Matt Oremland, PhD

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 and data scientist with 10+ years of experience designing, developing, and validating AI/ML models in the pharmaceutical industry. Hands-on experience building LLM-based production systems using OpenAI APIs, predictive modeling with scikit-learn, and deep learning frameworks including TensorFlow and PyTorch. Strong background in applying ML to unstructured and structured data problems in pharma, collaborating closely with engineering teams to bring models from prototype to production, and ensuring rigorous validation and performance monitoring.

PROFESSIONAL EXPERIENCE

Redica Systems
2025 – Present

Director, Data Strategy & Analytics — *Remote*

- Designed and developed an LLM-based classification model using the OpenAI API that reads unstructured text from FDA Form 483 documents and assigns severity levels (low, medium, high, critical) and weighted topic categories (e.g., Quality Unit, Documentation) - deployed into production across the entire Redica data platform
- Owned model design, prompt engineering, and validation methodology end-to-end, including confusion matrix analysis and accuracy benchmarking in collaboration with the data engineering team to ensure production implementation matched prototype performance
- Built end-to-end risk-scoring models using Python and scikit-learn, designing pipelines from raw data ingestion through feature engineering, model development, and operationalization
- Designed a reference data architecture to harmonize data from multiple pharmaceutical clients and external sources, resolving complex data quality and integration challenges

Tidal Wave Analytics LLC
2023 – Present

Consulting Data Scientist — *Saratoga Springs, NY*

- Delivered machine learning engagements across multiple industries using a range of frameworks including scikit-learn, TensorFlow, and PyTorch, selecting approaches based on problem type and scale
- Built and deployed a production classification model for a SaaS platform using Databricks - the model runs nightly with ongoing monitoring and iteration, demonstrating experience with the full ML production lifecycle
- Designed and built a RAG pipeline in Python using HuggingFace and PyTorch for a tax-tech startup - ingested a large corpus of IRS and state tax forms from public sources and built the full document retrieval and generation system to advise businesses on tax filing requirements
- Partnered with client engineering teams to translate ML solutions into production systems, collaborating on integration, validation, and performance requirements

Gilead Sciences
2023 – 2024

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

- Led development of analytics and modeling capabilities within a regulated pharmaceutical environment, ensuring data pipelines and model outputs met rigorous quality and compliance standards
- Built centralized data infrastructure supporting multiple downstream analytics and ML use cases, designing data models and ingestion pipelines for accessibility and reuse
- Established data governance frameworks including quality validation standards and metric definitions to ensure trust and reproducibility in analytical outputs

Takeda Pharmaceuticals
2021 – 2022

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

- Implemented predictive modeling programs using machine learning for manufacturing and quality decision-making, working cross-functionally with engineering on system integration and deployment
- Partnered with IT teams on analytics platform architecture and tooling decisions, ensuring ML capabilities were scalable and accessible to operational stakeholders
- Built real-time monitoring dashboards fed by predictive models, enabling continuous performance tracking and iterative improvement

Regeneron Pharmaceuticals 2016 – 2021

Senior Process Data Scientist *Rensselaer, NY*

- Developed and deployed predictive models for biopharmaceutical manufacturing using ensemble methods, PLS, and PCA - iterating on model performance and expanding scope based on business impact over a five-year period
- Built automated data pipelines to ingest, clean, and transform process data from multiple systems into formats optimized for model training and inference
- Collaborated with cross-functional stakeholders to identify opportunities for ML integration and translate operational questions into modeling problems

EDUCATION

Doctor of Philosophy (PhD) in Mathematics 2014

Virginia Polytechnic Institute & State University

SKILLS & TECHNOLOGIES

AI & Machine Learning: LLM Development (OpenAI API) • RAG Pipelines • Prompt Engineering • Scikit-Learn • TensorFlow • PyTorch • HuggingFace • Ensemble Methods • Classification & NLP

Programming & Platforms: Python • SQL • Snowflake • Databricks • AWS • Pandas • Data Pipelines

Model Validation & Operations: Model Validation & Benchmarking • Confusion Matrix Analysis • Production Model Deployment • Performance Monitoring • Cross-Functional Engineering Collaboration

Domain & Governance: Pharmaceutical Industry • FDA Regulatory Data • Data Governance • Compliance & Regulated Environments • Unstructured Text Analysis