

Vu Nguyen

Software Engineer · AI agents, agent infrastructure, end-to-end product delivery

713-598-9156 · vuluan@gmail.com · github.com/vu1n

EXPERIENCE

Phantom — Senior Data Scientist

2026 · Remote

- Sat with analysts and PMs who waited on someone else every time they needed a query turned into a shareable chart, and shipped the internal API that removed the handoff -- one POST returns a chart, Slack unfurl, or social image. Designed agent-first from the spec out: `/llms.txt` serves the full API reference as agent-readable markdown, so an agent self-authors against the same surface a human reads, and provenance (session, thread, requester) is first-class. A one-person tool took over a workflow that normally needs a BI team.
- Built and shipped an AI data exploration agent used in production for warehouse analytics: retrieval over warehouse metadata finds the right data, code-mode Python runs the analysis in a sandboxed container, and a templated report lands at the end. Delivered as both an interactive CLI and an **MCP server exposing 15 tools** for native Claude Code and Claude Desktop use.
- Tuned retrieval weights with Optuna + MLflow against a held-out eval set (recall@k, precision@k, MRR). A local DuckDB cache removed a concrete failure mode -- the agent paying for Snowflake queries against the wrong table before realizing it. Defense-in-depth at every agent surface: sandboxed execution, SQL AST validation, a tool-registry allowlist.
- Traced onboarding funnel drop-off in a new auth flow to missing event coverage, then shipped the instrumentation into the mobile wallet repo with mobile engineering rather than filing a ticket. Published the funnel, bot-detection, and screen-level audits into data engineering, analytics, and growth so three orgs worked from one shared reading of the funnel.

Independent Builder

2025 · Remote

- Built and operated trading and market-making bots across Hyperliquid, Pacifica, and Solana DeFi, focused on execution quality, risk limits, and monitoring.

Consensys (MetaMask) — Staff Data Engineer

2022–2024 · Remote

- Led schema-as-code for MetaMask analytics events: canonical schemas in Git-tracked config, Node.js tooling that validates changes, flags breaking modifications, and generates compile-time types so malformed events fail at the source. Enforced in CI via Docker + GitHub Actions.
- Shipped an auto-documented event browser so PMs and analysts could answer their own questions instead of queuing behind data engineering; drove cross-team adoption until data contracts were a normal part of the engineering workflow.

Squarespace — Senior Software Engineer

2020–2022 · Remote

- Built a unified analytics platform across Squarespace products during a cloud migration -- a canonical wide customer table with pre-aggregated attributes let most reporting hit one table with minimal joins, and reconciling against legacy reports is what earned stakeholder trust in it.

Earlier: **Facebook (Meta)** — Data Engineer, IC6 (2018–2019) · **Indeed** — Senior Data Engineer (2018) · **StockX** — Data Engineer Team Lead (2017–2018).

PROJECTS

Pillbox — local-first sandbox for coding agents: credentials it can use but never sees

Rust, Tokio, libkrun, Hudsucker, Cloudflare Durable Objects · github.com/vu1n/pillbox

- An MITM credential vault swaps stub keys for real ones at the host boundary, so a prompt-injected or leaked key inside the sandbox is worthless -- the agent sends a credential it never sees. Per-sandbox CA, destination-bound swaps, per-host OAuth minting.
- Sessions are an append-only typed event stream, not a terminal to sit in front of: subscribe streams JSON to a UI or orchestrator and --from SEQ late-joins from any point, so a frontend builds on the same surface you drive by hand. 512 commits, public.

Bacchus + Lum — running and supervising fleets of coding agents

Rust, tree-sitter, SQLite FTS5, jj · TypeScript, Tauri v2, React, tldraw, libkrun/E2B · github.com/vu1n/bacchus

- Bacchus coordinates subagents by lease rather than central dispatch -- TTL heartbeats, j j worktrees isolating each on its own branch, AST-aware retrieval, skill archetypes matched to task type, and Claude Code and Codex as interchangeable runners behind one coordination surface.
- Lum supervises them spatially: each agent on a canvas in its own microVM, pausing and resuming cleanly, with a structured status card as the primary surface and the terminal demoted to a drill-in.

brief — architectural decisions as governance, read-only to the coding loop

Python, Claude Code skills, GitHub Actions · github.com/vu1n/brief

- The party being checked can't edit the check: a ratified decision is read-only to coding work, and changing one means filing an amendment a human ratifies. Validated adversarially on a real Rust codebase -- handed agents tasks contradicting firm decisions, naive "docs are editable truth" framing let them reverse the decision **4 out of 4 times**; separation-of-powers framing produced **0 reversals across 8 runs** on two model tiers, every agent stopping to file an amendment instead.

Hyperstitch (Rubick) — non-custodial copy-trading platform, live on Hyperliquid mainnet

Go (DBOS), TypeScript (Effect), Python, DuckDB/DuckLake, Postgres, Pulumi/AWS · rubick.fi

- Solo build across five repos, zero to live on mainnet in five weeks, now serving invited users. Trade execution is a durable 14-state saga with crash-recovery replay and exchange reconciliation; the session signer is policy-scoped default-deny, so the engine that trades cannot move user funds by construction.
- **Also public** -- [Canopy](#) (codebase-indexing MCP server) and [Fando](#) (Claude Code skills framework, Codex as external reviewer); upstream PRs to [NousResearch/hermes-agent](#), [DefiLLama](#), and [deathbyknowledge/gsv](#).

EDUCATION

University of Texas at Austin — B.S., Computer Science

Georgia Institute of Technology — Online M.S., Computer Science (20 credit hours completed)

TECHNICAL

Languages	Python, TypeScript, Rust, Go, SQL
AI & Agents	MCP servers, subagent orchestration, agent skills; Claude Code and Codex daily; agent runtimes and sandboxing (Docker, libkrun microVMs, bubblewrap/Seatbelt); eval harnesses and failure-mode analysis (held-out gates, recall@k, MRR); structured outputs, retrieval, tool routing, agent memory; LoRA/QLoRA fine-tuning (Unsloth, PEFT), DSPy, Optuna, MLflow
Infrastructure	AWS, GCP, Docker, Kubernetes, Temporal, DBOS, Pulumi, GitHub Actions, PostgreSQL, Cloudflare Workers/Durable Objects
Data	Snowflake, BigQuery, ClickHouse, DuckDB, dbt, Parquet/Arrow, Polars