

Joe Laskowski

AI Solutions Engineering | LLM Evaluation | Agent Infrastructure

South Bend, Indiana · joe.laskowski79@gmail.com · linkedin.com/in/thebiglaskowski · github.com/thebiglaskowski · thebiglaskowski.com

SUMMARY

Twenty-five years in IT, the most recent stretch spent as the human in the loop for production LLM assistants — authoring the evaluation rubrics and failure taxonomy, building the Teams chatbots and Copilot Studio agents that deliver the models to technicians, and instrumenting whether any of it actually moved. Deep enterprise background underneath it: Microsoft 365, Active Directory, Intune, Entra ID, and roughly 82,000 lines of production automation across six languages. Nearly all of that built with AI in the loop — directing these systems to produce tested, production-grade work is the skill, and it is the same skill I am paid to evaluate.

PROFESSIONAL EXPERIENCE

AI Quality Analyst

Aunalytics — South Bend, IN · March 2026 – Present

- Human in the loop for two LLM-powered ticket enrichment assistants. Scored 2,500+ AI outputs against a five-dimension rubric I authored, using a coded failure taxonomy that converts individual bad answers into diagnosable model weaknesses. Core finding, stated in week one and confirmed by five months of coded data: the failure is in synthesis, not retrieval — completeness failures outnumber accuracy failures.
- Designed and built the Microsoft Teams chatbots technicians use to reach both systems: Azure Functions and Bot Service over an in-house LLM, with live third-party data wired in through Power Automate, circuit breakers and response caching on the API path, and a tenant allow-list plus three-layer auth. Refactored two near-identical 5,300-line bots into a single 2,500-line shared core at 95% statement coverage.
- Raised technician documentation quality 29% (2.37 to 3.05 on a five-point rubric) across 18 weekly coaching cycles covering 89 technicians and 7,371 graded entries, more than doubling the share of tickets carrying a usable resolution entry. Those notes are the model's training corpus, so the gain compounds.
- Measured an automated remediation program to 96% script success and 61% of tickets closing with zero technician touch, reclaiming 3.45 TB of disk across 493 tickets, with script failures driven to zero over the measurement window.
- Identified and corrected a data-export defect that had understated a reporting population by an order of magnitude. Suspended trend comparisons, rebuilt the baseline, and published the corrected figures rather than let a flattering artifact stand.
- Mined 134,000+ tickets across 15 months into 115 recurring types, then authored the standard resolutions and the docx-to-SQL parser that load them into the Azure SQL knowledge base the assistants read at enrichment time.
- Designed a six-agent Microsoft Copilot Studio family under a hard 8,000-character per-agent instruction budget, wrote the tooling to manage it, and authored the internal methodology guide for agent instruction design. Later ran a formal governance review of my own agents against vendor best practice, producing 11 prioritized configuration findings and a remediation program.
- Authored the function's operating standards end to end: scoring rubrics, failure codes, documentation style guides, a twelve-KPI measurement framework across three tiers, and the Azure architecture proposal for productizing the whole capability.

Internal Support Engineer

Aunalytics — South Bend, IN · August 2023 – March 2026

- Designed and built the company's user lifecycle automation end to end — nine PowerShell modules spanning Active Directory, Microsoft 365, and ticket-driven provisioning, with certificate-based authentication from Key Vault, schema-validated configuration, separate production and test environments, and a full Pester test suite.
- Built zero-touch Windows Autopilot deployment: certificate-only auth from Key Vault, a resumable checkpoint state machine with exponential backoff, and automated cleanup of stale Intune, Autopilot, and Entra records.
- Wrote the quarterly privileged-access review tooling — exporter, risk flagger, and quarter-over-quarter diff — plus a Microsoft 365 tenant audit module carrying CIS and vendor baselines with HTML reporting.

- Linux subject-matter expert for the Centralized Services team; worked part-time with the Security team as a SOC analyst on internal and client security needs.
- Delivered Tier 1 and Tier 2 support to internal staff and contributed to organization-wide technology rollouts.

