

Junkai Zhao

6418192530 | lanbodikai@gmail.com | [LinkedIn](#) | [GitHub](#) | [lanbodikai.com](#)

EDUCATION

UNIVERSITY OF CALIFORNIA, BERKELEY

Berkeley, CA

B.S. Industrial Engineering & Operations Research, Minor in EECS

May 2027

Relevant Coursework: Data Structures; Algorithms; Software Engineering; Operating Systems; Database Systems; Networks

TECHNICAL SKILLS

Languages: Python, Java, SQL, C, JavaScript, TypeScript

Web & Backend: React, Next.js, FastAPI, Node.js, PostgreSQL, Redis, SpringBoot

AI & Cloud: RAG, LLM APIs, structured generation, STT/TTS, GCP, AWS, Docker

Systems & Tools: Git/GitHub, GitHub Actions, Kotlin/Jetpack Compose, BLE, WebSockets

WORK EXPERIENCE

RISELINK | [Product Demo](#)

Remote/Sacramento, CA

Developer

Jul 2026 - Present

- Owned development of an Android companion app for an AI recording device, building BLE-to-Wi-Fi/FTP connectivity, recording controls, audio synchronization, and transcript workflows in Kotlin and Jetpack Compose.
- Designed a GCP transcription pipeline with Cloud Run, Cloud Tasks, Cloud Storage, Firestore, and Speech-to-Text; added a fast path for short recordings that reduced transcription latency by **~50%** in testing.
- Designed a resilient BLE to Wi-Fi hotspot to FTP connection flow around device constraints, adding retry and timeout handling to prevent stalled connection states.
- Worked directly with hardware and test engineers in Shenzhen, shipping APK iterations and debugging device integration across physical Android devices.

BERKELEY LAW LIBRARY

Berkeley, CA

IT / Software Assistant

Nov 2025 - Present

- Adapted Python/Beautiful Soup scraping pipelines across **200+** university faculty sites with differing HTML structures, reducing extraction work to seconds that otherwise required ~2–5 minutes of manual parsing per site.

NEURALSEEK | [Coding Agent Demo](#)

Remote, US

AI Agent Builder Intern

Mar 2026 - May 2026

- Built an AWS-hosted coding-practice agent with RAG over task presets, code execution and output evaluation, automated debugging feedback, and conversation-context compression for longer sessions.

SOMEIDEA AI | [Financial AI Demo](#)

Berkeley, CA

Full Stack Engineer Intern

Jan 2025 – Apr 2025

- Built a Llama 3.1-8B query parser that normalized vague financial questions into schema-validated JSON fields for company, date, and requested data, improving reliability of downstream database search.
- Developed a GPT-4o processing pipeline for generating structured balance sheets and cash-flow statements from financial data with schema validation and input normalization.

PROJECTS

MOUSEFIT STUDIO | TypeScript, Next.js, FastAPI, PostgreSQL | [Mousefit.pro](#)

- Built and deployed a full-stack gaming-mouse recommendation platform with a 434-model catalog collected from multiple sources, ranking products by hand size, grip style, shape, weight, and connectivity.
- Prototyped MediaPipe-based hand/grip analysis, then simplified the UX after testing exposed reliability issues in CV-based grip detection; **~80% testers** were satisfied with the resulting recommendations.

VOICE AI DEVICE PLATFORM | Python, C, WebSockets, Deepgram, GPT-5-Nano, Oracle Cloud | [Web-Demo](#)

- Built a streaming STT to LLM to TTS pipeline for a BK7258-based voice device, sending 16 kHz PCM audio over WebSockets for real-time speech processing and on-device responses.
- Reduced end-to-end response latency from **~7s to ~3.5s** by streaming 20 ms audio chunks to STT instead of waiting for complete recordings.