

Soliev

📍 Seoul, South Korea ✉ soliev.dev@gmail.com ☎ 010-2204-9798 🌐 soliev.dev 🌐 soliev 🌐 soliev01

Education

BS **Kyungpook National University**, Computer Science Daegu, South Korea
Sept 2017 – Feb 2022

Experience

Urban, Fullstack AI Engineer (Agentic Systems) Seoul, South Korea
Dec 2025 – present
4 months

- Developing autonomous AI agents using LangGraph and CrewAI to automate complex data analysis workflows.
- Architecting multi-agent systems for high-reliability infrastructure, handling 2M+ monthly API requests.
- Implementing advanced RAG pipelines with hybrid search and reranking to improve context retrieval accuracy.

Hoca Inc., Flutter & AI Developer Seoul, South Korea
Nov 2023 – Nov 2024
1 year 1 month

- Integrated LLM-based chat interfaces into mobile applications, utilizing function calling for local device actions.
- Developed cross-platform mobile app with 100K+ downloads, incorporating AI-driven user personalization.
- Optimized mobile-to-cloud AI inference latency for real-time agentic responses.

ITSROOM, Frontend Developer & AI Engineer Ulsan, South Korea
July 2022 – Oct 2023
1 year 4 months

- Built an AI-powered smart factory dashboard with integrated RAG for technical manual querying.
- Developed specialized "expert" agents to monitor real-time sensor data and generate automated maintenance reports.
- Designed intuitive React-based UIs for interacting with complex AI reasoning chains and decision logs.

Skills

AI Agents & RAG: LangChain, LangGraph, CrewAI, Function Calling, Vector Databases (Supabase/Chroma), RAG Optimization

Fullstack Tech: React, TypeScript, NestJS, Supabase, PostgreSQL, Python

Mobile & AI: Flutter, Dart, OpenAI API, Anthropic SDK, TensorFlow Lite

Infrastructure: Kubernetes, Docker, Git, CI/CD, AWS, Distributed Systems

Projects

LLM Observability & Evaluation Dashboard Remote
Jan 2025

A full-stack observability platform for monitoring AI agent reliability, cost, and accuracy in production.

- Built a high-performance logging backend using **FastAPI** and **Supabase** to track agent traces, tool calls, and per-request latency.
- Developed an automated "LLM-as-a-Judge" pipeline that scores agent responses for faithfulness and relevance using GPT-4o with structured reasoning.
- Designed an interactive analytics dashboard with **React** and **D3.js** to visualize the correlation between latency, token cost, and quality scores.
- Implemented real-time alerting for "hallucination triggers" when evaluation scores drop below a defined threshold (0.7/1.0).

Agent Inventory Manager

An autonomous storefront agent capable of managing inventory and processing orders via tool-calling.

- Engineered a **NestJS** and **LangChain** architecture allowing an LLM to interface directly with a PostgreSQL database via secure function calling.
- Reduced manual inventory updates by 60% by deploying an autonomous "Supervisor" agent to reconcile stock levels and generate restock alerts.
- Utilized **Supabase RLS** to ensure secure data access and isolation between different storefront tenants.