HARI CHANDANA VADIGE

Contact: (520)-283-5760| **Email:** chandana3940@gmail.com | LinkedIn | Portfolio

DATA SCIENTIST | MACHINE LEARNING ENGINEER | AI ENGINEER

Data Scientist and AI/ML Engineer who build intelligent systems that turn complex data into production-ready models and GenAI solutions across finance, healthcare, e-commerce, and SaaS. Experienced in driving the full ML lifecycle—from large-scale data pipelines and predictive modeling to RAG-powered AI platforms, MLOps automation, and cloud-native deployment on AWS, Azure, and GCP.

TECHNICAL SKILLS

- **Programming Languages:** Python, SQL
- **ML & DL:** Scikit-learn, TensorFlow, PyTorch, XGBoost, LightGBM, CatBoost, BERT, GPT Models
- **GenAI / LLMs:** OpenAI APIs, Hugging Face Transformers, LangChain, LangGraph, LangSmith, CrewAI, AutoGen, RAG Architectures
- **AI Agents:** LangGraph Agents, CrewAI Agents, Rule-based Decision Agents, Multi-Agent Evaluation Pipelines
- **Vector Databases:** FAISS, Pinecone, Chroma, Milvus
- **ML Algorithms:** Regression, Classification, Ensemble Methods, Gradient Boosting, Clustering, Time-Series Models, CNN/RNN architectures, Transformer-based models
- **Data Engineering:** PySpark, Kafka, Snowflake, BigQuery, Redshift, ETL/ELT Design
- **MLOps & Automation:** MLflow, Airflow, Kubernetes, Docker, Terraform, FastAPI, GitHub Actions, Azure DevOps, n8n, Cursor AI, MCP
- **Cloud:** AWS (S3, Lambda, ECS, SageMaker), Azure (Azure ML, AKS, Azure OpenAI), GCP (BigQuery, Vertex AI)
- **Evaluation & Safety:** SHAP, LIME, RAGAS, BLEU/ROUGE, ROC-AUC, Prompt Safety Testing, Bias Analysis
- **Visualization:** Tableau, Power BI, Plotly, Matplotlib, Seaborn
- **Security & Compliance:** HIPAA, PII/PHI Redaction, HITL Review Loops

PROFESSIONAL EXPERIENCE

Wells Fargo: AI-Driven Financial Intelligence & Risk Analysis Platform **Jan 2025 - PRESENT**
Role: AI Engineer

Project Overview

Built a GenAI-powered financial intelligence platform to assist risk analysts and portfolio managers with contextual analysis of financial statements, credit reports, and internal risk policies, enabling faster decision-making and improved risk transparency across lending and investment workflows.

Roles & Responsibilities

- Designed an **end-to-end Retrieval-Augmented Generation (RAG) architecture** to enable grounded financial question answering over structured (financial ratios, transaction summaries) and unstructured data (annual reports, credit memos, policy documents).
- Implemented **document ingestion and preprocessing pipelines** using PySpark, storing curated datasets in Amazon S3 and generating dense embeddings via transformer-based models for downstream retrieval.
- Built a **FAISS-based vector search layer** to support low-latency semantic retrieval, optimized for large volumes of financial documents with strict relevance and traceability requirements.
- Developed **LLM orchestration workflows using LangChain**, enabling prompt templating, context assembly, and tool calling for financial analysis use cases such as risk summarization, variance explanation, and covenant assessment.
- Modeled **stateful, multi-step reasoning flows using LangGraph**, allowing AI agents to decompose complex financial queries into retrieval, analysis, and validation steps before generating responses.
- Instrumented the GenAI pipelines with **LangSmith** to monitor prompt performance, agent execution paths, latency, and failure modes, enabling rapid iteration and debugging.
- Integrated **AI agents** to automate routine analyst workflows such as document triage, anomaly flagging, and comparative financial analysis, while keeping final decision-making human-in-the-loop.

- Exposed GenAI capabilities via **FastAPI-based services**, containerized using Docker and deployed on AWS ECS, ensuring secure and scalable access for internal applications.
- Conducted **RAG quality and safety evaluation** using RAGAS, along with prompt safety testing to mitigate hallucinations and ensure compliance with internal financial governance standards.

Environment: Python, SQL, XGBoost, Scikit-learn, Kafka, PySpark, FastAPI, MLflow, Airflow, Docker, AWS (S3, ECS, Lambda, SageMaker), SHAP

Johnson and Johnson: Predictive Clinical Risk & Patient Outcome Modeling Platform **Dec 2022 – Dec 2023**
Role: Machine Learning Engineer

Project Overview:

Developed a machine learning–driven clinical risk prediction platform to identify high-risk patient cohorts, support early intervention strategies, and improve outcome forecasting using structured electronic health record (EHR) and claims data across large healthcare systems.

