

J. Michael Brannigan

ML & DATA PLATFORM ENGINEERING CLOUD / BACKEND ARCHITECTURE PRODUCTION ML SYSTEMS

Alameda, CA mike@msxpert.com (310) 428-5500 vectorpanda.com msxpert.com/cv

Remote, Hybrid, or SF Bay Area

Principal engineer who owns data and ML systems end to end — ingestion, pipelines, model serving, and the observability and cost controls that keep them running. I turn ambiguous requirements, fragmented data, and operationally complex systems into documented, measurable software that teams can extend and trust. Currently shipping aviation-safety intelligence at Navi AI; independently operate a benchmarked vector-search service in production.

SELECTED WORK

Navi AI · Forward Deployed Engineer

Jun 2026 – Present

Aviation AI · flight telemetry + cockpit audio · domain LLM · customer-embedded delivery

- Own the insights layer over flight telemetry, cockpit audio, and flight-school syllabus data — dropout risk, student performance, and exceedance-trend detection, surfaced as signal instructors can act on.
- Own the customer requirements loop directly with **United Airlines, Embry-Riddle, Sling Pilot Academy, Utah State, and UND**, converting field requirements into shipped features on a predictable cadence.
- Joined a **\$6.7M**-seed team (United Airlines Ventures, New Vista Capital, Raptor Group) building the first generative-AI platform in commercial operation for pilot training.

Vector Panda · Founder / Sole Engineer

LIVE PRODUCT

May 2025 – Present

Managed vector database · Rust core, Python SDK · public API at vectorpanda.com

- Designed, built, and operate a managed vector-search service end to end: Rust query and index engine, veep Python SDK, tiered storage, public API, billing, and published benchmarks. Built solo using agent-assisted workflows with human ownership of architecture, data contracts, and production quality.
- Sustains **1,620 QPS** at **0.965 recall@100** on VectorDBBench (Cohere 1M), with **3.1 s** cold time-to-first-query.
- Built an auto-optimizer that benchmarks **11** index families — HNSW, Vamana, IVF-Flat, IVF-PQ, PQ, OPQ, ScaNN, SQ, LSH, PCA, brute-force — against a sample of each customer's own data and selects off the accuracy / recall / p95-latency Pareto frontier, making index choice a measured decision rather than a hand-tuned one.
- Drove serving cost for a 1M-vector collection to **\$17.14/month** with a hot / warm / paused tiering model — in-memory **<10 ms**, SSD **<100 ms**, archive — priced at \$5.99 / \$1.49 / \$0.09 per GB-month.
- Multi-format ingest (Parquet, Arrow IPC, HDF5, NumPy, Safetensors, JSONL, CSV, BVECS/FVECS/IVECS), ONNX Runtime inference with a bundled all-MiniLM-L6-v2 for local encoding, and MongoDB-style predicate filtering at query time.

Loti · Principal / Founding Engineer

Aug 2024 – May 2026

ML inference · image/video ingestion · AWS · observability · data contracts

- Built ML inference and image/video ingestion systems processing up to **700,000 images/day**, indexing social-platform content against Aurora/Postgres, DynamoDB, and Milvus.

- Built custom MTCNN-derived models for cosine-similarity matching and AI/deepfake content detection across Instagram, Facebook, Pinterest, TikTok, and YouTube.
- Unified fragmented schemas and "data puddles" into documented interop contracts, improving dashboarding, reporting, and engineering workflow coherence.
- Operated multi-service pipelines with OpenTelemetry, Terraform CDKTF, ECS/Docker, API Gateway, GitHub Actions, and monorepo patterns.

