CAPTAIN YAO

857-334-5758 | cz.ycaptain@gmail.com | github.com/ycaptain | **Available for full-time SDE roles starting Apr 2026**

🎓 EDUCATION

Northeastern University, San Jose, CA — Sep 2024 – Apr 2026 (Expected)
Master, Computer Science

University College Dublin, Dublin, Ireland — Sep 2015 – Jun 2019
Bachelor, Software Engineering

👥 EXPERIENCE

NinjaTools.ai San Francisco Bay Area — Mar 2024 – Present
Co-Founder & CTO   https://ninjatools.ai/

- **Business Impact:** Reached 3k+ paying users in first 3 months; scaled platform to handle 50k+ API calls/day with 99.9% uptime.
- **Multi-Model Orchestration:** Built a provider-agnostic layer integrating OpenAI, Anthropic (Claude), Google (Gemini), DeepSeek, Grok, and Perplexity via unified adapters; added structured output, tool/function calling, SSE streaming, retries/backoff, and cost/latency-aware routing with graceful failover.
- **Generative Suite:** Shipped AI Doc / chat with artifacts, image generation (Replicate FLUX-dev / OpenAI Images), video (Veo3, LTX Video, Kling), music (MusicGen), mindmap creator, and Chat-with-YouTube (transcript ingestion + chunking + RAG).
- **RAG & Search:** Implemented embeddings with pgvector on Supabase; hybrid search with citations and cache-first retrieval; Perplexity fallback when confidence is low to sustain answer quality.
- **Platform & Infra:** Next.js (App Router) + shadcn/ui + Tailwind on Vercel (serverless/edge); Supabase (Auth, Postgres, Storage, RLS); BullMQ + Redis workers; ffmpeg media pipelines; CRON/webhooks; PostHog analytics with feature flags/A/B tests.
- **Payments & Growth:** Stripe subscriptions with usage metering, credit wallet, and rate limiting; AppSumo code activation & entitlement sync; launch reached *#1 on AppSumo homepage* (Lifetime Deal).
- **Security & Compliance:** OAuth, JWT, RBAC, audit logging, PII redaction, secrets rotation, and data-residency options for enterprise workspaces.

TikTok Beijing, China — Jun 2018 – Jan 2024
Intern → L5 Senior Software Engineer   Web Infrastructure & PICO XR Teams

- **Career Progression:** Joined as intern (2018) → full-time (2019) → L4 team lead (2020) → L5 team leader (2022), managing 8 engineers. Consistently rated ″**Exceeds Expectations (E/E+)**″ in annual reviews(top 5% in org of 100+).
- **Modern.js Framework:** Spearheaded Modern.js, a meta-framework widely adopted in TikTok web infrastructure. Delivered enterprise scaffolding, state-management integration, and runtime optimizations; project achieved 4.8k+ GitHub stars.
- **Low-Code Editor Platform:** As a Web Infra team leader, built a low-code visual editor in Modern.js, NestJS, and Node.js; containerized services with Docker and deployed on Kubernetes (HPA, rolling/canary, resource limits, centralized logging/tracing). Drag-and-drop components, schema-driven config, and live preview cut feature delivery time by 40%.
- **AI Acceptance Platform:** Built an AI-powered validation platform with React, WebGL, and Golang; reduced design-to-dev review cycles by 30% via real-time feedback and automated testing.
- **Low-Code WebXR Engine:** Led a team of 8 engineers to build a low-code WebXR engine (Next.js, WebGL, NestJS) enabling rapid performance evaluation across VR, desktop, and mobile; significantly improved validation efficiency of multiple rendering engines.

⚙️ SKILLS

- **Languages:** TypeScript, JavaScript, Python, Golang, Java, C++, Objective-C
- **Frameworks:** Node.js, React, Next.js, NestJS, Prisma, Flask, Gin, Spring Boot, GraphQL
- **Infra/Tools:** Kubernetes, Docker, Git, Redis, BullMQ, RabbitMQ, Kafka, Pnpm, Supabase, FFmpeg, PostHog