

Justin Sublette

AI Systems Engineer | Agent Evaluation and Research Environments

j.sublette@vmosaic.com · 812-374-4565

Naples, Florida, USA · Willing to relocate

Open source: [Mathematical Research](#) · [Protocol Runner](#)

Profile

AI systems builder developing autonomous research environments and experiments that assess useful agent behavior. Connects durable scientific context, adaptive agent coordination and falsification with controlled evaluations. Investigates when successful execution or favorable scores fail to establish the intended result, then uses findings to improve systems and adoption decisions.

Experience

Founder and AI Systems Engineer | Vector Mosaic Inc. | 2025-present

Founded a software company spanning autonomous research, agent infrastructure and clinical software. Own problem definition, architecture, integration, experimental direction and acceptance of results, collaborating extensively with Codex on implementation, investigation and testing. Began practical software development in June 2025 through self-directed, AI-assisted learning. Selected systems below are projects within this work.

Mathematical Research / RH Mission | Building an autonomous scientific research system | 2026-present

- Created a system for pursuing a checkable proof or disproof of the Riemann Hypothesis with autonomous AI. Designed a durable Mission with an AI executive that selects and revises strategy, delegates investigation and criticism, and synthesizes results across replaceable conversations. My role is system design, integration and evaluation, not routine human mathematical strategy or proof adjudication.
- Separated cumulative scientific understanding from immediate strategy, retaining qualified connections, useful partial results and failed approaches. Built revisioned branches, evidence and candidate records, exact-source retrieval, checkpoints and atomic publication of an evidence result with its authored continuation.
- Built the research process around exact claim scope, dependencies and adversarial falsification. One campaign rejected an overbroad obstruction claim because its class had not been declared before the search, while retaining a narrower local lemma. Kept candidate preservation, role-separated review and mathematical admission distinct.
- Published the MIT-licensed implementation and [four reproducible cases](#) separating checked mathematics, historical research and authored store walkthroughs. One reference-assisted live successor retrieved retained evidence and improved a bound from 10 to less than 9 under the corrected assumptions. No general success rate or blind-discovery claim follows.

RH remains open. Public results cover the standalone release with fresh state and the selected examples.

CTX Engine | Testing whether repository context improves agent answers | 2025-2026

- Built context infrastructure combining keyword search, semantic search and dependency analysis, then evaluation harnesses comparing agent answers against ordinary repository search. Separated file discovery, answer quality, verification and time/token measurements.
- In 30 internal code-navigation tasks from the project's own repository, retrieved 69% of reference-relevant files versus 34% for keyword search, with 85% selection precision. A separate 46-task same-repository answer comparison recorded nine wins, one loss and 36 ties or mixed results under coarse scoring.
- Scored required facts, unsupported claims and acknowledgment of missing evidence. Protected reference answers from trial agents and froze source, model, prompt and scoring conditions for a sealed holdout. Excluded

two attempts invalidated by execution/output-contract failures, repaired the conditions and completed a clean rerun.

- Investigated why an experimental context format found more relevant evidence without improving answers. Added diagnostics for supplied evidence, agent-recorded file/evidence paths and the sequence of useful evidence discovery; kept missing time/token measurements explicit.
- Retained the standard configuration after an experimental replacement failed to demonstrate better answers on 24 held-out tasks from the same repository. Used the experiment to make an adoption decision and identify remaining measurement limitations.

Pre-Note Delta | Evaluating AI-generated clinical prose and AI-operated tests | 2026

- Created an Epic-integrated application showing chart changes since a clinician's prior encounter. Kept twenty deterministic, evidence-bearing summary sections separate from optional AI drafting, with source inspection and clinician review; the application did not autonomously write to the chart.
- Built browser and desktop test harnesses combining scripted navigation with an optional AI pilot for diagnosis and recovery. Replaced shared replay data with isolated patient-evidence collection after discovering unreliable provenance; a strict recollection ledger recorded 1,733 successful collections and eleven unresolved failures.
- Designed a 96-case matrix crossing scenario, baseline and drafting mode. Distinguished successful streams, adequate source inputs and faithful prose: 53 cases passed the quality audit, 42 were input-limited and one had corrupted output.
- Traced source-use failures to a digest-and-merge pipeline and replaced it with bounded direct source packets. A twelve-case diagnostic rerun captured all outputs: four used additional detail, six remained effectively summary-only and two lacked sufficient material. Kept those findings separate from later passing browser-behavior checks and the broader quality audit.
- In a sixteen-patient Epic sandbox cohort, all 3,599 evidence references loaded. Demonstrated AWS vendor infrastructure and prepared hospital pilot materials; no hospital rollout or commercial success was established. Retired the system in July 2026.

Protocol Runner | Preserving work through investigation and correction | 2026-present

- Built persistent orchestration for long agent workflows. Used 25 investigation assignments over 595 commits to produce 40 engineering-history accounts; separate AI review found three needing correction. Preserved the completed investigations and other accounts, corrected the three and obtained a fresh full-draft review before internal acceptance.
- Open-sourced [Protocol Runner](#) under MIT, with a real API/store recovery demo using simulated workers and separate live Codex CLI completion and cancellation checks. [Evidence](#).

Technical foundation

Python, TypeScript, Node.js, SQLite, PostgreSQL; evaluation harnesses, protected references, held-out comparisons, failure analysis; semantic/keyword retrieval, dependency graphs, content-addressed evidence; Playwright, browser debugging protocols, Git, MCP, Linux and remote execution. Experience evaluates existing models and their surrounding systems; no hands-on model construction or training yet.

Earlier experience and education

Manufacturing and pipeline construction, 2012-2018: shift-team leadership and field coordination. Independent market research and trading, 2018-2025.

CCNA: certified April 7, 2025; valid through April 7, 2028. **Indiana University, SPEA:** policy analysis studies, 2013-Spring 2015; no degree completed. **Columbus North High School:** Academic Honors diploma.