

Vu Nguyen

Product Engineer — full-stack, AI-native developer tools, data platforms

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

EXPERIENCE

Phantom — Senior Data Scientist

2026

- Built a production AI data-exploration agent — RL-tuned retrieval to find the right data in a large warehouse, sandboxed code-mode Python to analyze it, an interactive HTML report at the end — with an eval + observability layer (citation-scoring tuned via Optuna + MLflow against a held-out retrieval set; per-turn TurnTrace JSONL). Shipped as a CLI and a 15-tool MCP server.
- Root-caused a DEX-aggregator fee-mechanism volume cliff — traced it to fee vaults and program addresses the models didn't know about, modeled an untracked bot pattern into a canonical dataset; recovered significant production volume. When event coverage was the gap, shipped the instrumentation directly in the mobile wallet repo rather than filing a ticket.

Independent Builder — Software & Data Engineer

2025–Present

- 12+ active projects, several open-source from inception — a coding-agent runtime (Pillbox), codebase intelligence (Canopy), multi-agent orchestration (Bacchus), a real-time agent-first data service (Cinco), and a stealth trading stack (order-execution, market-intel, product). End-to-end across Rust, TypeScript/React, Go, and Python.

Consensys (MetaMask) — Staff Data Engineer

2022–2024

- Led a schema-as-code initiative for analytics events across a large user base — canonical schemas in Git, Node.js validation and compile-time client types enforced in CI, plus a browsing UI so PMs and analysts could navigate events without reading YAML. Drove data contracts to first-class.
- Contributed to the broader data-platform and governance model — new product initiatives shipped with reliable, privacy-respecting tracking from day one.

Squarespace — Senior Software Engineer

2020–2022

- Designed a cloud analytics schema for a warehouse migration; built ETL/ELT pipelines and reconciled outputs against legacy reporting.

Wanderjaunt — Senior Software Engineer

2019–2020

- Built the end-to-end data and analytics stack across Redshift, BigQuery, and Snowflake.

Meta (Facebook) — Data Engineer, IC6

2018–2019

- App-responsiveness performance analysis; matrix-profile anomaly/motif detection on mobile time-series.

Indeed — Senior Data Engineer

2018

- Led GDPR compliance across company-wide data pipelines.

StockX — Data Engineer, Team Lead

2017–2018

- Led the data team through the marketplace's early scale-up.

SELECTED PROJECTS

Pillbox — a real runtime for coding agents: microVM sandboxes, a credential vault the agent uses but never sees, sessions as an OpenTelemetry event stream, multi-runner, §0 multiplayer trust layer. (Rust)

github.com/vu1n/pillbox

Canopy — token-efficient codebase intelligence for agents: tree-sitter symbols, token-aware budgets, recursive-LM path-finding; beats grep. (Rust) github.com/vu1n/canopy

Bacchus — multi-agent orchestration: lease-based coordination, jj worktrees per agent, single-agent eval discipline. (Rust) github.com/vu1n/bacchus

Cinco — real-time market-data service with an agent-first API (filtered, programmatic, confidence-scored); Temporal + ClickHouse, venue-isolated.

Gremlin — turns real user sessions into Playwright/Maestro tests and regression signal; early MVP. (TypeScript) github.com/vu1n/gremlin

The Frontier of Evals — a React portal over the awesome-evals corpus (catalog, landscape map, guided reading); shipped upstream. vu1n.github.io/awesome-evals

EDUCATION

University of Texas at Austin — B.S., Computer Science · Georgia Institute of Technology — Online M.S., Computer Science (in progress)

TECHNICAL

LANGUAGES TypeScript / React / Next.js, Python, Rust, Go, Swift, SQL, Solidity

INFRA Kubernetes, Temporal, Docker, libkrun microVMs, OpenTelemetry, Cloudflare Workers / Durable Objects, Pulumi, GitHub Actions

AI / AGENTS LLM agents, multi-agent orchestration, MCP servers, codebase indexing, eval (Optuna + MLflow), retrieval / RAG, DSPy RLM, per-turn tracing

DATA ClickHouse, Snowflake, DuckDB, Postgres, dbt, schema-as-code / data contracts, real-time event pipelines