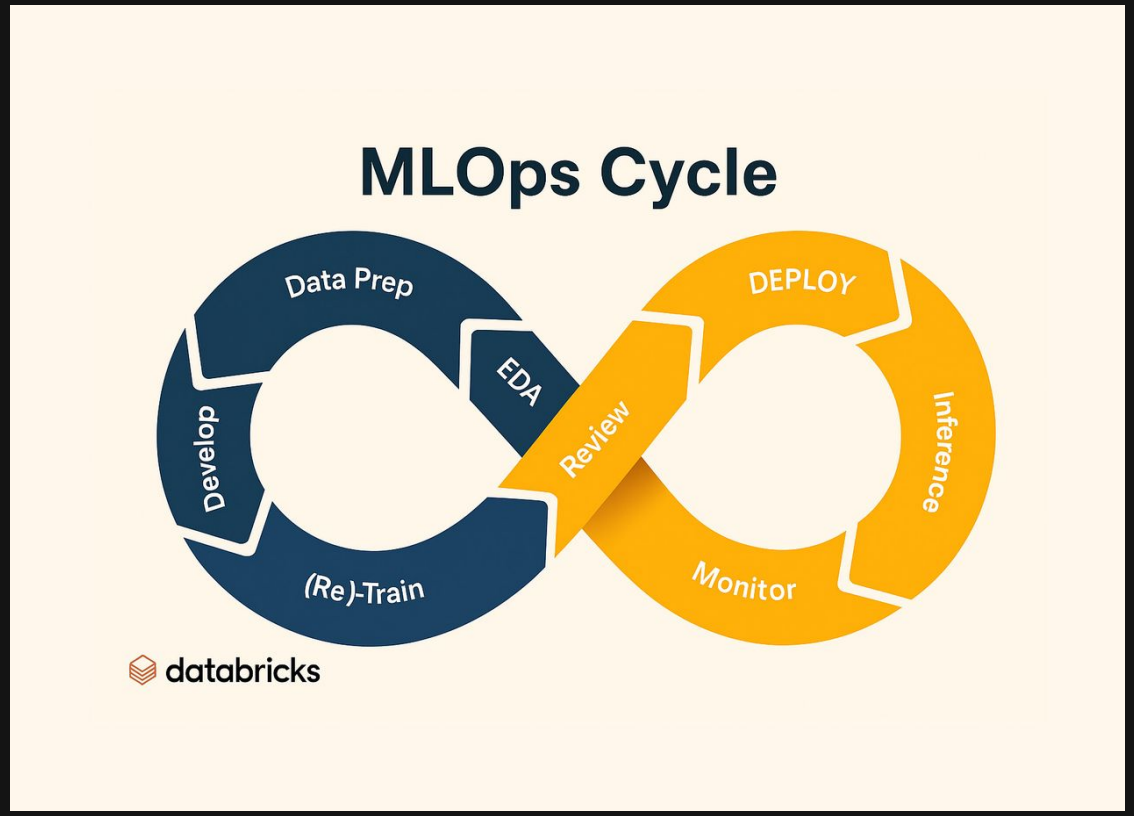


SEP 25 2025

From Lab to Launch: Enabling Continuous AI with MLOps

Chaitanya

About Me

Background

I’ve worked as a Machine Learning Engineer and MLOps Specialist for few years. My passion for intelligent systems started early, inspired by a love of solving problems with technology. That curiosity now drives me to design pipelines and deploy models that scale.

