

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

I build the systems that decide whether AI output is any good, and then I build the things that deliver it. 25 years in IT. Right now I'm the human in the loop for a pair of production LLM assistants, which means I wrote the rubrics, coded the failure taxonomy, shipped the chatbots the technicians actually use, and then measured whether any of it moved. Deep enterprise background underneath that: Microsoft 365, Active Directory, Intune, Entra ID, and roughly 82,000 lines of production automation across 6 languages. Nearly all of it built with AI in the loop, which I'd rather say out loud than bury. Getting these systems to produce tested, production-grade work is the skill, and it happens to be the one I get paid to evaluate.

PROFESSIONAL EXPERIENCE

AI Quality Analyst

Aunalytics — South Bend, IN · March 2026 – Present

- The human in the loop for 2 LLM-powered ticket enrichment assistants. I wrote the 5-dimension rubric we score them against and the coded failure taxonomy that turns a bad answer into something the data science team can act on. The finding that's held up for 5 months: the model finds the right documentation and then stops short. Synthesis is the problem, not retrieval.
- Designed and built the Microsoft Teams chatbots technicians use to reach both systems. Azure Functions and Bot Service over an in-house LLM, live third-party data wired in through Power Automate, circuit breakers and response caching on the API path, tenant allow-list and 3-layer auth. Refactored 2 near-identical 5,300-line bots into a single 2,500-line shared core at 95% statement coverage.
- Technician documentation quality up 29% (2.37 to 3.05 on a 5-point rubric) across 18 weekly coaching cycles covering 89 technicians and 7,371 graded entries, with the share of tickets carrying a usable resolution entry more than doubled. Those notes are what the model trains on, so it compounds.
- Built the parser and reporting that measured a new automated remediation feature through its first week in production: 493 tickets, 96% script success, 61% closing with zero technician touch, 3.45 TB of disk reclaimed. The automation was another team's build; the instrumentation and analysis were mine.
- Caught a data-export defect that had been quietly misstating a reporting population by roughly 10x. Suspended the trend comparisons, rebuilt the baseline, published the corrected numbers. The old numbers looked better.
- Kicked off our standard-resolutions program with a discovery script I wrote that mined 134,000+ tickets across 15 months into 115 recurring types, then worked with a small team to turn the highest-volume ones into published resolutions. Also built the docx-to-SQL parser that loads them into the Azure SQL knowledge base the assistants read at enrichment time.
- Designed a 6-agent Microsoft Copilot Studio family under a hard 8,000-character per-agent instruction budget, built the tooling to manage it, and wrote the internal methodology guide for agent instruction design. Later ran a governance review of my own agents against vendor best practice and produced 11 prioritized findings with a remediation plan.
- Wrote the operating standards for the function: scoring rubrics, failure codes, documentation style guides, a 12-KPI measurement framework, and the Azure architecture proposal for turning the whole thing into a product.

Internal Support Engineer

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

- Built the company's user onboarding and offboarding automation end to end. 9 PowerShell modules covering Active Directory, Microsoft 365, and provisioning driven straight off the ticket, with certificate-based auth from Key Vault, schema-validated config, separate prod and test environments, and a full Pester 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, quarter-over-quarter diff) plus a Microsoft 365 tenant audit module carrying CIS and vendor baselines.
- Linux subject-matter expert for the Centralized Services team. Worked part-time with the security team as a SOC analyst on internal and client work.
- Tier 1 and Tier 2 support for internal staff, and a hand in most organization-wide 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, SLA maintenance. Owned staff performance, mentoring, and training.
- Co-lead engineer for internal departments. User access administration, Microsoft 365 licensing, and whatever engineering support came up.

Service Desk Lead

Aunalytics · October 2019 – January 2022

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

Managed Services Technician

MicroIntegration · March 2017 – September 2019

- Closed 70 to 100 tickets a week on remote first-level support, working complex hardware and software faults and escalating what needed it.

Field Technician Manager

Realm Technologies · July 2007 – March 2017

- Ran day-to-day operations for break/fix technicians nationwide. Recruited and certified more than 200 techs doing warranty repair for Dell, Sony, IBM, HP, and Lexmark.
- Managed dispatch, parts logistics, and vendor relations. Kept the company intranet and public site running.

Before that: Hardware Field Technician, On Sight / Qualxserv / Dell (2005–2007). Logistics Coordinator and Deployment Technician, Compaq / TCML at Whirlpool Corporation (2001–2002), where I supervised a 3-person team through a 6,000-machine multi-site rollout including world headquarters.

