

LOKESH CHOWDHARY YELLAMANCHALI

Boston, MA | lokeshcy510@gmail.com | +1 857-385-2500 | LinkedIn:

PROFESSIONAL SUMMARY

Generative AI and AI/ML Engineer with experience delivering production-ready solutions across Databricks, Azure, AWS, and GCP. Skilled in building RAG pipelines, LLM-powered applications, predictive models, ML workflows, and scalable data pipelines using Python, PySpark, Spark, Delta Lake, MLflow, and cloud AI services. Proven ability to improve search relevance, automate document intelligence, optimize model deployment, and reduce cloud processing costs. Strong background in healthcare AI, MLOps, vector search, prompt engineering, and enterprise AI solution delivery

TECHNICAL SKILLS

Languages & Databases: Python, SQL, PySpark, R, Java, Scala, Bash, SQL, NoSQL

AI/ML & Gen AI: Machine Learning, Deep Learning, LLMs, RAG, Prompt Engineering, Embeddings, Vector Search, Semantic Search, Predictive Modeling, NLP, Anomaly Detection

Frameworks & Libraries: Scikit-learn, TensorFlow, PyTorch, Hugging Face, LangChain, LlamaIndex, MLflow, Pandas, NumPy

Databricks & Big Data: Azure Databricks, AWS Databricks, Apache Spark, Delta Lake, Unity Catalog, Delta Live Tables, Databricks Workflows, Model Serving, Vector Search

Cloud & MLOps: Azure, AWS, GCP, Azure OpenAI, SageMaker, Vertex AI, BigQuery, Docker, Kubernetes, Airflow, Git, CI/CD, Model Monitoring

Visualization & Governance: Power BI, Tableau, Looker, Data Quality, Data Lineage, AI Governance, Model Explainability, Bias Assessment

PROFESSIONAL EXPERIENCE

Generative AI Engineer | Databricks

June 2024 - Present | USA

- Designed and deployed production-ready **Generative AI and RAG solutions** using **Azure Databricks, AWS Databricks, PySpark, Delta Lake, MLflow, embeddings, vector search, and LLMs**, reducing enterprise document search time by **40%**.
- Built scalable **LLM-powered applications**, including enterprise chatbots, semantic search tools, document summarization workflows, and AI knowledge assistants for business users.
- Processed over **10M+ records daily** using **Apache Spark, PySpark, SQL, and Delta Lake** across **Azure Databricks and AWS Databricks** environments.
- Implemented **Databricks Vector Search** and embedding-based retrieval, improving search relevance by **45%** compared to keyword-based search methods.
- Automated summarization and Q&A for **10,000+ enterprise documents**, reducing manual review effort by **60%**.
- Deployed AI inference workloads using **Databricks Model Serving**, achieving response latency below **2 seconds** for business-critical applications.
- Optimized Spark jobs, SQL queries, cluster configurations, and workload execution, reducing pipeline runtime by **35%** and cloud compute costs by **20%**.
- Implemented **MLflow, Unity Catalog, Databricks Workflows, Delta Live Tables, and CI/CD pipelines** to support model governance, secure data access, and production AI operations.

AI/ML Engineer | Verily

Mar 2022 - Aug 2023 | India

- Designed and deployed scalable **AI/ML solutions** for healthcare and life sciences use cases using **Python, SQL, Scikit-learn, TensorFlow, PyTorch, GCP, and AWS**, improving prediction accuracy by **25%**.
- Built end-to-end ML pipelines for data ingestion, preprocessing, feature engineering, model training, validation, deployment, and production monitoring.
- Processed large-scale healthcare datasets, improving model-ready data quality by **40%** through automated validation, cleansing, and transformation workflows.
- Automated ML workflows using **MLOps, CI/CD, Git, Docker, Airflow, and MLflow**, reducing manual effort by **45%** and deployment cycle time by **30%**.
- Deployed production-ready ML models on **GCP and AWS** for batch and real-time inference, supporting reliable healthcare analytics and operational decision-making.
- Implemented model monitoring for **data drift, accuracy, latency, prediction quality, and performance degradation** to ensure stable production AI operations.
- Applied responsible AI practices, including **explainability, bias assessment, validation, and governance**, to support trustworthy healthcare AI solutions.
- Collaborated with cross-functional healthcare, data science, engineering, and product teams to convert business requirements into production-ready ML solutions.

EDUCATION

Northeastern University - Boston, MA

Master of Science in Data Analytics Engineering - May 2025

KL University - Vijayawada, Andhra Pradesh, India

Bachelor of Technology in Computer Science Engineering - March 2023

CERTIFICATIONS

- Databricks Certified Generative AI Engineer Associate
- Microsoft Certified: Azure AI Engineer Associate
- Google Cloud Professional Machine Learning Engineer