EDUCATION

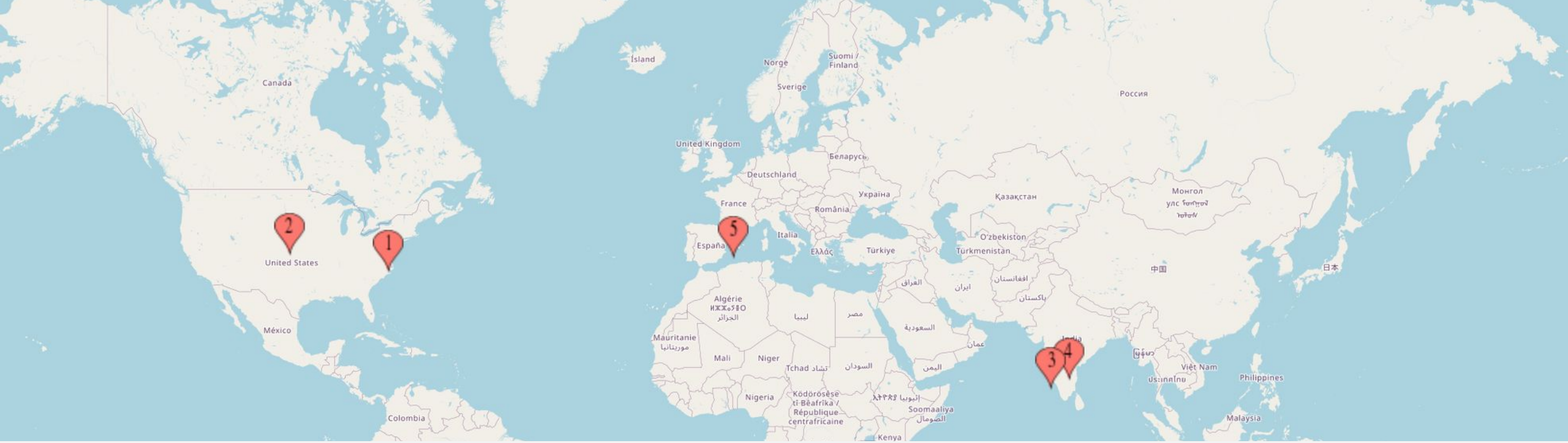
- VIRGINIA TECH (MASTER’S)

EXPERIENCE

- SERVICE CENTER METALS, VIRGINIA
- VIRGINIA TECH TRANSPORTATION INSTITUTE, VIRGINIA
- DELL TECHNOLOGIES , INDIA

PASSIONS

- TRAVELING , EXPLORING NEW PLACES



Conferences and Invited Talks

1. **The Current State of AI in Transportation through Topic Modeling**
   AI Expo 2024, Washington, D.C, USA

2. **Advancing Agriculture Communication with a Secure Web Platform (“Ag Corp”)**
   NAPDC 2024, Nebraska, USA

3. **Sentiment Analysis of Online Customer Reviews (Flipkart Case Study)**
   ICONAT 2023, Goa, India

4. **Identifying Operational Regimes of Electric Arc Furnace via ML Clustering**
   ICETCI 2023, Hyderabad, India

5. **Parking Demand Hotspots & Surge Prediction with ML**
   Urban Transitions Conference 2024, Sitges, Spain

6. **From Clicks to Insights: Analysing Online Customer Reviews**
   Technology Analysis & Strategic Management Journal, 2025

Agenda

1. Project Objectives
2. Model In Action
3. Modeling Pipeline
4. Deployment Pipeline
5. Key Technical Decisions

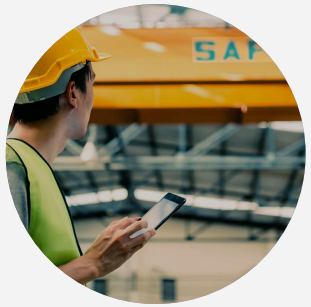
Projects:

- Computer Vision Object Detection for Safety Compliance
- LLM-Powered Maintenance Chatbot

Computer Vision Object Detection for Safety Compliance

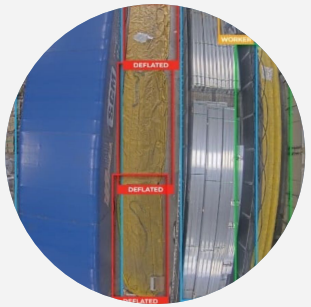
Business Problem : Proactively detect safety hazards during truck loading—where workers operate at heights up to ten feet—to reduce fall risks, improve compliance, and ensure a safer workplace.

