Stuart Sul

+1 (650) 407-8124 | Stanford, CA | stuartsul@stanford.edu | stuartsul.com | github.com/stuartsul | linkedin.com/in/stuartsul

Education

Stanford University
09/2024 - 06/2026 (expected)

M.S. in Computer Science — Stanford, CA

- Recipient of the merit-based Kwanjeong scholarship ($60,000)

Seoul National University
03/2016 - 08/2022

B.S. in Computer Science and Engineering — Seoul, South Korea

- GPA: 4.21/4.3 (class rank: **1st**)

Work Experience

Software Engineer
01/2024 - 04/2024

Linq (Techstars '23) — Cambridge, MA (remote)

- Increased Infrastructure as Code (IaC) coverage from **0% to 95%** for existing AWS resources using Terraform.
- Optimized monitoring of customers' large language model (LLM) usage by implementing the Elasticsearch-Fluentbit-Kibana (EFK) stack, Prometheus, and Grafana, achieving a 95% automatic failure recovery.

Co-Founder and CTO
07/2021 - 08/2023

Blux (formerly Z.Ai) — Seoul, South Korea

- Co-founded and grew Blux to serve **10M+ monthly users** with $1M+ in annual revenue; Blux is an AI SaaS startup specializing in real-time recommender systems for e-commerce.
- Secured $1.7M in funding and received a $10M valuation in South Korea.
- Designed a Kubernetes-based e-commerce recommender engine that handles 1B+ requests monthly and trains/deploys 100+ ML models daily (featured in the official AWS Blog).
- Developed and documented the Blux SDK in Python, Java, Kotlin, JavaScript, TypeScript, PHP, and Swift.
- Achieved 90% automation of SaaS product delivery by developing microservices for customer data collection and recommender model serving using FastAPI and gRPC.
- Recruited and led an agile team of 15+ engineers ranging from juniors to a university CS professor.

Research Assistant
06/2020 - 09/2020

Architecture and Code Optimization Lab, Seoul National University — Seoul, South Korea

- Reduced main memory access by 44% in commercial deep learning recommendation models (DLRMs).
- Designed and implemented a novel embedding clustering algorithm for DLRM inference in C++ (GitHub).
- Co-authored an academic paper accepted at ASPLOS 2021, a top-tier CS systems conference (paper).

Research Assistant
12/2019 - 02/2020

Music and Audio Research Group, Seoul National University — Seoul, South Korea

- Implemented an audio processing architecture using CNN, cGAN, super-resolution, and the Griffin-Lim algorithm, contributing to a commercial singing voice synthesis project.
- Led a team of 4 assistants in labeling audio and MIDI data with millisecond-precision using Logic Pro and Python, preparing 30+ songs for model training.

Selected Projects

SampyoNet (Python; GitHub): Reduced inspection time by 70% in concrete manufacturing by developing a TensorFlow-based deep learning model for gravel quality assessment; adopted by Sampyo, a leading cement manufacturer in Korea.

LLVM Compiler Optimization (C++; GitHub): Optimized LLVM compiler using C++ to enhance assembly code performance; secured 2nd place among 13 teams in an LLVM optimization competition at Seoul National University.

Selected Skills

Languages: Python, Java, C, C++, JavaScript, TypeScript, HTML, CSS, SQL, Shell

Libraries & Frameworks: FastAPI, Flask, React.js, Redis, PostgreSQL, NumPy, Pandas, Spark, TensorFlow, PyTorch

DevOps: Linux Administration, Docker, Kubernetes, Helm, Karpenter, AWS, Azure, Terraform, Jenkins, Argo CD, Apache Airflow, MLflow, Elasticsearch, Logstash, Fluentbit, Kibana, Prometheus, Loki, Grafana, Milvus, TensorFlow Serving