

# Junhyeok Lee

📞 (+82)10-3757-0231 | 📩 wnsx0000@gmail.com | 🏷 https://wnsx0000.github.io/ | 💬 wnsx0000 | 💬 Junhyeok Lee

## Education

### Soongsil University

BS in Software

GPA: 4.39/4.5 | Major GPA: 4.41/4.5 | Rank: 2/128

Seoul, Republic of Korea

Mar. 2021. - Present

## Research Interests

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

## Experience

### Advanced Intelligent Computing Architecture Lab (AICA)

Undergraduate Research Intern, Advisor: Youngho Gong

Soongsil University

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)

MIT Han Lab Online Course

Online

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)

Open Source Contributor

Online

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

Capstone Project

Soongsil University

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 to ensure safe storage and sharing of Vector DB
- 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

Course Project

Soongsil University

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

Course Project

Soongsil University

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% improvement in accuracy on Sudoku Extreme dataset compared to the baseline by conducting various experiments

## Awards and Scholarships

### Scholarship for Academic Excellence

Awarded to the top students based on the highest GPA in the department

Soongsil University

2021, 2024, 2025 (All Semesters)

### Academic Excellence Scholarship for Peer Tutoring

Awarded to outstanding tutees for active participation and high academic achievement

Soongsil University

2024, 2025

### TOPCIT Excellence Scholarship

Awarded for outstanding performance in TOPCIT (Level 4)

Soongsil University

2025

**Encouragement Award in 2024 ICPC Asia Seoul Preliminary Contest**

Awarded for algorithmic problem-solving skills in the collegiate contest

*Soongsil University*

2024

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

Awarded 1st Prize for developing a creative project using Arduino

*Soongsil University*

2021

## Skills

---

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

**Language** Korean (Native), English (Intermediate)