Key Performance Indicators -



Worker Safety

Operational Monitoring



Airbag Compliance

THE PILOT


SAFETY CRITICAL: FALL RISK DETECTED
ELEVATION:
~14 FT
WARNING: RIGHT-SIDE AIRBAGS ABSENT

Model In Action

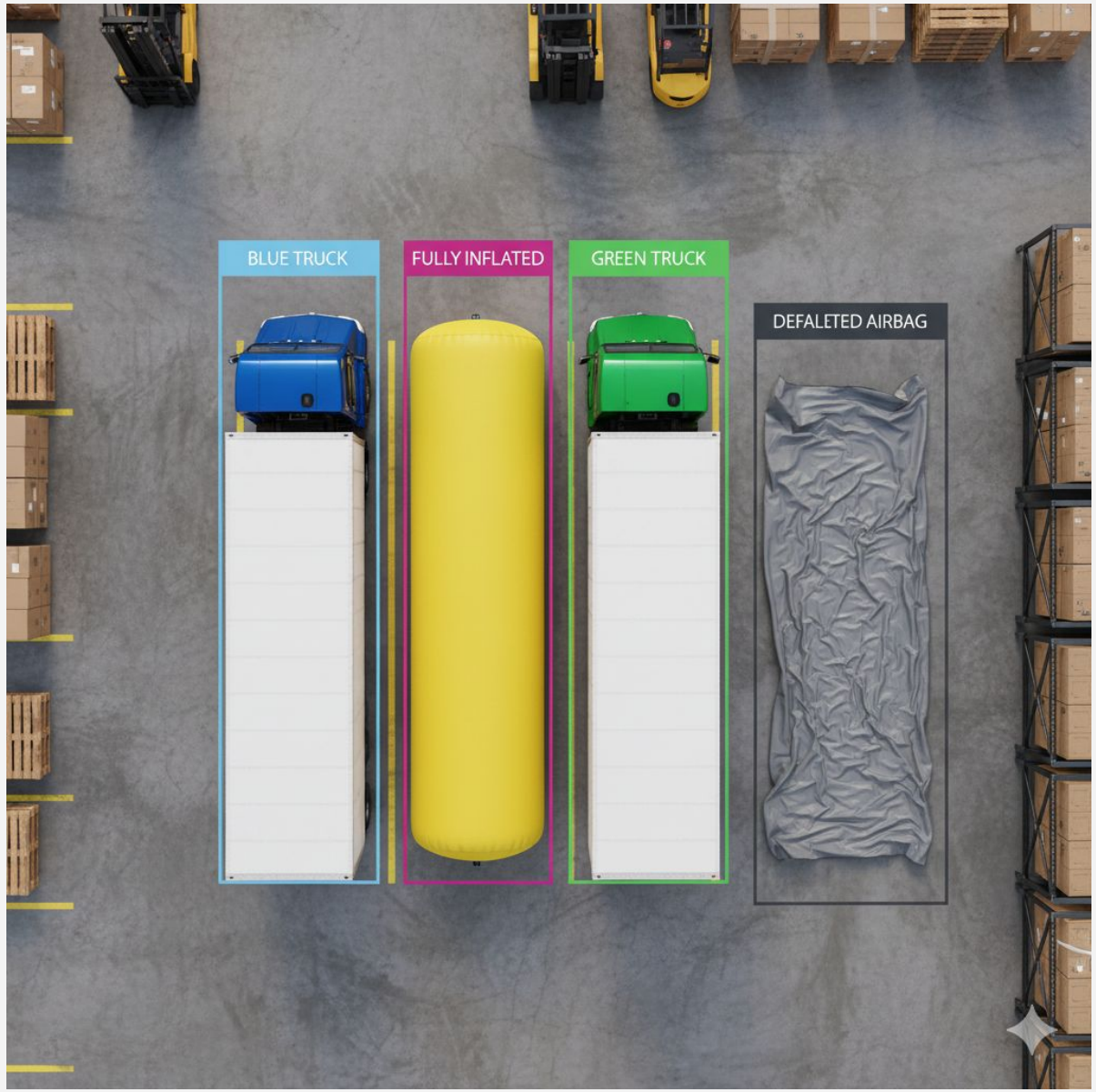
The image is tagged to indicate a truck is positioned in the loading bay, and the airbag is not inflated.

