

Brian Wang

88brianw@berkeley.edu | github.com/brianBoolean88 | linktr.ee/briantruefalse

EDUCATION

University of California, Berkeley

B.S. Electrical Engineering & Computer Sciences (EECS)

Berkeley, CA

Expected Dec 2027

- **GPA:** 3.73
- **Relevant Coursework:** CS 61A, CS 61B, CS 61C, CS 70, EE 64, EE 66, ENGIN 125

EXPERIENCE

BlueRobins (Skydeck Startup)

May 2026 – Present

Software Engineer

Remote

- **AI:** Engineered an AI retrieval system using PostgreSQL (pgvector) and custom RPC functions for real-time cosine similarity search across input embeddings for an AskRobins helper bot. Raised automated answers from 36% to 80% by reworking embeddings.
- **Mobile:** Migrated the company's React web app into a React Native/Expo mobile application (100k+ lines), adapting UI & data flows for mobile to connect mentors & students on our project-based education platform.
- **Infrastructure:** Hardened production security via AWS secret rotation, Supabase Row-Level Security (RLS), and isolated CRON/JWT auth, reducing credential exposure across 200+ real customers.
- **Incident Response:** Diagnosed & resolved a production outage by tracing AWS Route 53, EC2, ECS, ALB & Supabase Disk I/O, restoring service in ~30 minutes and preventing recurrence via optimization & Supabase Realtime.

Berkeley AI Research Labs (BAIR)

August 2025 – June 2026

MobileCybench Research Intern

Onsite

- Researched LLM-driven cybersecurity vulnerability detection benchmarking on Android Emulators under Prof. Dawn Song's Lab (Partnered with Stanford SAIL) w/ 30+ researchers & 5+ industry/graduate leads
- Injected synthetic vulnerabilities (CVE-2024-44080) in sandboxed Jitsi Meet APK, automated exploit workflows in Prosody shell for user/room creation, and improved CI/CD pipelines (error-logging, Windows/WSL compatibility)

Virtual Medical Missions

July 2024 – August 2025

Backend Developer & CS Lead

On-site

- Engineered Emergency SOAPNote system (Quart/Flask, PostgreSQL, Bootstrap HTML) for VMM's clinical partner network, serving Breathe Clinic Corona.

Agile Studios

July 2022 – Jan. 2026

Roblox Indie Developer Lead

Hybrid

- Directed a team of 10+ developers building state-driven NPCs & physics-based systems for Project Agile to reach 60K+ YouTube views.

PROJECTS

Empath AI: Fine-Tuned Text-Generation Model Therapist | *Mistral-7B, FlashAttention 2, Next.js, FastAPI*

- Served the fine-tuned model behind FastAPI on a full-stack NextJS application, manually removed Triton references in FlashAttention2 Library for Windows Compatibility
- Fine-tuned Mistral-7B on a set of 8,500+ curated emotional-text samples, accelerating training 4x via FlashAttention 2, GPU memory optimization (CUDA), batching, PyTorch Inference, BitsAndBytes Quantization

UnstableREM: LLM Tool-Calling Pipeline for Procedural World Generation | *OpenRouter, MediaPipe, Lua*

- Architected an OpenRouter LLM tool-calling pipeline translating real-time eye-tracking data (MediaPipe) into procedurally generated Roblox environments; open-sourced the project, gaining 3,000+ YouTube views

CVision: Serverless Resume Screening | *Next.js, FastAPI, AWS Lambda/S3/DynamoDB*

- Architected a serverless, full-stack resume screening platform on AWS, utilizing S3 Bucket, DynamoDB, and Lambda

TECHNICAL SKILLS

Languages: Python, Java, C++, JavaScript, SQL, Dart, LuaU, Shell

ML/Infra: PyTorch, Mistral-7B fine-tuning, FlashAttention 2, quantization, batched inference, pgvector/vector search, BitsAndBytes, MediaPipe

Frameworks/Services: React, Next.js, Flask, FastAPI, AWS, Supabase, Firebase, Algorithms And Data Structures

Tools: Git, Unit Test Coverage, Pull Request Development, Docker, Agentic Workflows, Git Bash

Other: REST API design, Software Testing, Google Certificates (Cybersecurity, Python, AI, IT)