Roles & Responsibilities

- Partnered with clinical stakeholders to translate healthcare outcome and utilization problems into **supervised ML formulations**, including risk classification, readmission prediction, and length-of-stay forecasting.
- Designed **feature engineering pipelines** for high-dimensional EHR and claims datasets, incorporating temporal aggregation, comorbidity indices, and treatment history while ensuring data consistency and auditability.
- Built and evaluated **multiple classical ML models** (Logistic Regression, Random Forest, XGBoost, LightGBM), selecting optimal models based on ROC-AUC, precision-recall trade-offs, and clinical interpretability.
- Developed **time-series modeling approaches** to capture patient trajectory trends and early deterioration signals across longitudinal healthcare records.
- Implemented **model explainability using SHAP** to surface feature contributions at patient and cohort levels, supporting clinician trust and regulatory review requirements.
- Orchestrated **end-to-end ML pipelines** using Azure ML and Airflow, covering data ingestion, training, validation, and batch inference workflows.
- Containerized training and inference workloads using Docker and deployed scalable services on **Azure Kubernetes Service (AKS)** for production-grade execution.
- Provisioned and managed cloud infrastructure using **Terraform**, enabling reproducible, version-controlled environments across development, staging, and production.
- Established **model versioning, experiment tracking, and lifecycle governance** with MLflow, ensuring traceability and rollback capabilities for regulated healthcare deployments.
- Built **performance monitoring dashboards** in Power BI to track model drift, prediction distributions, and outcome alignment over time.
- Implemented **data security controls**, including PII/PHI redaction and access governance, ensuring compliance with HIPAA and internal healthcare data policies.

TARGET: Personalization & Demand Intelligence Platform for E-Commerce **Oct 2021 – Nov 2022**
Role: Data Scientist

Project Overview

Built a large-scale machine learning platform to power product personalization, demand forecasting, and customer behavior intelligence for an e-commerce ecosystem, enabling improved conversion rates, inventory efficiency, and customer lifetime value.

Roles & Responsibilities

- Designed ML solutions for **personalized product recommendations**, search relevance optimization, and customer segmentation using behavioral, transactional, and clickstream data.
- Implemented **collaborative filtering and matrix factorization models** to capture user–item affinity, complemented by gradient boosting models to incorporate contextual and temporal signals.

- Built **customer clustering pipelines** to segment users based on purchasing behavior, engagement patterns, and lifecycle stages, enabling targeted marketing and merchandising strategies.
- Developed **time-series forecasting models** to predict product-level demand across categories and regions, supporting inventory planning and supply chain optimization.
- Engineered scalable **data pipelines using PySpark**, ingesting real-time and batch data into BigQuery for feature generation and analytical modeling.
- Trained, validated, and deployed ML models using **Vertex AI**, enabling automated retraining, versioning, and managed inference endpoints.
- Established **experiment tracking and model governance** using MLflow to compare model variants and ensure reproducibility across environments.
- Orchestrated end-to-end ML workflows with **Airflow**, coordinating data preparation, training, evaluation, and batch scoring jobs.
- Evaluated recommendation and forecasting performance using **domain-specific metrics** such as Precision@K, Recall@K, MAP, and forecast accuracy, aligning model outcomes with business KPIs.
- Collaborated with product and growth teams to integrate ML outputs into **customer-facing experiences**, including personalized homepages, category ranking, and promotional targeting.

State Farm: Product Intelligence & Experimentation Platform for SaaS Applications **Jan 2020 – Sep 2021**
Role: Associate Data Scientist | Claims fraud detection • Risk segmentation • Predictive analytics • Operational optimization
Domain: Insurance | Cloud: GCP

Project Overview

Developed a data science–driven product intelligence platform to analyze user behavior, measure feature adoption, and generate predictive insights for a B2B SaaS product, enabling data-backed product decisions, improved user engagement, and reduced churn.

Roles & Responsibilities

- Collaborated with product managers and engineering teams to define **data science problem statements** around user engagement, feature adoption, churn risk, and monetization funnels.
- Built **user-level analytical datasets** by aggregating product usage logs, event streams, and subscription data, enabling consistent cohort and lifecycle analysis.
- Applied **exploratory data analysis (EDA)** techniques to uncover usage patterns, behavioral drop-offs, and leading indicators of customer churn.
- Developed **classification and regression models** to predict churn probability, expansion likelihood, and long-term user value, supporting proactive product and customer success interventions.
- Performed **clustering and segmentation analysis** to identify distinct user personas based on behavior, feature usage intensity, and lifecycle stage.
- Designed and analyzed **A/B experiments** to evaluate the impact of new features, UI changes, and pricing experiments, translating statistical outcomes into actionable product recommendations.
- Implemented **time-series and trend analysis** to monitor feature adoption over time and assess the long-term impact of product releases.
- Evaluated model performance using **ROC-AUC, precision-recall, and lift metrics**, ensuring models were both statistically sound and business-relevant.
- Built **executive-ready dashboards in Tableau**, presenting insights on engagement, retention, and experiment outcomes to product and leadership stakeholders.
- Communicated analytical findings through **clear narratives and data storytelling**, enabling non-technical teams to act confidently on insights.

A C A D E M I C S

- **Master's in information science – Specialization in ML** **Jan 2024 –Dec 2025**
  University of Arizona, Tucson, USA