Bounding Box -

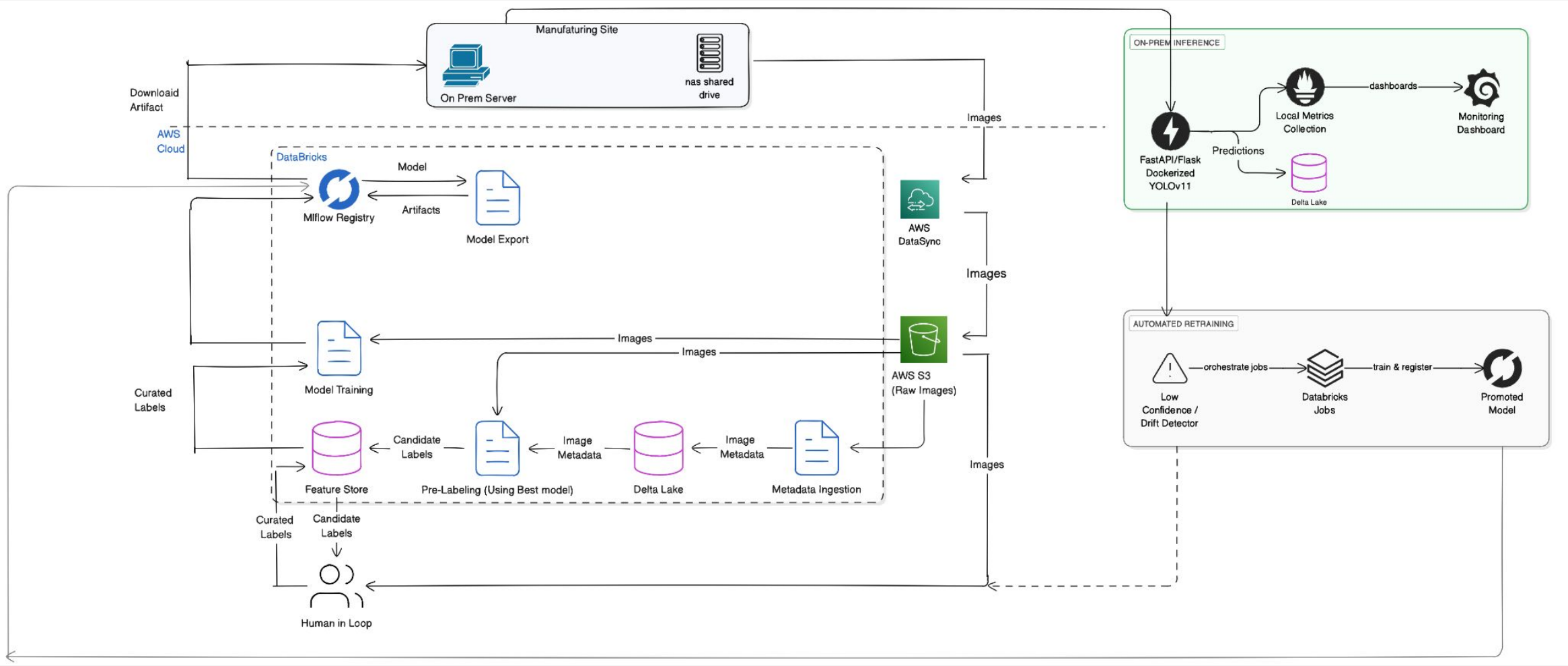
Blue, Green -> Truck Present

Pink -> Airbag Inflated Status

Black -> Airbag Deflated Status



Modeling Pipeline



Deployment Pipeline

1. Developer Workspace

2. Github Actions

3. CI Pipeline
   - Code Quality
   - Run Tests
   - Build Package

4. CD Pipeline
   - Setup CLI
   - Deploy Bundle (DAB)
     - Deploy to ACC
     - Deploy to PRD
   - Tag Version

5. Databricks Job Scheduler

6. Docker Registry + On Prem Inference

7. Lake House Monitoring

Key Technical Decisions

DAB Bundle = Infrastructure Package

- Contains job definitions, code, and configurations
- Gets deployed to Databricks workspace
- Creates actual Databricks jobs and schedules

Scheduled Jobs = ML Pipeline Execution

- Runs biweekly
- Executes the actual ML workflow
- Trains models, processes data, deploys models
- Managed by Databricks job scheduler

CI/CD = Code Deployment Pipeline

- Runs immediately when code changes
- Deploys the DAB bundle (job definitions)
- Does NOT run the ML pipeline
- Updates what will run, not when it runs

Supervised Drift Detection

- Human in Loop
- Relabeling images with low confidence and trigger pipeline

** Implemented Ultra-Fast Python Package Manager

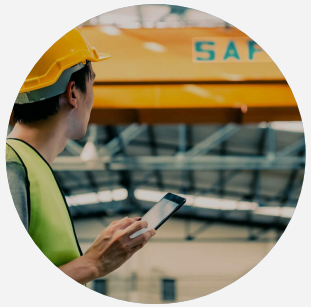
Business benefits

- Generated $500K in recurring annual savings.
- Computer Vision significantly improved daily airbag compliance rates, surging from below 25% to consistently surpassing 90% daily compliance throughout the pilot.
- The system detected almost 100 instances of improperly deployed airbags, triggering 253 violations across five bays, prompting immediate alerts and safer operations.
- 312 days with no incidents

LLM-Powered Maintenance Chatbot (SQL + RAG)

Business Problem : Build an LLM-powered maintenance assistant that enables engineers to query both structured data and unstructured documents in natural language, providing faster, more accurate, and explainable answers to critical maintenance questions.

