

# Ensuring ESG Integrity in Carbon Markets through Llama 3 and Apache Spark

Carbon credits are traded in global markets, but greenwashing and fraudulent transactions undermine the integrity of such markets. By leveraging big data / AI techniques, we seek to identify suspicious transactions and enhance transparency in carbon credit trading.

## What is greenwashing?

**Misleading** the public to believe that a company or other entity is doing more to **protect the environment** than it is

Source: UN Climate Action

### Why does it matter?



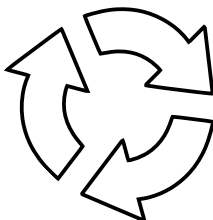
Promotes false solutions to combat climate change



Distracts from and delays the pursuit of concrete, credible actions

## A tool for transparency and accountability

Regulatory authorities



Environmental NGOs & advocacy

Business

## Data

### Voluntary Registry Offsets Database

#### UC Berkeley Carbon Trading Project

- **Structured** data source
- 10,145 records from:
  - American Carbon Registry
  - Climate Action Reserve
  - Gold Standard
  - Verra

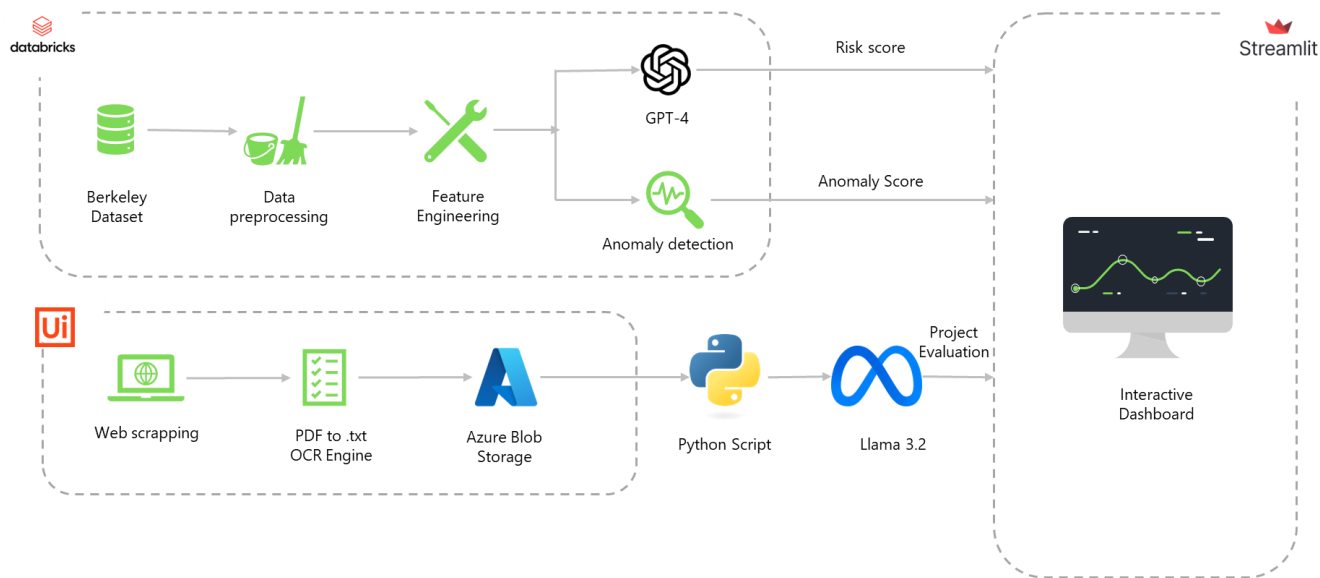
### Verra Project Design Documents (PDDs)

#### Verra Registry

- **Unstructured** textual data
- Publicly available reports in **PDF** format

**60%** of the world's **2000** largest publicly traded companies have established net-zero targets by 2030

# Methodology



## Data Ingestion & Cleaning

- Apache Spark
- UiPath UI Automation
- UiPath OCR Engine
- Azure Blob Storage

## Feature Engineering

- Apache Spark

## Dashboard Deployment

- Streamlit

## Fraud Detection Modeling

- Isolation Forest from PySpark ML
- Llama
- OpenAI

# Key Takeaways



Brings big data and Gen AI tools to carbon markets



Helps stakeholders avoid reputational and financial risks



Enhances accountability and supports ESG integrity in global climate initiatives

TEAM 1



Anne Hackman  
hackm030@umn.edu



Apurva Baru  
baru0016@umn.edu



Fnu Aastha  
aasth003@umn.edu



Qirui Zheng  
zhen0780@umn.edu



Sameeksha Mohan  
mohan225@umn.edu



Shang-Chien Wang  
wan02526@umn.edu