

JASON YI

408-706-6183 | jasonyi2023@gmail.com | [linkedin.com/in/jasonyi33](https://www.linkedin.com/in/jasonyi33) | github.com/jasonyi33 | jasonyi.ai

EDUCATION

University of California: Berkeley | GPA: 4.00

Berkeley, CA | Grad: May 2028

B.S. in Electrical Engineering and Computer Sciences | Dean's List

Activities: UpSync Tech Consulting, Generative AI @ Berkeley, AI Entrepreneurs @ Berkeley, Codex Student Lab @ OpenAI

Coursework: Machine Learning, Neural Networks, Operating Systems, Internet Architecture & Protocols, Computer Architecture

PROFESSIONAL EXPERIENCE

TikTok | *Software Engineer Intern*

San Jose, CA | May 2026 - Aug 2026

- Built TypeScript/React monetization dashboard modules for 50K+ PGC creators to track rewards, subscriptions, and revenue.
- Integrated Go REST APIs with client-side data models and validation, reducing subscription analytics load latency by 18%.
- Implemented deep-link routing, state handling, and analytics logging for Creator Rewards points, raising tool navigation by 22%.

Adobe | *Machine Learning Engineer (Contract)*

San Jose, CA | Feb 2026 - May 2026

- Built Python graph-RAG research system processing 500+ PDFs with hierarchical ingestion and LLM-based triple extraction.
- Implemented graph traversal retrieval across 20K+ entities to support multi-hop reasoning over domain-specific documents.
- Designed retrieval evaluation harness against baseline RAG, reducing token usage by 32% and improving accuracy by 18%.

Google | *Software Engineer (Contract)*

San Francisco, CA | Sept 2025 - Dec 2025

- Built full-stack AI interview platform with Next.js, Python, Gemini, and 15+ REST APIs for candidate assessment workflows.
- Designed modular evals layer with 30+ scoring modules, API contracts, and Langfuse tracing, cutting debugging time by 40%.
- Shipped document parsing and structured extraction pipeline with validation checks, reaching 95%+ accuracy on 500+ PDFs.

VoiceBit | *Software Engineer Intern*

Berkeley, CA | Sept 2025 - Dec 2025

- Built FastAPI/PostgreSQL analytics dashboards modeling call outcomes & revenue trends for 220+ restaurant operators.
- Recovered 12% of lapsed restaurant customers with RAG-based SMS win-back engine using churn triggers & campaign analysis.
- Shipped Twilio campaign APIs with cohort segmentation, message validation, & delivery tracking, raising pilot retention by 14%.

Avantree | *Software Engineer Intern*

San Jose, CA | May 2025 - Aug 2025

- Built React/FastAPI review analytics platform for 70+ products, cutting workload by 92% for 15+ PMs and saving \$250k annually.
- Designed customer-review data pipeline with CRM sync, preprocessing, validation, & hourly API ingestion across 90K+ reviews.
- Fine-tuned LLM classifier with statistical evaluation, reaching 97% accuracy and improving issue response speed by 40%.

UC Santa Barbara | *Machine Learning Research Intern*

Santa Barbara, CA | June 2024 - Aug 2024

- Benchmarked CNNs, MLPs, and SNNs in PyTorch/SNNtorch to study accuracy-latency tradeoffs for audio classification.
- Built MFCC ML pipelines with model profiling and optimization, reducing inference latency by 37% on 10K+ audio samples.
- Evaluated model performance with accuracy, latency, and confusion-matrix analysis; presented findings at ICOPH Bangkok.

PROJECT EXPERIENCE

VoiceReach - Backed by OpenAI | *Co-Founder, Technical Lead* | [Link](#)

July 2025 - Oct 2025

- Built React Native voice-first field app with FastAPI backend, Whisper transcription, and GPT-4o extraction for 400+ users.
- Increased documentation efficiency by 78% by automating transcription, structured case-note generation, and location tagging.
- Improved report triaging by 53% with AI-generated summaries, real-time search, consent-based photos, and priority filters.
- Deployed with SF NGOs, won 1st/300+ and \$5K OpenAI startup grant, acquired by City of SF (Dept for Emerging Technologies).

Prereq - 2nd @ TreeHacks 2026 | *Hackathon Team Lead, Full-Stack Engineer* | [Link](#)

Feb 2026

- 2nd out of 1000+ (Education Track), awarded \$1,000, for building a live AI lecture copilot with personalized knowledge graphs.
- Reduced live latency to sub-500ms by designing a WebSocket push architecture with Redis caching and PostgreSQL backend.
- Built full-stack system with Next.js, Express, and Zoom RTMS to transcribe lectures, extract graphs, and update mastery maps.

MediSnap - 2nd @ CalHacks 2025 | *Hackathon Team Lead, Back-End Engineer* | [Link](#)

Oct 2025

- 2nd out of 2000+ (Snap AR Track), awarded \$2,000, for an AR medical assistant combining voice and AI clinical reasoning.
- Built scalable Express backend, integrating Snap ASR, Gemini, and stateful memory to deliver sub-1500ms clinical workflows.
- Engineered custom MediaPipeHands CV pipeline with pressure estimation to enable anatomical overlays and technique feedback.

SKILLS AND INTERESTS

Languages: Python, Java, C++, C, Swift, JavaScript, TypeScript, Golang, SQL, Rust, HTML/CSS

Frameworks: React.js, Next.js, React Native, Node.js, Express.js, FastAPI, Flask, Spring Boot, SwiftUI, UIKit, PostgreSQL

ML/Data: PyTorch, TensorFlow, Hugging Face Transformers, RAG, scikit-learn, NumPy, Pandas, LangChain, Langfuse

Systems/Tools: Docker, AWS, Kubernetes, Git, Linux, Terraform, CI/CD, Redis, OAuth, WebSockets, REST APIs, MicroServices

Interests: Weightlifting, Running, EDM/House Music, Sci-Fi/Fantasy, Sitcoms, Basketball, Football, Dogs, Malatang, Chipotle