Key Performance Indicators -

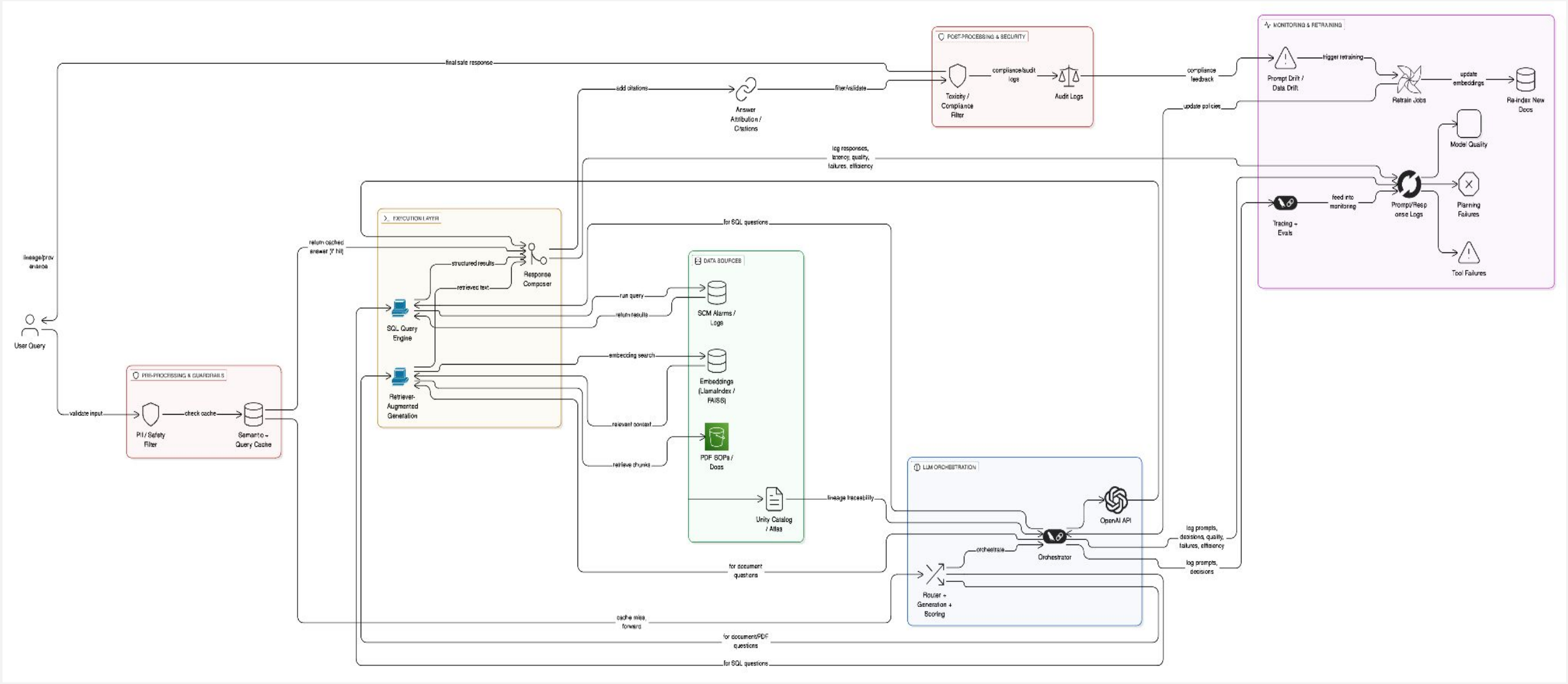


First-Time Resolution Rate

Compliance & Traceability

Maintenance Downtime Reduction

Modeling Pipeline



Deployment Pipeline

1. **Developer Workspace**
2. **CI Pipeline (GitHub Actions)**
3. **CD Pipeline (GitHub Actions)**
4. **Databricks Job Scheduler**
5. **Inference Runtime**
6. **Retrieval Subsystem**
7. **Observability & Monitoring**
8. **Feedback & Retraining Loop**

Key Decisions

- **Why hybrid (SQL + RAG)?**:

  Structured vs. unstructured queries → maximum coverage.

- **Why guardrails , audit logs?**:

  Ensure compliance , Essential for regulated environments (traceability, accountability).

- **How we measure drift**:

  Sql - No of failed executions

  RAG - Feedback from end user

Business Benefits

First-Time Fix Rate

- Current industry average = **~65%**.
- With contextual answers + source-backed instructions, chatbot can raise this to **85-90%**, reducing repeated interventions by **25%**, saving **~$100K annually** in labor and spare parts.

Knowledge Democratization

- Cuts onboarding for new maintenance engineers from **2 months to ~3 weeks**, reducing training costs by **50%** per hire.
- At scale this equals **$150K+ in annual savings**.

End of Workflow: Thanks