Michael Martin

Murfreesboro, TN • (615-691-1505) • Michael.wmartin91@gmail.com
https://www.linkedin.com/in/6424235/

SUMMARY

Data Engineer and AI Engineer with 3+ years deploying production LLM systems, building scalable data pipelines, and delivering analytics solutions that drive measurable business impact. Proven track record architecting enterprise AI tools, modernizing legacy infrastructure, and translating complex technical concepts for cross-functional stakeholders. Combines physics background with expertise in Python, cloud platforms, and machine learning to solve real-world problems from concept to production.

TECHNICAL SKILLS

Languages: Python, PySpark, SQL, Bash
AI & ML: LLM Engineering (Llama, GPT, Mixtral), HuggingFace, Prompt Engineering, RAG, DeepEval, scikit-learn
Data Engineering: Apache Airflow, Databricks, Snowflake, ETL Pipelines, Docker
Cloud & Infrastructure: AWS, Azure, On-Premises Deployment
Analytics & Visualization: Tableau, Power BI, Excel
Practices: Agile, DevOps, CI/CD, Cross-functional Collaboration, Technical Leadership

PROFESSIONAL EXPERIENCE

Consultant – CGI | Knoxville, TN | 2022 – 2025

LLM & AI Projects

Architected and deployed enterprise survey categorization system using Llama 3.3 70B (on-prem) with custom HuggingFace integration, processing 240K+ employee and client survey responses across 3 classification levels with 93% accuracy

- Reduced manual categorization timeline from 3 months of nights/weekends to 3 weeks, eliminating hundreds of SME and Marketing team hours through automated prompt-engineered classification
- Served as technical liaison between Marketing, Business, and Engineering teams; presented workflow and validation results to SVP leadership in go/no-go decision meeting
- Built solution using on-premises infrastructure to maintain data privacy for confidential internal feedback

Developed natural language to SQL proof-of-concept system enabling non-technical staff to query databases using GPT-3.5, demonstrating feasibility of conversational data access

- Led 4-person team over 1.5 months; personally built Python-SQL connector layer for ElephantSQL integration
- Designed custom orchestration pipeline: natural language → GPT translation with embedded schema context → SQL execution → results delivery via CLI

Engineered real-time Responsible AI evaluation pipeline using DeepEval framework to assess open-source LLM outputs (primarily Mixtral) across 8 compliance criteria: relevance, contextual recall, faithfulness, hallucination detection, PII (regex-based), toxicity, and bias

- Implemented automated blocking mechanism preventing response generation when any safety metric is flagged during prompt or completion phase
- Delivered production-ready backend system for company's open-source model hub supporting AI governance initiatives

Data Engineering & Modernization

Led SAS-to-PySpark migration for medical claims pricing platform on Databricks, independently converting 2,500+ lines of legacy code over 4 months to process 116K+ weekly claims

- Architected modern data pipeline: CSV ingestion → Databricks transformation → enterprise database export for downstream claims analyst consumption
- Part of 4-person team modernizing pricing infrastructure across multiple client implementations

Rebuilt legacy claim pricing logic from SAS to modularized Python, creating client-specific implementations across 8 healthcare organizations

- Developed boutique pricing rules engine accommodating unique client business logic while maintaining scalable core architecture
- Delivered performance improvements that reduced processing burden for claims adjustment teams

Migrated 7 legacy SQL workflows to Apache Airflow DAGs (60+ workflows team-wide) within 1 month, contributing to $120K annual licensing cost savings

- Designed self-managed Airflow infrastructure with improved pipeline observability, monitoring, and scheduling capabilities
- Translated complex legacy scheduling logic to modern orchestration framework supporting claims processing operations

Analytics & Visualization

Developed claims processing dashboards in Tableau and Power BI for operations managers and claims analysts, visualizing key performance metrics

PROJECTS

AllergyFind – Food Allergen Safety Platform | allergyfind.com • *Python, FastAPI, Grok API, Brave Search, JavaScript, SQLite*

- Engineered full-stack web application enabling users to safely identify allergen-free menu items across 156+ restaurants, supporting FDA's Big 9 allergens
- Built intelligent caching system using SQLite and designed multi-stage pipeline: user query → Brave Search API → LLM-powered menu parsing (Grok) → filtered results in 15-20s
- Architected fail-safe parsing strategy prioritizing false negatives over false positives to ensure user safety; developed 1,500+ lines of custom PDF/web parsing logic before pivoting to LLM solution

Medicare Fraud Detector – ML Anomaly Detection Pipeline | *Python, Snowflake, scikit-learn, OpenAI GPT-4, Isolation Forest*

- Designed end-to-end fraud detection pipeline analyzing 850K+ Medicare claims using Isolation Forest (contamination=0.01) with 3 engineered features (payment amount, length of stay, cost per day)
- Integrated GPT-4 to auto-generate plain-language fraud summaries for non-technical QA teams, reducing analysis time and improving stakeholder communication
- Flagged 3% of claims as anomalous using rule-based detection (0-day high payments, duplicate claim IDs) combined with ML scoring

Clanker – Privacy-First Local LLM Assistant | *Python, llama.cpp, Llama 3.1 (8B), Qwen 2.5 (14B), GGUF*

- Developed offline-capable AI assistant using quantized open-source models (Q4_K_M) running entirely on local hardware via llama.cpp for complete data privacy
- Implemented conversation memory system and GUI-based model switching between Llama 3.1 (8B) and Qwen 2.5 (14B) for task-specific optimization

VOLUNTEER EXPERIENCE

Mentor – Congressional App Challenge | L&N STEM Academy | 2023 – 2025

- Mentored high school students on app development, design, and technical presentation skills
- One mentee secured internship at CGI; another won first place in district competition

EDUCATION

Master Certificate in Data Science | Middle Tennessee State University | 2021
Relevant Topics: Predictive Modeling, Model Deployment, Data Understanding

Bachelor of Science in Physics | Middle Tennessee State University | 2019
Relevant Courses: Modern Physics, Quantum Mechanics, Optics, Computational Physics