SHREYANS SATPATHY

+1 (213) 425-1297 | shreyans12@g.ucla.edu | LinkedIn | Los Angeles, CA

ML & AI Engineering

UCLA Master's Candidate (Graduating Dec 2026) and **AI/Data Science Engineer** with **2.5+ YOE** building production AI systems on robust data foundations. Proven track record of shipping RAG pipelines, LLM applications, and ML models while architecting the scalable data platforms (Databricks, Azure, Spark) that power them.

Education

UCLA ANDERSON SCHOOL OF MANAGEMENT, *Masters of Science in Business Analytics* *Sep 2025 – Dec 2026*
- Customer Analytics, Data Management, Forecasting, Optimisation & Simulation, Recommender Systems, Prescriptive Analytics

VELLORE INSTITUTE OF TECHNOLOGY (India), *B.Tech in Computer Science and Engineering* *Aug 2019 – Jun 2023*
- Algorithms, Distributed Computing, Operating Systems, Database Management Systems, Machine Learning, Artificial Intelligence

Experience

MERCEDES-BENZ RESEARCH & DEVELOPMENT Bengaluru, India

Data Science Engineer *Aug 2023 – Aug 2025*
- Architected **FACTS**, a unified field quality analytics platform consolidating warranty data from **10+ sources** (Xentry, VeDoc, TIPS, GuK) into a single lakehouse; **delivered $1.6M in annual cost savings** and **reduced claims lifecycle time by 40%**.
- Built a **RAG-based Workshop Co-Pilot** using Databricks Vector Search, hybrid retrieval (semantic + keyword), and Unity Catalog metadata filtering; enabled technicians to query historical repair data in real-time, **reducing diagnostic research time and senior staff escalations.**
- Developed **auto-clustering pipelines** for multilingual workshop claims using LLM-driven text translation and **fine-tuned BERT** for automotive-specific context; achieved **91% clustering accuracy** across 6 languages.
- Implemented **ML classification model** on warranty claims to auto-identify Top Issues and Severity levels, enabling proactive quality containment and prioritized root-cause investigation.
- Designed **A/B experiments** benchmarking GenAI (LLM) models vs. legacy rule-based systems for anomaly detection; achieved **91% accuracy** in identifying defect patterns and fraud anomalies.
- Engineered **end-to-end data pipelines** using Databricks (Delta Lake), Spark, and Azure Data Factory with **Medallion Architecture** (Silver → Gold → Platinum); processed millions of warranty records with incremental ingestion and partition-optimized Spark jobs.
- Built **KPI dashboards** in Tableau/Power BI tracking defect frequencies, resolution time, and warranty costs; improved critical KPI visibility by **25%** for cross-functional stakeholders (RD, GSP, QEC).

Data Engineering Intern *Jan 2023 – Aug 2023*
- Optimized data ingestion using **parallelized API calls on Spark**, reducing pipeline runtime by **90%** and enabling faster iteration for downstream ML models.
- Built **data quality validation frameworks** catching logic and schema issues pre-production; reduced pipeline failure rates by **30%**.
- Re-architected **semantic models** with supplier analytics teams, improving dashboard latency and ensuring consistent data accuracy for reporting layers.

Skills

AI/ML: LLMs, RAG Pipelines, Vector Search, LangChain, BERT Fine-Tuning, Scikit-learn, MLflow, Forecasting/Time Series, A/B Testing
Data Engineering & Platforms: Apache Spark, PySpark, Databricks (Delta Lake), Azure (ADF, Synapse, ADLS Gen2), Kafka, Airflow, ETL/ELT, Data Lakehouse, Medallion Architecture
Languages & Tools: Python, SQL, PySpark, GCP, AWS, Tableau, Power BI, Git, CI/CD, REST APIs
Domain: AI Systems, Data Platforms, Machine Learning, Product Analytics, Data Visualization, Data Modeling

Certifications

- **Databricks Certified Data Engineer Professional** *May 2025*
  Spark, Delta Lake, and Lakehouse architecture.
- **Microsoft Certified: Azure Data Engineer Associate (DP-203)** *Nov 2024*
  Data pipelines and analytics on Microsoft Azure.

Academic Projects

AI in MedTech Strategic Analysis | UCLA Biodesign & BCG
- Conducted EDA on investment trends and operational costs, **identifying key correlations** between funding stages and AI adoption rates to support hypothesis-driven strategic recommendations.
- Built a quantitative framework forecasting **AI integration** ROI; translated **cost-benefit metrics** into capital allocation guidance.