Service Desk Location Lead — Kalamazoo

Aunalytics · January 2022 – August 2023

- Led the Kalamazoo team through Tier 1 and Tier 2 delivery — incident management, escalation, and SLA maintenance — while owning staff performance, mentoring, and training.
- Co-lead engineer for internal departments: user access administration, Microsoft 365 licensing, and general engineering support.

Service Desk Lead

Aunalytics · October 2019 – January 2022

- Led a team of technicians and engineers supporting the service desk's largest accounts, including environments north of 1,500 users.
- Owned customer relationships through the client IT point of contact — service delivery, escalation communication, and ticket follow-through.

Managed Services Specialist

MicroIntegration · March 2017 – September 2019

- Resolved 70 to 100 tickets per week across remote first-level support, investigating complex hardware and software faults and escalating what required it.

Field Technician Manager

Realm Technologies · July 2007 – March 2017

- Ran day-to-day operations for break/fix technicians nationwide, recruiting and certifying 200+ technicians delivering warranty repair for Dell, Sony, IBM, HP, and Lexmark.
- Managed service call dispatch, parts logistics, and vendor relations; maintained the company intranet and public website.

Earlier: Hardware Field Technician, On Sight / Qualxserv / Dell (2005–2007). Logistics Coordinator and Deployment Technician, Compaq / TCML at Whirlpool Corporation (2001–2002) — supervised a three-person team through a 6,000-machine multi-site rollout including world headquarters.

SELECTED PERSONAL PROJECTS

Self-directed work, built outside the day job.

Nexus Signals — market analysis platform. Python 3.12 / FastAPI / PostgreSQL / Redis with a React 19 and TypeScript front end. Domain-driven design across six bounded contexts, CQRS, seven market-data providers behind circuit-breaker fallback chains, 23 technical indicators feeding a scoring model, LLM-driven sentiment analysis, paper trading with live P&L, WebSocket streaming, 43 REST endpoints. 3,764 passing tests; mypy --strict with zero Any types.

cowork-rag + cowork-graph — private retrieval infrastructure (MCP). Two complementary retrieval systems over one corpus, both exposed as Model Context Protocol servers and shared across Claude, Codex, Grok Build, and Hermes — the host supplies the answer model, these supply the evidence. cowork-rag runs fully local: Nomic embed-text-v1.5 (768-dim) and a MiniLM cross-encoder reranker in-process via FastEmbed ONNX over ChromaDB, with no GPU server or API key. Hybrid retrieval fuses a SQLite FTS5 keyword lane with the vector lane by reciprocal-rank fusion so exact identifiers stay findable, then reranks the pool. Layered on top: map-reduce document summaries, a GraphRAG entity and relationship layer clustered into summarized communities, date-aware filtering, domain scoping, local OCR and image captioning (Tesseract + Florence-2), cross-session memory, and an optional PII detector that masks identifiers before they reach a host's context. Sync is mtime-then-content-hash with chunk-level embedding reuse; generation sits behind a provider-neutral interface defaulting to a nightly local llama.cpp Qwen3-8B. cowork-graph is the structured half — a SQLite knowledge graph over the same Markdown answering what embeddings are bad at: what's active, what's blocked, what links to what, what got decided. 2,651 files, 98,242 chunks, 38,345 graph entities across 150 communities, 15 MCP tools.

Claude Sentient — agent orchestration layer (public). An autonomous development loop for coding agents — INIT, UNDERSTAND, PLAN, EXECUTE, VERIFY, COMMIT, EVALUATE — with hard quality gates on lint, test, build, and git between stages. Nineteen slash commands, fifteen hooks, nine agent roles, nine language profiles with Python environment detection across conda/venv/poetry/pdm, twelve JSON schemas enforcing cross-module integrity, 924 tests, and a self-improvement mechanism that proposes its own rule updates after corrections. Used it for months, then removed

it once native platform features turned the abstraction layer into drag. github.com/thebiglaskowski/claude-sentient

Documentary production pipeline — three chained tools. Script to finished 1080p video, automated end to end: TTS narration with WhisperX sentence-level beat alignment, a browser-based manager for the ~150 image generations an episode consumes, and Remotion-based assembly with motion, transitions, and per-era color grading. 90% test coverage on stage one; five episodes shipped.

