Junhyeok Lee

📱 (+82)10-3757-0231 | ✉ wnsx0000@gmail.com | 🏠 https://wnsx0000.github.io/ | ⭘ wnsx0000 | in Junhyeok Lee

Education ─────────────────────────────

Soongsil University   *Seoul, Republic of Korea*
BS in Software   Mar. 2021. - Present
GPA: 4.39/4.5 | Major GPA: 4.41/4.5 | Rank: 2/128

Research Interests ─────────────────────

Efficient ML   Efficient Inference, Model Compression, Efficient LLM/VLM, Efficient model serving

Experience ────────────────────────────

Advanced Intelligent Computing Architecture Lab (AICA) ⬀   *Soongsil University*
Undergraduate Research Intern, Advisor: Youngho Gong   Dec. 2024 - Present
- Participated in a reading group focused on efficient LLM serving and efficient VLM/VLA models
- Participated in research on efficient VLM/VLA models

TinyML and Efficient Deep Learning Computing (Fall 2024) ⬀   *Online*
MIT Han Lab Online Course   Dec. 2024 - Feb. 2025
- MIT Han Lab online course covering pruning, quantization, knowledge distillation, and efficiency of LLMs and vision models
- Synthesized key concepts from the course into a 100-page document typeset in LaTeX ⬀

PyTorch Tutorial Translation Project of Open Source Contribution Academy (OSSCA)   *Online*
Open Source Contributor   Apr. 2025 - May. 2025
- Translated official PyTorch tutorials from English to Korean and refined existing translations
- Participated in the peer review process
- Collaborated with the community via GitHub workflows

Selected Projects ──────────────────────

(In Progress) Secure Shared-RAG VLA Model Serving System   *Soongsil University*
Capstone Project   Dec. 2025 - Present
- A project developing a secure client-server system for VLA models utilizing RICL-based shared RAG
- Architected the overall serving system and system workflow
- Leading the implementation of the client-side VLA model serving pipeline, focusing on data privacy in security-sensitive environments.

PixelOn: On-device Pixel Art Generation Service ⬀   *Soongsil University*
Course Project   Sep. 2025 - Dec. 2025
- An on-device AI service project that integrates diffusion model-based pixel art generation into the open-source editor Piskel
- Searched and evaluated various open-source diffusion models and LoRAs to operate efficiently under limited hardware resources
- Successfully implemented the serving pipeline and shared the extended feature with the Piskel open-source community

ToTRM: Tree-of-Thought Tiny Recursive Model ⬀   *Soongsil University*
Course Project   Sep. 2025 - Dec. 2025
- A research project enhancing the Tiny Recursive Model (TRM) architecture by integrating the Tree-of-Thought (ToT) mechanism
- Modified the PyTorch model implementation to incorporate various branching and merging policies
- Achieved a 5%p improvement in accuracy on Sudoku Extreme dataset compared to the baseline by conducting various experiments

Awards and Scholarships ─────────────────

Scholarship for Academic Excellence   *Soongsil University*
Awarded to the top students based on the highest GPA in the department   2021, 2024, 2025 (All Semesters)

Academic Excellence Scholarship for Peer Tutoring   *Soongsil University*
Awarded to outstanding tutees for active participation and high academic achievement   2024, 2025

TOPCIT Excellence Scholarship   *Soongsil University*
Awarded for outstanding performance in TOPCIT (Level 4)   2025

1st Prize in the Software Department Creative Engineering Design Exhibition *Soongsil University*

Awarded 1st Prize for developing a creative project using Arduino 2021

Skills

Programming Python, C/C++, Java, PyTorch

Language Korean (Native), English (Intermediate)