghoon Park, Hojung Cha

The 23rd ACM International Conference on Mobile Systems, Applications, and Services (MobiSys '25)

Best Paper Award, Invited to GetMobile

Vega: Fully Immersive Mobile Volumetric Video Streaming with 3D Gaussian Splatting

Gunjoong Kim*, Seonghoon Park*, Jeho Lee, Chanyoung Jung, Hyungchol Jun, Hojung Cha

The 31st ACM Annual International Conference on Mobile Computing and Networking (MobiCom '25)

Invited to GetMobile

Panopticus: Omnidirectional 3D Object Detection on Resource-constrained Edge Devices

Jeho Lee, Chanyoung Jung, Jiwon Kim, Hojung Cha

The 30th ACM Annual International Conference on Mobile Computing and Networking (MobiCom '24)

3DGS on Your Phone: Towards Fully Immersive Mobile Volumetric Video Streaming

Gunjoong Kim, Seonghoon Park, Jeho Lee, Chanyoung Jung, Hyungchol Jun, Hojung Cha

ACM GetMobile: Mobile Computing and Communications, Vol. 30, No. 1 (Invited)

VFM in Your Hands: Optimizing Real-Time Scene Understanding for Mobile Augmented Reality

Jeho Lee, Chanyoung Jung, Gunjoong Kim, Jiwon Kim, Seonghoon Park, Hojung Cha

ACM GetMobile: Mobile Computing and Communications, Vol. 29, No. 4 (Invited)

SKILLS

Programming

Python, C++, Java

Language

English (Intermediate), Korean (Native)

Frameworks

Android, QNN, PyTorch, LiteRT (Tensorflow Lite), TensorRT, CUDA

Hardware Platforms

Qualcomm Snapdragon SoCs (with Hexagon NPUs), NVIDIA Jetson Series

Interests

On-device AI, NPUs-aware Optimization, Graph Optimization