JOSHUA KO

Santa Clara, CA | joshua.ko.jko@gmail.com | in/joshua-ko-jko | github.com/joshuakojko | joshuako.dev

EDUCATION

San Jose State University   San Jose, CA
B.S. Computer Science   ***Expected Graduation: Dec 2026***
- **Coursework**: Computer Systems, Data Structures and Algorithms, Object-Oriented Design, Computer Architecture
- **Organizations**: CS Club Backend Dev, SCE Dev, ML Club Project Member

EXPERIENCE

Software Engineering Fellow   July 2024 – Sept. 2024
Headstarter AI
- Integrated Firebase Firestore with Next.js to deploy a real-time pantry management application on Vercel
- Utilized zero-shot prompting via Llama 3.1 model for recipe generation feature based on user-selected ingredients
- Developed an AI Flashcard Generator application using Next.js, shadcn/ui, Firestore, Clerk, and OpenAI API
- Defined task-oriented prompting with OpenAI structured outputs to generate accurate flashcards from source docs

Software Engineering Intern   June 2024 – Aug. 2024
Software & Computer Engineering Society
- Built a FastAPI server to stream YouTube videos with RTMP to a Raspbery Pi receiving the stream with VLC Player
- Optimized video cache management for dockerized streaming server, implementing preemptive cache downsizing
- Added custom Prometheus metrics to monitor server performance, download speed, and cache performance
- Aggregated Prometheus metrics using PromQL queries into Grafana dashboard for real-time data visualization
- Refactored, tested, and secured printer API endpoint, implementing server-side validation to enforce print limits

Undergraduate Research Assistant   Sep. 2023 – May 2024
Engineering Success Program
- Led a collaborative research project investigating current application and limitations of AR/VR/XR technology in university education through the Meta Quest 3; presented research findings to faculty in end of year symposium

PROJECTS

Car Dealership Voice Agent - Headstarter Hackathon Finalist | *Next.js, MongoDB, Deepgram, Twilio, OpenAI*
- Utilized a Generative AI phone call toolkit with Twilio Media Streams to create an AI car dealership voice agent
- Used Deepgram for STT and TTS and OpenAI for chat completion with tool calling to schedule appointments
- Jointly developed an analytics dashboard for call logs, metrics, and scheduled appointments with Recharts

Patent Mate - Stanford LLM x Law Hackathon 4 | *Next.js, React, TypeScript, LlamaIndex, Pinecone, OpenAI, Groq*
- Co-developed Patent Mate alongside devs and IP lawyer to streamline legal patent prefiling service with LLMs
- Leveraged Groq API for patentability criteria analysis and comprehensive patenteability report generation
- Utilized LlamaParse for patent ingestion, OpenAI for embeddings, and Pinecone for vector storage in RAG pipeline
- Performed similarity search against Pinecone database to identify prior art and provide attorney recommendations

Ginder (Open Source Recommender Chatbot) | *Next.js, React, TypeScript, Drizzle, Flowise, Cloudflare (Workers, AI)*
- Led frontend development for Ginder, a personalized chatbot enabling developers to discover Open Source repositories based on their GitHub repository metrics, skills, and interests.
- Configured GitHub OAuth to authenticate repository fetching for analysis, summarization, and vector embedding with Cloudflare Vectorize, enabling accurate Open Source project recommendations
- Collaborated on LLM orchestration flow using Flowise for personalized recommendations and integrated WebSocket protocol for low-latency, real-time response streaming

SKILLS

Languages: Java, Python, JavaScript, TypeScript, SQL
Frameworks: React.js, Express.js, Next.js, Hono, Flask, FastAPI
Developer Tools: Git, Docker, Postman, PostgreSQL, Pinecone, SQlite, Firebase, Prometheus, Grafana, Vercel