

SUMMARY

AI systems engineer building production-grade agent harnesses and LLM products across orchestration, tool use, sandboxed execution, context and memory, evals, observability, and human-in-the-loop controls. Ships TypeScript, Python, and Rust systems against explicit reliability, latency, cost, and security constraints.

EDUCATION

Northeastern University, San Jose, CA

Sep 2022 - Aug 2025

Master of Science in Computer Science (GPA: 4.0/4.0)

University of California, Irvine, Irvine, CA

Sep 2015 - Mar 2019

Bachelor of Arts in Quantitative Economics (GPA: 3.5/4.0)

TECHNICAL SKILLS

LLM & Agent Systems: Agent harnesses, multi-agent orchestration, tool use/function calling, long-horizon workflows, context engineering, memory/RAG, evals/graders, guardrails, human-in-the-loop, observability, MCP

Programming Languages: TypeScript, Python, Rust, JavaScript, SQL, HTML/CSS

Product & APIs: Next.js, React, Node.js, FastAPI, Fastify, Hono, REST/WebSocket APIs, Vercel AI SDK, Playwright

Cloud & Data: AWS (CDK, Lambda, SQS, DynamoDB, CloudWatch), GCP, Docker, PostgreSQL, SQLite, Redis

PROFESSIONAL EXPERIENCE

SiriusMindshare LLC, San Jose, CA

Oct 2025 - Present

Machine Learning Engineer | *Next.js, TypeScript, AWS, Vercel AI SDK*

- Architected and shipped a production multi-agent platform with **5 role-specific agents** and **10 registered tools** across marketing, design, bookkeeping, support, and sales; encoded per-agent personas, context, turn limits, and tool permissions.
- Enabled prompt-prefix reuse and bounded long-session context growth by separating **approximately 4K static tokens** from dynamic state and triggering **4-stage compression at 80% capacity**.
- Built a **5-step agent execution graph** with dependent/parallel orchestration, human approval gates, cross-agent notifications, streaming events, and SQLite-backed workflow and audit state.
- Co-authored a paper on a generative AI/LVLM framework for retail attention-quality assessment, accepted for publication at **IEEE CCET 2026**; extended the platform's evaluation approach to multimodal workflows (forthcoming proceedings).

1 Thing Against Racism Organization, Monroe, NJ (Remote)

Jan 2024 - Apr 2025

Full-Stack Software Developer Intern | *React, React Native, Flask, Firebase, GCP*

- Delivered React/React Native clients and Flask/GCP APIs with Firestore synchronization, OAuth, encryption, notifications, and AI-assisted moderation; reduced frequent-query latency by **40%** through caching.

PROJECTS

AgentFlow | *TypeScript, Bun, React, Hono, WebSocket*

2026

- Engineered an agent harness and execution runtime with **5 node types** and **3 edge modes** for sequential, parallel, conditional, and state-machine workflows; exposed role-scoped tools, REST/WebSocket APIs, and a React UI.
- Implemented code-enforced guardrails with automated and human quality gates, filesystem and shell policies, and token, turn, and edit budgets inherited across **3 workflow/role/step scopes** at runtime.
- Built evaluation, recovery, and debugging infrastructure with cross-examination, checkpoint recovery, blind review, runtime validation, isolated worktrees, and evidence bundles; maintained **1,670+ unit/e2e tests** across **6 packages**.

Agent Bridge | *Rust, Python, TypeScript, Docker, AWS*

2026

- Designed a capability-gated sandbox orchestration layer exposing **5 operations** across **8 provider backends** and **9 capability flags**, blocking unsupported compute, network, storage, and secret operations before execution.
- Delivered a Rust CLI/runtime, Python/TypeScript SDKs, and **5 MCP tools** with consistent schemas across Rust, Python, and TypeScript, plus structured errors, retries, audit trails, OTel telemetry, and cost accounting.
- Enforced preflight security with image allowlists, signature/vulnerability admission, fail-closed isolation, no-network execution, least-privilege IAM, and **3 hot/warm/cold artifact tiers**.

Kaggle Team | *TypeScript, Bun, Zod, Kaggle API*

2026

- Built a multi-agent experimentation loop coordinating **3 specialized agents** across a **4-phase** research-to-QA workflow; emitted **4 structured artifacts** plus scorecards and cost summaries per round.
- Scheduled **2 isolated engineers** with O(N+E) dependency resolution and **16 Zod-validated tools**; used CV-to-leaderboard gaps and community deltas for QA while parallelizing reads and serializing writes.

AI Diary Companion | *Next.js, TypeScript, Fastify, Prisma, SQLite*

2026

- Delivered a full-stack AI journaling product with **6 persona companions**, SSE-streamed chat, structured diary generation, safety/drift anchors, and **50+ guided prompts** for reflective writing and check-ins.
- Implemented typed memory retrieval that ranks the **top 10 items** by importance/recency, extracts conversation memories, triggers decay/merge above **30 stored items**, and caps chat context at **20 messages**; validated with **170+ unit/e2e tests**.

Additional open source: *Agent Hook (6-source event-driven agent control plane) | jsfetch (Python Playwright CLI/MCP for JS-rendered pages)*