Pareto Health · Data Team Lead / Senior ML & Analytics Engineer

Mar 2021 – Aug 2024

Healthcare data · ML · lake architecture · compliance-sensitive systems

- Led a **team of 4** building a from-scratch ingestion platform and **50+ TB** S3/DynamoDB data lake for healthcare underwriting and supervised learning workflows.
- Automated manual Excel-based underwriting into an API-driven CRM adjunct, improving throughput and decision quality **20-fold**.
- Integrated a commercial claims library spanning **10 years** of history across **300 million lives** into the underwriting feature set.
- Maintained strict PHI/PII segregation, tokenization, and HIPAA-sensitive data handling across dozens of ingest sources.
- Shipped ML predictions into legacy product systems via AppSync, GraphQL, TypeScript Lambda, EMR/Hadoop, PySpark, and CatBoost.

Little Caesars / Microsoft · Azure ML & DevOps Engineer

Mar 2020 – Jun 2020

- Built a PaaS ML pipeline forecasting hourly "Hot-N-Ready" demand across **3,500** global locations, integrating weather, events, and POS sales data via Kafka and Azure Functions.

LaVid Technologies / Anonifi · Technical Co-Founder / CTO

Jul 2016 – 2022

Fintech · privacy · alternative credit · backend/data architecture · hybrid cloud

- Co-founded LaVid, a B2B microfinance lending platform built to undercut predatory lending, then led the pivot to Anonifi — a direct-to-consumer service letting the chronically under-banked build verifiable alternative credit profiles from PII they control.
- Sole in-house technologist: built the web, mobile (Ionic), server infrastructure, and data systems from scratch — Docker Compose, Node.js, Postgres, Elasticsearch/Kibana, and FIDO2/WebAuthn.
- Built custom ML tooling to algorithmically surface credit opportunities from movement and activity data that users self-collected and submitted anonymously.
- Designed hybrid infrastructure with on-prem XenServer coordination and burst capability to AWS, Azure, and GCP. Work in this space produced **US Patent 11,023,942**.

Expensify · Software Engineer

Feb 2018 – Jun 2018

- Engineer-at-large across mobile apps, C++/SQLite core, DevOps tooling, and PHP/React backoffice systems; contributed to Project 10X scale-out ahead of a national media campaign.

Sony PlayStation · Senior UI Architect

Jun 2015 – Jul 2016

- Modernized PlayStation Vue UI systems from extJS/jQuery to AngularJS/HTML5/CSS3; built a high-performance Node.js IP geolocation service and prototyped a React re-architecture.

CORE STACK

DATA & ML	SQL (advanced — data modeling, query tuning, warehouse design), Python , PySpark, EMR/Hadoop, Airflow, Kafka, SageMaker, Azure ML, Databricks, CatBoost, ONNX Runtime, vector search (HNSW, IVF-PQ, ScaNN), embeddings, feature pipelines, MTCNN-derived inference
BACKEND	Rust , TypeScript, Node.js, FastAPI, GraphQL, PostgreSQL/Aurora, DynamoDB, SQLite, S3 data lakes, API design, data contracts, monorepos
CLOUD & OPS	AWS, GCP, Azure, Terraform / CDKTF , Docker/ECS/Lambda, API Gateway, GitHub Actions, OpenTelemetry, XCP-ng, secure bastions, bare-metal colo
PRACTICE	Agent-assisted engineering (Claude Code, Codex, MCP/OpenRouter, Beads) with human ownership of architecture, security boundaries, data contracts, and production quality

CREDENTIALS & EARLIER WORK

• US Patent 11,023,942 — secure data cryptography • AWS Certified Cloud Practitioner • Prevailed <i>pro se</i> in domain-name defense; secured a Reverse Domain Name Hijacking finding • Commercial Pilot, Instrument Rated; CFI / CFII / MEI • FAA Inspection Authorization; Airframe & Powerplant Mechanic	Earlier engineering work includes Pala Interactive, GlobalCare Analytics, Palmetto Solar, Trailer Park, Kodak, Warner Bros., Toyota/Lexus, and Blizzard Entertainment — across web, mobile, streaming, gaming, healthcare, fintech, and data systems. Longer work history, legacy project detail, references, and role-specific technical deep dives available on request.
--	--