SELECTED PERSONAL PROJECTS

Built on my own time. Several ship under AutoScript Studio LLC, an Indiana company I started to hold the work that outgrew being a hobby. Full write-ups, with architecture and stack detail for each: thebiglaskowski.com/resume.

cowork-rag + cowork-graph. Two search systems over the same corpus of my own documents, one semantic and one structured, both wired into Claude, Codex, Grok Build, and Hermes over MCP. Hybrid retrieval fuses an FTS5 keyword lane with the vector lane by reciprocal-rank fusion, then reranks with a cross-encoder. GraphRAG, date-aware filtering, local OCR, and PII masking on top. Runs entirely on my own hardware.

FastEmbed ONNX · ChromaDB · SQLite FTS5 · local llama.cpp Qwen3-8B · 2,651 files, 98,242 chunks, 38,345 entities · 15 MCP tools

PixelDrops production pipeline. A research question goes in and a finished, published documentary comes out. 5 tools pass work down a chain, a runner carries a per-video manifest between them, and an agent team drives it, stopping at 7 human approval gates. TTS narration with WhisperX force-alignment produces the beat timestamps everything downstream keys off; Remotion assembly handles motion, transitions, kinetic type, per-era color grading, and data motifs bound to specific beats.

pixeldrops-research-mcp → scriptcast-cli → shotlist → framecast-cli → yt-publish-cli · 4 channels · 5 episodes shipped

Nexus Signals. A market analysis platform for crypto and stocks. Domain-driven design across 6 bounded contexts with CQRS, 7 market-data providers behind circuit-breaker fallback chains, 23 indicators feeding a normalized signal score, LLM sentiment analysis, and full paper trading over WebSocket. I got halfway into v1, saw how I should have built it, and started over.

Python 3.12 / FastAPI / PostgreSQL 16 / Redis 7 · React 19 / TypeScript · 3,764 passing tests · mypy --strict, 0 Any types

Claude Sentient. An autonomous development loop for coding agents, built before the tooling shipped one. 7 stages with hard quality gates between them, 9 language profiles that detect their own environment, and a mechanism that watched me override it and proposed edits to its own rules. I ran it for months, then pulled it out of everything once native features made it drag.

19 slash commands · 15 hooks · 9 agent roles · 897 tests, all passing · public at github.com/thebiglaskowski/claude-sentient

NexusLedger. One ledger covering personal accounts and 2 companies, where the entities can't bleed into each other and a closed month can't quietly move. Plaid ingestion, rules-based categorization, reconciliation, reporting. Checksummed migrations, period locks, synthetic transfer pairs, and a doctor command that catches drift before it reaches a report.

Python CLI · Plaid · SQLite · 3 legal entities · 662-test suite

Also: PromptVault, a prompt compiler with a pgvector semantic library and a 30-case regression suite. A 13-package SEO content pipeline that scores keywords by expected revenue and grades finished pages for AI-search citation-readiness. DropTok, a TikTok Shop operations platform on the live marketplace APIs. An encrypted 3-layer backup system with Slack

alerting and SHA-256-verified restore round-trips. xrp-ticker, a terminal portfolio tracker reading the XRP ledger over WebSocket.

TECHNICAL SKILLS

AI and LLM: LLM output evaluation and rubric design; prompt and agent-instruction engineering; Model Context Protocol (MCP) server development; RAG and GraphRAG pipelines; embedding, reranking and hybrid retrieval; Microsoft Copilot Studio; Anthropic and OpenAI APIs; agent orchestration; 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) daily since 2007 and exclusively for 16 years; Proxmox and virtualization; systemd; shell tooling; self-hosted infrastructure, backup and monitoring

Data and storage: PostgreSQL, SQLite, Redis, ChromaDB, pgvector, pandas, FTS5

Engineering: Pester, pytest, Vitest, REST and GraphQL, circuit breakers and retry design, FastAPI, React, Remotion, Git

How I work: Spec before implementation; test-driven where it counts; dry-run defaults on anything destructive; schema-drift guards; documentation as a deliverable

COMMUNITY, EDUCATION, AND CERTIFICATION

Co-founder, Michiana InfoSec (2019) — the regional information security community for South Bend and the surrounding area. I ran it through the early years and handed the reins to a new admin a few years back. It kept going without me.

Working toward the Claude Certified Architect — Foundations credential. Technical training, Productivity Point International. Mentor in the Aunalytics internal mentoring program.