



JOSHUA KAPPLER

San Francisco, CA • 512-210-9934 • Joshua.Kappler@gmail.com

github.com/JoshKappler • joshuakappler.com • linkedin.com/in/josh-kappler • youtube.com/boffy

AI engineer building autonomous agents from scratch with direct SDK calls, no frameworks like LangChain. Full-stack TypeScript and Python with multi-provider LLM orchestration. Grew a YouTube channel to 2.1M subscribers before moving to engineering full-time.

PROJECTS

- **memo-engine** (private, NDA): AI deal analysis and investment memo platform for a private credit investment firm. Ingests messy deal documents (PDFs, Excel models, Word drafts, .msg Outlook emails), runs **contextual retrieval** over pgvector (1024-dim) with per-chunk Sonnet prefixes cached via **ephemeral prompt caching**, produces institutional-grade memos with **every claim cited back to source**. Durable pipeline orchestration via **Vercel Workflow DevKit**, forced `tool_use` + Zod for ~40-field structured extraction. Next.js 16, React 19, Neon Postgres, Anthropic SDK, Voyage AI embeddings, Puppeteer PDF export.
- **autohack**: Autonomous vulnerability hunter across HackerOne, Immunefi, Huntr, Bugcrowd, and Intigriti. **12-state findings pipeline** with 60-minute hunts, **adversarial self-review** filters false positives before submission, cross-hunt memory records outcomes for learning. TypeScript monorepo, Next.js 15, SQLite + Drizzle, PTY-based Claude spawning, tRPC.
- **property-leads** (private client): Autonomous lead-finding pipeline for a real-estate cash buyer. **4-agent chain** on hourly Vercel cron: Haiku orchestrator, Sonnet research and scoring, Sonnet writer gated by a Haiku reviewer. Resend outreach at **\$0.22 per 33-property run**. Next.js 16, Neon Postgres, Anthropic SDK.

EXPERIENCE

Independent AI Engineer

2024 - Present

Self-employed, San Francisco

- Ship autonomous AI agent systems end to end. Direct SDK calls over framework abstractions. State machines for entity lifecycles. Learning loops that record outcomes and feed them back into future runs.
- Build multi-provider LLM orchestration: Anthropic API and Claude Agent SDK, OpenAI SDK, Groq, OpenRouter, with fallback routing, prompt caching on stable system prompts, and `tool_use` / function calling for structured output.
- Monorepo architecture with npm/pnpm workspaces, tRPC APIs, Drizzle + SQLite/Postgres, real-time dashboards. Validation at boundaries with Zod (TS) and Pydantic (Python).

Founder & Content Creator, YouTube (Boffy)

2019 - Present

Independent digital media brand, 2.1M subscribers, 270M+ lifetime views

- Grew channel from zero to 2.1M subscribers and 270M+ views across 136 long-form videos covering gaming, hardware, and consumer tech.
- Hired and directed a team of 3 editors and designers through the full content lifecycle from scripting to post-production.
- Negotiated and delivered brand partnerships with **RedMagic**, **Wargaming**, **GeoGuessr**, **YouTooz**, **Factor**, **GamerSupps**, and others, managing outreach, contracts, scripted integrations, and deliverables.
- Built and operated a branded e-commerce storefront around content drops. Ran platform ops for a commercial Minecraft server.

SKILLS

- **Languages:** TypeScript, Python, JavaScript, Bash, HTML/CSS
- **AI / LLM:** Anthropic SDK (API + Claude Agent SDK), OpenAI SDK, Gemini, Groq, OpenRouter, Ollama, `tool_use` / function calling, prompt caching, multi-provider routing
- **Backend & Data:** Next.js 16, React 19, FastAPI, tRPC, Node.js, PostgreSQL, SQLite, Drizzle ORM, Zod, Pydantic, SSE streaming, WebSockets
- **Testing & Tooling:** Playwright, Vitest, pytest, pnpm / npm monorepos, Turbopack, Tailwind CSS v4, shadcn/ui, ffmpeg, faster-whisper, PTY spawning