

Weibo Zhang

me@weibozhang.com | weibozhang.com | [GitHub](#) | [LinkedIn](#)

EXPERIENCE

Software Developer

Jan 2025 – Present

Bluon

- Bluon's **RAG** chatbot struggled with **70%** retrieval relevance from unstructured transcript data in **Pinecone**; processed **50k+** call recordings through a multi-stage pipeline using **OpenAI** transcription, LLM filtering, and structured **JSON** summarization with field tags to improve retrieval relevance to **92%**, validated by technician review
- Third-party cookies were blocked across partner iFrames and mobile WebViews, breaking user authentication; built **JWT**-based auth using the **PostMessage API** for cross-origin token exchange, migrating to **Zustand** client state with bearer tokens
- Architected a **PostgreSQL** schema supporting **200k+** chats with relational tables for messages, parts, tool calls, AI responses, shared chats, and user feedback; built **REST APIs** and analytics endpoints
- Designed and built AI tool-calling with **AI SDK** and streaming, integrating Bluon's existing APIs as callable tools for model lookup, **Pinecone** manual retrieval, image uploads, and file uploads; implemented tracking for tool call counts and user feedback to monitor usage and improve reliability
- A breaking change in the **AI SDK** altered how chats and messages were stored; led the production database migration, building type-safe conversion and parsing with **Zod** to ensure zero data loss

Software Developer Intern

June 2023 – Dec 2024

Bluon

- Worked under the SVP of Technology on AI and machine learning projects for Bluon's HVAC tech support platform serving **10k+** monthly active users
- Designed UI/UX prototypes in **Figma** for the new Bluon.com and developed an OCR HVAC nameplate reader

Co-founder & Web Developer

Oct 2022 – June 2023

Moonwish Studios

- Co-founded a web development studio, shipping **8** production websites for local businesses and generating **\$5k+** in revenue
- Built mobile-responsive apps with **React**, **TypeScript**, **Next.js**, **Tailwind**, and **Supabase**; established a **Figma** design system that cut design cycles to **3–4 days**
- Owned end-to-end client relationships from scoping and contracts through delivery and iteration

PROJECTS

miaomo.ai — AI Research Assistant

In Progress

- Building a NotebookLM-style research tool with RAG-based document Q&A using **Next.js**, **TypeScript**, vector database, and LLM APIs
- Implementing semantic search and context-grounded responses as part of a university team project

TECHNICAL SKILLS

Languages: TypeScript, JavaScript, Python, SQL

Frameworks: React, Next.js, Node.js, Tailwind CSS, AI SDK

Databases & Infrastructure: PostgreSQL, Supabase, Pinecone, Zustand

Tools: Git, GitHub, Vercel, Figma, Zod, REST APIs

AI/ML: OpenAI API, Anthropic API, RAG pipelines, prompt engineering, streaming, model evaluation

EDUCATION

California State University, Long Beach — B.S. Computer Science

Aug 2022 – May 2026

Relevant Coursework: Data Structures & Algorithms, Databases, Operating Systems, Distributed Systems, Software Engineering