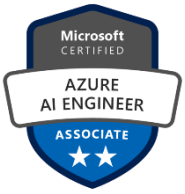
AI/ML Engineer
Lokeshwar Kodipunjula
Email: lokeshwarkodipunjula@gmail.com
Contact: +1 6823901628



Professional Summary:

- AI/ML Engineer with 4 years of experience designing, developing, and deploying scalable machine learning and NLP solutions across insurance, healthcare, and fintech domains.
- Proven expertise in building supervised, unsupervised, and deep learning models using Python, scikit-learn, XGBoost, TensorFlow, and Keras to solve complex business problems such as fraud detection, patient risk prediction, and document classification.
- Strong hands-on experience with NLP frameworks like spaCy, Transformers (Hugging Face), and NLTK for extracting insights from unstructured text data in clinical, customer service, and financial domains.
- Developed and optimized real-time and batch data pipelines using PySpark, Apache Kafka, and Apache Flink to handle large-scale inference workloads and feature engineering efficiently.
- Skilled in deploying ML models as RESTful microservices using FastAPI and Docker, orchestrated via Kubernetes (AWS EKS) for production-grade resilience and scalability.
- Experienced in implementing robust MLOps practices using MLflow, Airflow, CI/CD pipelines, and model monitoring stacks (Prometheus, Grafana) to ensure model traceability, reproducibility, and performance over time.
- Hands-on with LLM-driven and RAG-based intelligent systems, including embeddings, vector search, and function/tool calling, to enhance retrieval accuracy and contextual automation in enterprise AI use cases.
- Built and tuned model evaluation frameworks (offline / online / golden set / HIL) and monitoring dashboards to measure quality, latency, cost, and safety metrics for production AI models.
- Applied Responsible AI, observability, and guardrail frameworks to ensure fairness, transparency, privacy compliance, and auditability of deployed AI solutions.
- Demonstrated ability to integrate model outputs with BI tools (Power BI, Tableau) and backend systems, enabling real-time decision support and actionable insights for business stakeholders.
- Deep understanding of regulatory and compliance frameworks (HIPAA, RBI) with hands-on implementation of security best practices using TLS encryption, HashiCorp Vault, and audit logging.
- Worked in Agile/Scrum teams alongside data scientists, DevOps, product managers, and domain experts to deliver and deploy end-to-end AI solutions.

Technical Skills:

Category	Tools & Technologies
Programming Languages	Python (3.x), SQL, R, Java
ML Libraries	scikit-learn, XGBoost, LightGBM, SHAP
Deep Learning	TensorFlow, Keras, PyTorch
NLP & Model Explainability	spaCy, Transformers (Hugging Face), NLTK, SHAP, LangChain, OpenAI API, Embeddings, RAG pipelines
Data Engineering	PySpark, pandas, NumPy, SQLAlchemy, Apache Kafka, Apache Flink, Databricks, Delta/Unity Catalog
ETL & Workflow	Apache Airflow
Model Tracking & MLOps	MLflow, CI/CD, Git, Bitbucket
APIs & Web Frameworks	FastAPI, RESTful APIs
Cloud Platforms	AWS (S3, EC2, Lambda, EKS), Azure (HIPAA-compliant), Redis
Containerization	Docker, Kubernetes (K8s), AWS EKS
Monitoring & Observability	Prometheus, Grafana
BI & Reporting	Power BI, Tableau, Jupyter Notebook, Markdown
Security & Compliance	HIPAA, RBI guidelines, HashiCorp Vault, TLS encryption
Development Methodologies	Agile, Scrum
Version Control & Tools	Git, Bitbucket, JIRA, Confluence

Professional Experience:

AIG, NY **Nov 2024 – Till Date**
AI-ML Engineer
Project: SentinelAI
Responsibilities:

- Designed and implemented supervised and unsupervised ML models using Python, scikit-learn, and XGBoost to predict claim fraud likelihood, reducing false positives by 23%.
- Developed deep learning pipelines using TensorFlow and Keras for computer vision-based document classification and anomaly detection in scanned claim forms.
- Prototyped RAG (Retrieval-Augmented Generation) pipelines combining OpenAI embeddings and vector databases (FAISS/Chroma) for policy document summarization and claim reasoning tasks.
- Developed and integrated evaluation framework for NLP models including golden set comparisons and human-in-the-loop (HIL) feedback loops, improving accuracy metrics by 18%.
- Implemented Responsible AI guardrails ensuring privacy compliance, safety checks, and explainability for production AI features, aligning with internal audit and risk policies.
- Built and deployed NLP pipelines using spaCy and Transformers (Hugging Face) for named entity recognition and sentiment analysis from customer support transcripts.
- Engineered real-time data pipelines with Apache Kafka and batch ETL flows using PySpark for preprocessing and feature engineering on large-scale policy and claim datasets.
- Created and scheduled end-to-end workflows using Apache Airflow, ensuring reliable ML pipeline orchestration across training, validation, and deployment phases.
- Containerized ML microservices using Docker and deployed scalable inference endpoints using Kubernetes on AWS EKS, optimizing latency for fraud detection scoring.
- Utilized MLflow for experiment tracking, model versioning, and lifecycle management during collaborative model development cycles.
- Integrated model outputs into business dashboards using Power BI and exposed insights via RESTful APIs for consumption by underwriting and fraud analytics teams.
- Conducted regular model monitoring using Prometheus and Grafana to track drift and retraining triggers, ensuring model relevance in dynamic environments.
- Collaborated cross-functionally with data scientists, DevOps, and domain experts in Agile/Scrum settings to iterate on solution design and delivery.

Environment: Python, scikit-learn, TensorFlow, Keras, PySpark, Apache Kafka, Airflow, Docker, Kubernetes, AWS EKS, MLflow, Power BI, Prometheus, Grafana, Agile.

VNS Health, NY **Jan 2024 – Oct 2024**

Python Data Scientist

Project: Clinical Insight Engine – NLP & ML for Patient Risk Prediction

Responsibilities:

- Designed and implemented NLP pipelines using spaCy and scikit-learn to extract clinical entities and sentiment from provider notes, enabling integration with structured risk models.
- Built and validated patient risk prediction models using XGBoost, Random Forest, and Logistic Regression on top of feature sets engineered from time-series vitals, lab results, and NLP-derived features.
- Preprocessed large-scale EHR datasets (HL7/C-CDA formats) using pandas, NumPy, and SQLAlchemy, ensuring data consistency and quality for modelling workflows.
- Developed and maintained automated model training pipelines using Airflow and version-controlled experimentation with MLflow, ensuring reproducibility across clinical releases.
- Tuned and evaluated model performance using stratified cross-validation, SHAP for explainability, and AUROC/F1 metrics to support clinical adoption and trust.
- Deployed models into a clinical decision support interface using FastAPI and Docker, ensuring scalable inference in production aligned with HIPAA compliance.
- Collaborated with clinical informaticists and data engineers in Agile sprints, documenting findings and presenting model behavior using Jupyter, Markdown, and Tableau dashboards.

Environment: Python 3.10, Jupyter, pandas, NumPy, scikit-learn, spaCy, XGBoost, SQLAlchemy, PostgreSQL, MLflow, Airflow, Docker, FastAPI, SHAP, Git, Tableau, Agile (Scrum), HIPAA-compliant cloud (Azure).

Axis Bank, India **July 2021 – Aug 2023**

Jr. ML Engineer

Project: Real-Time Transaction Fraud Detection Engine

Responsibilities:

- Designed and trained supervised machine learning models using Python (scikit-learn, XGBoost) on historical transactional datasets to detect patterns indicative of fraudulent activity.
- Developed real-time scoring pipelines using Apache Kafka, Apache Flink, and Redis, enabling sub-second inference latency for high-volume transaction streams.
- Performed feature engineering on streaming and batch datasets using PySpark and SQL, deriving risk indicators such as transaction velocity, merchant risk scores, and device fingerprinting.
- Implemented and maintained model versioning, drift detection, and retraining pipelines using MLflow, Airflow, and Docker within a CI/CD framework.

- Collaborated with security and compliance teams to ensure the fraud detection logic adheres to RBI guidelines and internal governance protocols, implementing necessary audit trails and encryption mechanisms using HashiCorp Vault and TLS.
- Deployed models to production using RESTful APIs via FastAPI, containerized with Docker, and orchestrated in Kubernetes (K8s) clusters for scalability and resilience.
- Monitored model performance in production using Prometheus and Grafana, tracking precision, recall, latency, and false positive rate, while participating in monthly A/B testing cycles to evaluate the effectiveness of new model versions and rule-based overrides.

Environment: Agile, Scrum, AWS (S3, EC2, Lambda, EKS), Docker, Kubernetes, Airflow, CI/CD, JIRA, Bitbucket.

Education:

Master's: University of Texas at Arlington - Arlington, Texas, United States
Bachelor's: Kingston Engineering College - Vellore, Tamil nadu, India

Certification:

- Microsoft Certified: Azure AI Engineer Associate
- **LinkedIn:** https://www.linkedin.com/in/lokeshwar-k-753a95386/