Matt Oremland, PhD

Senior Machine Learning Engineer

mattoremland@gmail.com • (518) 560-1473 • Saratoga Springs, NY 12866
https://www.linkedin.com/in/matthew-oremland-31a312101

PROFESSIONAL SUMMARY

Machine learning engineer with 9+ years building production ML systems and data-intensive applications across pharmaceutical manufacturing and SaaS environments. Proven track record deploying end-to-end ML pipelines, agentic AI solutions, and backend services using Python, cloud platforms, and modern ML frameworks. Strong foundation in software engineering with experience building full-stack applications, REST APIs, and scalable data architectures. PhD in Mathematics with expertise in optimization algorithms and computational modeling.

PROFESSIONAL EXPERIENCE

Tidal Wave Analytics LLC
January 2025 – Present

Consulting Data Scientist / Founder — *Saratoga Springs, NY*

- Built production random forest model for Paradigm (building supply SaaS) that groups quotes into projects using text and numerical features; deployed to Databricks for nightly batch processing with monitoring and alerting
- Engineered RAG system for Denari tax tech startup using HuggingFace and PyTorch; designed document ingestion pipeline downloading IRS and state tax forms, built vector store for semantic search, and implemented query generation for tax filing recommendations
- Developed full-stack daily puzzle game (Escalators) from scratch: Python backend for puzzle generation algorithms, React/Node.js/Vite/Tailwind frontend, Google OAuth integration, and Supabase for data persistence; running in production for 2+ months with competitive leaderboards
- Designed and implemented end-to-end ML workflows including data preprocessing, feature engineering, model training, validation, and deployment across multiple client engagements

Redica Systems
April 2025 – Present

Director, Data Strategy & Analytics — *Remote*

- Built production LLM classifier using OpenAI API to analyze unstructured FDA 483 inspection reports; engineered prompt templates for multi-class severity classification and weighted topic extraction; designed validation framework using confusion matrices and collaborated with data engineering on deployment
- Developed Python-based risk scoring models analyzing FDA enforcement patterns across pharmaceutical facilities; designed feature engineering pipeline processing regulatory text, facility attributes, and historical violation data in Snowflake
- Architected reference data harmonization system consolidating multi-source client data (Excel, databases, APIs); built custom matching algorithms and inference logic for missing values using fuzzy string matching and domain knowledge
- Created reusable Python visualization library for generating client-specific analytics outputs; designed modular architecture supporting multiple chart types, branding configurations, and export formats

Gilead Sciences
January 2023 – March 2025

Director, Data & Analytics (Medical Affairs) — *Foster City, CA*

- Built centralized data lake consolidating healthcare professional (HCP) and healthcare organization (HCO) data from 5+ disparate sources; designed ETL pipelines, schema normalization, and data quality validation logic
- Developed automated data processing workflows for HCP/HCO profile enrichment including affiliation mapping, relationship graph construction, and engagement tracking; reduced manual data operations by 70%
- Implemented data governance framework with formal data lineage tracking, quality metrics, and access controls for healthcare reference data serving Medical Affairs organization

Takeda Pharmaceuticals May 2021 – December 2022

Associate Director, Digital & Data Science *Lexington, MA*

- Built predictive ML models for manufacturing process optimization using ensemble methods, PLS, and PCA; deployed models driving real-time quality decisions in production environment
- Designed and implemented digital dashboards for real-time manufacturing process monitoring; integrated sensor data streams, statistical process control algorithms, and alerting logic
- Engineered automated data pipelines for Annual Product Quality Review (APQR) generation; built data extraction, transformation, statistical calculation, and report generation workflows reducing review time from weeks to days

Regeneron Pharmaceuticals May 2016 – May 2021

Senior Process Data Scientist *Rensselaer, NY*

- Built automated feed quantity optimization system for bioreactor manufacturing; developed optimization algorithms considering process constraints, material availability, and quality targets; system changed business decisions for production planning
- Engineered predictive modeling platform for product quality prediction using ensemble methods and partial least squares regression; designed feature selection pipeline, cross-validation framework, and model performance monitoring
- Developed automated statistical limit calculation system processing historical batch data; implemented outlier detection, distribution fitting, and confidence interval estimation algorithms; reduced manual calculations from weeks to hours
- Built end-to-end data pipelines for manufacturing data ingestion, cleaning, feature engineering, and storage; designed data quality checks, error handling, and logging for production reliability

EDUCATION

Doctor of Philosophy (PhD) in Mathematics 2014

Virginia Polytechnic Institute & State University

Dissertation: Optimization and control of agent-based models in biological systems. Published 7 peer-reviewed papers applying genetic algorithms and multi-objective optimization to computational biology.

SKILLS & TECHNOLOGIES

Programming & Software Engineering: Python • JavaScript/TypeScript • Node.js • React • SQL • REST APIs • Git

Machine Learning & AI: Scikit-Learn • TensorFlow • PyTorch • HuggingFace • OpenAI API • Random Forest • Ensemble Models • PLS • PCA • RAG Systems

Data Platforms & Cloud: Snowflake • Databricks • AWS • Supabase • Sigma • ETL/ELT Pipelines

ML Operations: Model Deployment • Pipeline Automation • Data Quality Validation • Production Monitoring • A/B Testing

Specialized Domains: Agent-Based Modeling • Optimization Algorithms • Statistical Process Control • FDA-Regulated Manufacturing • Data Governance

PUBLICATIONS

- "Optimization and control of agent-based models in biology: a perspective" — Bulletin of Mathematical Biology, 2017
- "Agent-based models and optimal control in biology: a discrete approach" — Mathematical Concepts and Methods in Modern Biology (Elsevier), 2013 (Textbook chapter)