NexusLedger — multi-entity finance system. Plaid ingestion, rules-based categorization, reconciliation, and reporting across three legal entities. Checksummed migrations, period locks, pending-to-posted carryover, and an integrity-check command that catches drift before it reaches a report. 662-test Python suite.

PromptVault. Deterministic prompt compiler and semantic library — card/template engine, multi-pass critic, 92 system cards, 27 templates, and a 30-case regression harness diffing output against stored baselines. Node 22, pnpm monorepo, Fastify, Next.js, PostgreSQL 16 with pgvector HNSW indexes, Redis 7, Drizzle.

SEO content pipeline. Thirteen-package TypeScript monorepo: keyword discovery, gap analysis, revenue-per-keyword scoring, competitive briefs, AI copy and image generation, static build, and GEO scoring — grading pages for citation-readiness across generative search engines. OpenAI + Anthropic, DALL·E 3 + Stability AI, Serper.dev.

DropTok. TikTok commerce platform on live marketplace APIs: multi-shop management, AI listing generation, content pipeline, AI-scored opportunity board, creator CRM, a fulfillment queue built as an explicit state machine, and token-spend tracking. TypeScript end to end; React 18 / Vite 6 / Zustand over Express 4 / Drizzle / better-sqlite3.

BudgetLedger. Budgeting application on live Plaid sync with CSRF, rate limiting, and optional TOTP MFA. Pay-period planning assigns bills to the paycheck that covers them and flags negative paychecks before they land. Flask 3.1, SQLAlchemy 2.0, Chart.js.

xrp-ticker. Terminal portfolio tracker reading balances directly off the XRP ledger over WebSocket, aggregated across self-custody wallets, with live exchange pricing, four cyclable sparkline renderers, and auto-reconnect. Python 3.11+, Textual, uv. Public on GitHub.

SKYNET backup. Encrypted, deduplicated, snapshot-preserving backup across three redundancy layers, driven by systemd timers and Windows Task Scheduler, with green/yellow/red Slack alerts on every run and restores round-tripped and SHA-256 verified byte-identical. restic, Bash.

TECHNICAL SKILLS

AI and LLM: LLM output evaluation and rubric design; prompt and agent-instruction engineering; Model Context Protocol (MCP) server development; Microsoft Copilot Studio; Anthropic and OpenAI APIs; retrieval and knowledge-gap analysis; agent orchestration and multi-agent design; LLM observability

Languages: PowerShell, Python, TypeScript, JavaScript, SQL, Bash

Cloud and platform: Azure (Functions, Bot Service, SQL, Key Vault, Application Insights), Entra ID, Microsoft 365, Intune, Active Directory, Power Automate

Systems: Linux (RHEL, Ubuntu, CentOS) as a daily driver since 2007 and exclusively for sixteen years; Proxmox and virtualization; Bash and shell tooling; self-hosted infrastructure, backup, and monitoring

Engineering: Pester, pytest, Vitest, REST and GraphQL integration, circuit breakers and retry design, pandas, FastAPI, React, Remotion, Git, SQLite, PostgreSQL, Redis

Practice: Specification before implementation; test-driven development; dry-run defaults on destructive operations; schema-drift guards; documentation as a first-class deliverable

COMMUNITY, EDUCATION, AND CERTIFICATION

Co-founder, Michiana InfoSec (2019) — the regional information-security community serving South Bend and the surrounding area. Helped establish the group and ran its early years, including its first corporate-sponsored sessions; handed off to new leadership and the group continues without me.

Claude Certified Architect — Foundations (in progress). Technical training, Productivity Point International. Mentor, Aunalytics internal mentoring program.