

TRAVIS SMITH

AI SOFTWARE ENGINEER · AGENTIC SYSTEMS ARCHITECT

Nashville, TN · me@thisistravissmith.com

(615) 339-8828 · thisistravissmith.com

SUMMARY

AI engineer building agentic systems for high-stakes, documentation-heavy workflows — where accuracy, auditability, and trust decide whether the system ships. Eight years of clinical-documentation domain experience. Founder of Exigere Solutions (proprietary agentic systems) and Etzira Consulting (client-owned AI for small businesses). Works AI-augmented: directs LLM tooling to move fast on implementation while owning the architecture, evaluation, and safety decisions.

SKILLS

AGENTIC SYSTEMS	multi-agent orchestration, LangGraph, MCP server development (FastMCP), tool design, task routing, workflow state, human-in-the-loop confirmation gates
RAG & MEMORY	vector embeddings, semantic retrieval, pgvector / FAISS, persistent context stores, retrieval grounding
LLM EVALUATION	rubric-based scoring, frontier-model guardrail red-teaming, adversarial red/blue/reconcile audits, CI-gated eval harnesses, hallucination & failure-mode analysis, LLMops (tracing, drift monitoring, model pinning)
MODELS & RUNTIMES	Anthropic Claude (Opus / Sonnet / Haiku), OpenAI GPT & Codex, open-weight models (Llama, Hermes) via Ollama / vLLM, Whisper STT, local vision-language models
ENGINEERING & INFRA	Python, TypeScript / Next.js, SQLite / PostgreSQL, Docker / Compose, k3s, Proxmox, WireGuard / OPNsense, Ansible / NixOS, Stripe, Vapi, Git
SECURITY & OSINT	Sherlock, Maigret, Holehe, theHarvester, WhatsMyName, CCPA / GDPR takedown automation, cosign / Sigstore signed-artifact supply chain, Vault-issued mTLS
DOMAIN	clinical documentation, medical transcription, healthcare operations, privacy-aware (PHI-adjacent) workflow design

EXPERIENCE

- Exigere Solutions — Founder & AI Systems Architect

2023–present

 - Founder and lead architect of a proprietary offline-first agentic platform for high-stakes, compliance-sensitive operations — on-prem multi-agent orchestration, local RAG, and human-in-the-loop governance.
 - Developed a cross-model adversarial evaluation method — red / blue / reconcile audits — to stress-test agentic systems for enforcement gaps, fail-opens, and tests-green paths that don't enforce the guarantee they claim.
- Etzira Consulting — Founder & Principal Engineer

2026–present

 - Founded and run an agentic-services consultancy delivering client-owned AI systems to small businesses — 17 packaged offers across four families (front desk, growth, backbone, field-sales intelligence), live at etziraconsulting.com.
 - Built the multi-tenant platform underneath (TypeScript · Next.js · Claude API · Vapi · Stripe): every client tenant is a JSON config file, so new tenants and whole verticals ship with zero code changes — five products composed from one agent engine, with a TCPA/PEWC compliance core, persistent opt-out store, and a QA gate agents must pass before go-live.
 - Built its outbound engine: homepage scans disqualify AI-equipped prospects, leads are scored and tiered, and every draft enforces CAN-SPAM in code — a live run over 120 real Nashville SMBs qualified 88.
- AI Evaluation & Training — Independent Contractor

2025–present

 - Assess LLM outputs for instruction-following, factual accuracy, reasoning quality, and safety using structured, rubric-based approaches — focused on AI-coding, agentic-reasoning, and documentation tasks, where domain accuracy is the hard part.
 - Compare competing model responses to surface hallucinations, weak completions, and edge-case failures, and write feedback that measurably improves model behavior.

- On contract with DataAnnotation, Google Machine Learning, Anthropic, and OpenAI.
- Contract work spans systems-security / red-team testing of guardrails on newly released frontier models, domain-specific task evaluation, and evaluation of full departmental agentic-systems models.

eTran Business Solutions — Healthcare Documentation & Operations Specialist

2016–2024

- Produced high-accuracy clinical documentation and medical transcription under fast-turnaround and strict-confidentiality requirements.
- Firsthand knowledge of where documentation workflows break down — provider patterns, QA bottlenecks, the gap between how notes are written and how they need to read — the domain grounding behind the AI systems designed since.

SELECTED PROJECTS

Full write-ups: thisistravissmith.com

Field Intel — AI sales-call scoring & pre-visit intelligence

python · fastapi · whisper · ollama · react

- Locally-owned alternative to cloud “virtual ridealong” SaaS: transcribes and scores field sales calls against the client’s own script with a local LLM — every evidence quote is machine-verified against the transcript, so the model can’t fake receipts; consent detection is deterministic.
- Pre-visit homeowner dossiers + an offline vector map from 504 real county parcels and permits, zero cloud calls; public A/B demo (94% vs 31% on-script) live at thisistravissmith.com.

EndpointDrift — clinical-trial outcome-switching auditor

python · langgraph · pgvector · local VLM

- Catches drug trials “moving the goalposts”: compares endpoints pre-registered on ClinicalTrials.gov against the published paper, flagging omitted, added, demoted, or changed outcomes; grounding enforced at the type level — no finding without verbatim evidence quotes.
- Froze a reproducible 200-trial corpus (sha256 manifest) from live government APIs; converted 38 COMPare-audited trials into 358 provenance-traceable gold labels; 101 tests, \$0/month self-hosted stack.

OSINT Fusion — local-first personal-exposure scanner & takedown engine

python · local LLM · OSINT toolchain

- Fuses manual-input OSINT tooling (Sherlock, Maigret, Holehe, theHarvester) into a single local, zero-cloud scanner: typed identifiers in, verified exposure findings out, auto-drafted CCPA takedown requests; live run surfaced 102 exposures.
- Hardened through two independent audit lanes — 10 cross-model review passes plus an adversarial verification swarm that caught defects the first lane missed, including a false-clean failure (rate-limited checks misreported as determinate negatives); 1,400+ tests green.

Second Brain — local-first personal memory & intelligence system

python · sqlite · local LLM / vision

- Privacy enforced by architecture, not query filters: fail-closed sensitivity tiering, physically separated stores, an encrypt-only vault the always-on process cannot decrypt, and an egress kill-switch.
- Hardened through repeated adversarial red-team/blue-team review; ~118 tests lock the privacy invariants — a cross-tier leak is a hard test failure.

Private AI Lab — on-prem AI inference lab (reference architecture)

proxmox · vLLM · wireguard / opnsense

- Segmented, zero-egress on-prem architecture for local LLM and agentic workloads — VLAN trust zones, default-deny egress, no internet-facing services (identity-gated WireGuard access only).
- Threat-model-driven security: signed-artifact supply chain, sandboxed model serving, encryption at rest, 3-2-1 backup, Vault-issued mTLS; fully reproducible via IaC (Ansible/NixOS, Compose/k3s).

ADDITIONAL SYSTEMS 42-agent factory with automated QA gating · offline multi-agent orchestrator · sports-market consensus engine · voice & booking agents · production client sites

EDUCATION

University of Washington — B.A. Psychology, focus Industrial Psychology

Seattle, WA

Organizational systems, human factors, and decision-making — applied to human-centered AI design.

CONTINUING STUDY ongoing applied research — transformer internals and attention variants, state-space architectures (Mamba-class), quantization and inference optimization for on-prem deployment, adversarial red/blue evaluation for agentic systems, multi-agent failure taxonomies, and current AI-safety literature.