

Derrick Karake

Oklahoma City, OK · 405-887-9145 · karakederrick@gmail.com · github.com/derrickkarake

SUMMARY

Full-stack data scientist building production, **AI-integrated systems** for energy operations. Primary architect of a real-time frac data acquisition platform at Devon Energy — agent-based ingestion, automated fallback, AI-powered PDF/CSV backfill, and a React monitoring app — delivering **\$2.5M+ in projected annual savings**. Also shipped an ML-driven Azure casing forecasting platform supporting a **\$350M procurement program**.

TECHNICAL SKILLS

AI / ML	LLM agents, OpenAI SDK, Pydantic AI, Hugging Face Transformers, PyTorch, scikit-learn, time-series forecasting
Data	Snowflake (CTEs, parameterized SQL, pooling), OSIssoft PI, PySpark, Pandas, schema mapping / extraction
Azure	App Service, Blob Storage (SAS URLs), Container Registry, Azure AD / MSAL, Logic Apps, Microsoft Graph API
Languages	Python, TypeScript, JavaScript, SQL, Bash
Frontend	React 18, Vite, Material-UI, Recharts, Plotly.js, React Query, React Hook Form, Zod
Backend	Node.js, Express, Flask, REST APIs, node-cron, multi-agent pipeline design
DevOps	Docker (multi-stage), Nginx, Azure Container Registry, CI/CD, Playwright E2E, Vitest

EXPERIENCE

Full-Stack Data Scientist — Devon Energy

Oklahoma City, OK · January 2023 – Present

Real-Time Frac Data Acquisition Platform (flagship system) — sole architect of an end-to-end ingestion, validation, and AI-assisted backfill platform replacing a third-party vendor pipeline, supporting live frac fleet operations across Devon's field footprint.

- **Integrated-agent ingestion architecture:** Python/PySpark agents pull ~80 high-frequency telemetry files per frac fleet daily, run gap/spike validation, and stream clean data into the **OSIssoft PI** historian in near-real-time.
- **Automated fallback subsystem:** primary/secondary ingestion paths with health checks, automatic failover on stream drop, retry with exponential backoff, and self-healing recovery on node or device loss — keeps ingestion flowing when upstream sources degrade.
- **AI-powered backfill engine:** LLM-based agents (OpenAI SDK) parse **unstructured vendor PDF reports and heterogeneous CSV exports**, extract telemetry fields, and map varied service-provider schemas into PI-compatible format for historical gap recovery.
- **React monitoring app:** operator-facing frontend (React + Node.js) for live ingestion status, data-quality inspection, manual backfill triggers, and fallback/failover visibility — used daily by completions engineers.
- Partnered with IT to replace third-party cellular links with direct VPN (field data reliability **<80%** → **>97%**); eliminated the third-party data software dependency — **\$40K/month** recurring savings and projected **\$2.5M/year** in acquisition-fee savings from in-house transmission.

Casing Procurement & Demand Forecast Platform (Azure, production) — React/TypeScript + Node.js platform supporting a **\$350M casing procurement program**; migrated a legacy Python/Streamlit prototype into a containerized Azure deployment.

- Built a production forecasting model combining **decision trees and logistic regression**, fully deployed on **Azure** with end-to-end **MLOps** (automated retraining, drift monitoring, model observability) — re-scores casing needs by SKU (OD, weight, grade, connection) across 30–180 day and full-year horizons as drilling and completion operations shift day-to-day, producing exact per-SKU counts and delivery windows for the \$350M program.
- **Multi-stakeholder input pipeline:** drilling engineers submit per-well casing plans, supply chain teams feed in-flight inventory and buffer thresholds, and casing providers contribute lead-time and shipment commitments — all rolled into each forecast run so the model reflects real-time field, inventory, and vendor reality.
- **Year-ahead delivery projection layer** schedules rig-level casing deliveries across the full forward program, auto-rebalancing whenever drilling order, completion pace, or provider lead times shift — preventing stockouts while minimizing carrying cost.
- Azure stack — App Service, Container Registry, Azure AD SSO (3-tier RBAC), Blob Storage (SAS URLs), Logic Apps + Graph API cron pipeline; 15+ Snowflake tables across 4 enterprise DBs with parameterized SQL/CTEs, in-memory caching, and full audit trail.

Data Scientist — RMS

Oklahoma City, OK · April 2022 – November 2023

- Fine-tuned, trained, and deployed a **BERT** question-answering model extracting structured fields from **5,000+ TIFF images/day** (F1 = 0.90), replacing a manual review process.
- Implemented time-series forecasting for daily inbound document volume; built Sisense and Power BI dashboards surfacing operational performance.

EDUCATION

M.S. Artificial Intelligence · In Progress

October 2025 – Present

M.S. Data Science — Oklahoma Christian University, Edmond, OK

2021 – 2022 · GPA 3.78

B.S. Computer Science — Oklahoma Christian University, Edmond, OK

2016 – 2021