

# ALEX YANG

## SOFTWARE ENGINEER

### CONTACT

- 📞 15760676533
- ✉️ yanglongwei06@gmail.com
- 📍 China
- 🌐 <https://github.com/Alex-wuhu>

### PROFILE SUMMARY

Software Engineer with experience at Samsung Electronics and Core Developer at Apus Network (top AI project by staking in the Arweave ecosystem).

Specialized in AI infrastructure, on-chain contract development, and model fine-tuning, with practical work on WASI-NN, llama.cpp, vLLM, sglang, and frameworks like Tinker and Slime.

### EDUCATION

2019 - 2023

**HANYANG UNIVERSITY**

- Software Engineering
- full university scholarship
- GPA: 3.4

### SKILLS

- **Programming Languages**  
C++, Python, Rust, C#, Solidity, Lua
- **AI / ML & Infrastructure**  
AI inference infrastructure, Model fine-tuning, Ray, WASI-NN, llama.cpp, vLLM, sglang
- **Systems & Confidential Computing**
  - Bare-metal GPU setup, GPU TEE deployment (production), Linux systems
- **Blockchain & Web3**
  - Smart contracts, On-chain protocol development, Arweave ecosystem

### LANGUAGES

- Chinese: native
- English: Fluent(IELTS 6.5)
- Korean: Fluent(Topic 5)

### WORK EXPERIENCE

#### Apus Network

2025.02 - Present

Core Developer & dev rel

- Core developer of the leading AI project by staking in the Arweave ecosystem.
- Worked on AI infrastructure (WASI-NN, llama.cpp, vLLM, sglang) and AI model fine-tuning using internal datasets and frameworks such as Tinker and Slime.
- Deployed NVIDIA GPU TEE environments on bare-metal systems and delivered AI inference services to production.
- Developed and maintained on-chain contracts (Lua) for token and protocol logic.
- Actively involved in developer relations, serving as an X Spaces host, speaker, and live-coding workshop instructor.
- Collaborated with multi-national teams and organizations across the Web3 and AI ecosystem.

#### Samsung Electronics

2023.07 - 2025.02

Software Engineer

- Developed TV client applications using C# and .NET.
- Delivered performance improvements through memory and CPU optimization.
- Collaborated with Baidu to build an AI agent chat application, including deployment and RAG integration on Baidu Cloud.

### PROJECTS

#### Digital Twins (with Forward Research)

- Co-developed a Digital Twin system with the Forward Research team, providing AI infrastructure and secure inference environments.
- Designed infrastructure to persist prompts and model context permanently on Arweave.
- Fine-tuned a custom model using Tinker with internal datasets.
- Deployed inference workloads in GPU TEE environments and hosted public X Spaces to present the system.

#### AI Inference Service (OpenAI-Compatible)

- Built and deployed an OpenAI-compatible AI inference service powered by sglang.
- Implemented deterministic inference and GPU TEE-protected execution on bare-metal systems.
- Integrated on-chain contracts to meter usage and calculate inference costs.