

Matt Davis

Senior AI Product Engineer · *Agentic AI, LLM platforms and MCP*

Eugene, OR · me@mattdav.is · linkedin.com/in/dattmavis

SUMMARY

AI product engineer who builds production LLM systems end to end and gets them adopted. I co-lead AI product and engineering for ABACUS, Avolution's enterprise architecture platform: agentic tooling on the Claude Agent SDK, Python and FastAPI AI services on AWS Bedrock with LangGraph and Langfuse, and MCP servers that put enterprise data in front of AI assistants. Before AI product I was a sales engineer for Fortune 500 integrations, and before that, at 22, I grew a living-room phone repair startup into a 12-person, two-store business doing about \$1.3M a year, and ran it for nearly a decade.

EXPERIENCE

Senior AI Product Engineer, Avolution

Jul 2025 – Present · Eugene, OR

- Co-lead AI product and engineering for ABACUS with a fellow senior AI lead, hands-on across Python and TypeScript services, LLM integration, cloud infrastructure, security and observability, shipping on versioned enterprise release cycles.
- Built ABACUS Code, a multi-user agentic AI scripting platform on the Claude Agent SDK (TypeScript, Hono, React) with Entra ID SSO, an MCP endpoint over personal API keys and a self-improving skills library. Cut a 26-hour client integration to about 12 minutes at roughly \$0.20 per run; used daily across consulting, pre-sales and support, averaging nearly three new projects every working day.
- Built a remote MCP server in Python (FastMCP, OAuth 2.1 with dynamic client registration, revocable tokens) that lets AI assistants build EA models, diagrams and dashboards, with review-before-commit, an audit log and a time-saved impact report.
- Lead engineering on AI-assisted data import for the next-generation platform: FastAPI and Pydantic endpoints for LLM column mapping, type detection and pgvector duplicate detection, idempotent async jobs with leases and retries, and a 22-workbook golden evaluation set scoring 97–98% on attribute mapping.
- Built LLM observability for the AI service on Langfuse: trace design across the LangGraph agent runner and tool nodes, per-customer project isolation, OpenTelemetry separation and LLM-as-judge evaluators for groundedness and mapping quality.
- Built a MEGA-to-ABACUS importer that recreates MEGA diagrams 1:1 as live architecture models (components, connections and viewpoints), verified by an automated render-and-pixel-diff loop instead of AI vision: 35 of 35 diagrams imported at a 2.6% mean pixel difference. Demoed it live to resolve the key objection in a multi-year, six-figure new-logo deal.
- Built a migration tool (Python CLI plus web app) that converts existing ABACUS models to the next-generation platform; imported a 304K-element, 2.1M-value, 2,831-diagram model exactly, with a 329 MB upload converted in about 150 seconds after cutting it to 21 MB in the browser.
- Built an AI image importer (VB.NET, AWS Bedrock) that turns architecture diagrams into model elements and connections: 43 of 44 elements and 37 of 37 connections on a 44-element infrastructure test.
- Cut AI Assistant cost per query by over 40% in A/B testing through prompt caching, with no change in output quality and roughly double the concurrent capacity; also stopped a runaway-context cost spike in ABACUS Code with auto-compaction and read caps.

- Co-own the roadmap and engineering direction for the ABACUS AI Assistant on AWS Bedrock, with multi-tenant access control, prompt-injection guardrails and full LLM observability; led a security review of the AI layer and set the remediation order.
- Lead company-wide AI adoption and governance: Claude rollout and seat management, AI policy, data-handling standards, staff enablement and spend analysis. Named an Avo-Legend for Innovation at Speed (2026), representing North America.

Sales Engineer, Avolution

Apr 2022 – Jul 2025 · Austin, TX

- Technical lead across pre- and post-sales for the ABACUS enterprise architecture platform: demos, training, proofs of concept and custom integrations for Fortune 500 clients.
- Built API connectors for NIST NVD, GCP, ServiceNow and Azure, plus integrations and automations in VB, C# .NET and Python.
- Presented to everyone from operational teams to the C-suite, and guided clients on EA frameworks including TOGAF and BPMN.

Owner, Screen Geeks, LLC

Apr 2012 – Oct 2021 · Lufkin, TX

- Founded at 22 and grew from a living room to 12 employees, two stores and about \$1.3M in total annual revenue, including \$300K a year in online parts and device sales; sold both stores in 2021.
- Owned the P&L, hiring and operations, and managed both locations directly for seven years.
- Won and ran B2B contracts with local government, schools and other repair shops.
- Designed and built the company’s online platforms in JavaScript, HTML and CSS.

SKILLS

AI & LLM engineering	Agentic AI · Claude Agent SDK · Anthropic API · LangGraph · LiteLLM · Model Context Protocol (MCP) · FastMCP · Prompt engineering · Prompt caching · Tool use and structured output · LLM evaluation (golden datasets, LLM-as-judge) · Vector search (pgvector) · Prompt-injection guardrails
LLM observability	Langfuse · OpenTelemetry
Models & platforms	AWS Bedrock · Claude (Sonnet, Haiku) · Amazon Nova · Titan embeddings
Languages	Python · TypeScript · JavaScript · SQL · VB.NET · C#
Backend & APIs	FastAPI · Pydantic · Node.js · Hono · REST APIs · OData · OAuth 2.1 / OIDC (Entra ID SSO) · JWT · Streaming (NDJSON) · Async job queues
Cloud & data	AWS (Bedrock, EC2, S3, RDS, IAM, Secrets Manager, SSM) · Docker · Docker Compose · Caddy · Cloudflare · PostgreSQL · Redis · SQLite
Frontend	React · Preact · Vite · Tailwind CSS
Quality & delivery	pytest · mypy · ruff · Vitest · Playwright · pre-commit · GitHub
Enterprise architecture	ABACUS · TOGAF · BPMN · ServiceNow · Azure · GCP