

JOE KRACZ

Phone: (215) 718-7124 **Email:** joseph.kracz@gmail.com **Website:** www.joekracz.com

Full-stack engineer specializing in AI-native products: LLM agent architecture, sandboxed execution, and production infrastructure. Most recently Director of Engineering at an early-stage startup, owning both code and technical direction.

EXPERIENCE

Pull Systems **Palo Alto, CA 11/2023 – Present**

Director of Engineering (Dec 2025 – Present) | Software Engineer (Apr 2024 – Dec 2025) | Product Manager (Nov 2023 – Apr 2024)

- Led full-stack platform architecture pivot post-CTO departure, contributing to 3 enterprise wins in 5 months
- Architected multi-tenant AI agent platform (TanStack Start, Convex, Claude Sonnet) with streaming chat, tool-calling, and persistent investigation workflows, enabling quality engineers to run natural-language diagnostics on fleet data
- Implemented sandboxed agent code execution (Docker + Daytona) with per-tenant isolation, artifact capture, and controls
- Built artifact pipeline enabling "ask, run code, then visualize" workflows with file capture and chart/table rendering
- Productized agent workflows into scheduled automations with lifecycle management, timeout handling, and cancel/resume support for reliable recurring analyses
- Added production AI observability via Langfuse, and PostHog to track model/tool latency, token costs, and failure modes
- Designed Terraform-based self-hosted enterprise deployment on AWS for secure, customer-owned installations
- Built vehicle event analysis system with millisecond-level downsampling and multi-vehicle overlay, used across a fleet of 200K+ vehicles, enabling the identification of a critical fault that avoided \$3M in downstream costs
- Implemented statistical cluster recommendation engine grouping vehicles by fault and telematics history to surface actionable patterns for quality engineers

Independent Technical Consultant **Remote 2023 – Present**

- Advising enterprise clients on LLM integration, agent architecture, and production AI deployment
- Freelance web development for small businesses and startups across Next.js, Astro, and Webflow

Founder — Cause Compass **Sunnyvale, CA 05/2023 – 11/2023**

- Built LLM-powered validation pipeline (OpenAI API) to scrape, verify, and structure 1.5M+ nonprofit records at scale
- Developed full-stack MVP in Next.js + TypeScript with automated scraping, enrichment pipelines, and AI-generated data

Senior Technical Product Manager — Reach Financial **Remote 08/2022 – 03/2023**

- Defined the strategy and roadmap for a new, all-purpose lending product
- Conducted loan application A/B tests, increasing bank account linkage conversion by 15%

Product Manager — Even Financial **New York, NY 04/2021 – 08/2022**

- Launched a marketplace integration that yielded \$300K within the first quarter of launch
- Developed and marketed a debt optimizer API product for fintech businesses

Product Manager — Azlo **Birmingham, AL 08/2019 – 04/2021**

- Launched a credit card program to 5,000+ beta customers in less than four months
- Built lending referral partnership programs driving over 8,000 applications for Paycheck Protection Program loans

Software Engineering Intern — Alphadyne Asset Management **New York, NY 06/2018 – 05/2019**

- Designed and built a relational reference data management system that enabled users to access and manipulate trade data
- Implemented data pipelines to integrate with a new financial management software with the existing system

Software Engineer at Con-Edison **New York, NY 06/2017 – 04/2018**

- Saved the company millions in potential damages by implementing a data pipeline and alert algorithm around IoT sensor data

EDUCATION

NEW YORK UNIVERSITY – B.S. Computer Science, Minors in Finance & Math **New York, NY 2015 – 2019**

SKILLS

Languages: TypeScript, Python, SQL

Technologies: React, Next.js, SvelteKit, GraphQL, AWS, Convex, TanStack Start, OpenTelemetry, Docker, Terraform

AI/LLM: Tool Design, Prompt Engineering, RAG, MCP Integration, Langfuse, Anthropic API, OpenAI API, Vercel AI SDK, OpenRouter