

# Scaling AI Workloads with Apache Airflow

Shubham Mehta  
Sr. Product Manager, AWS

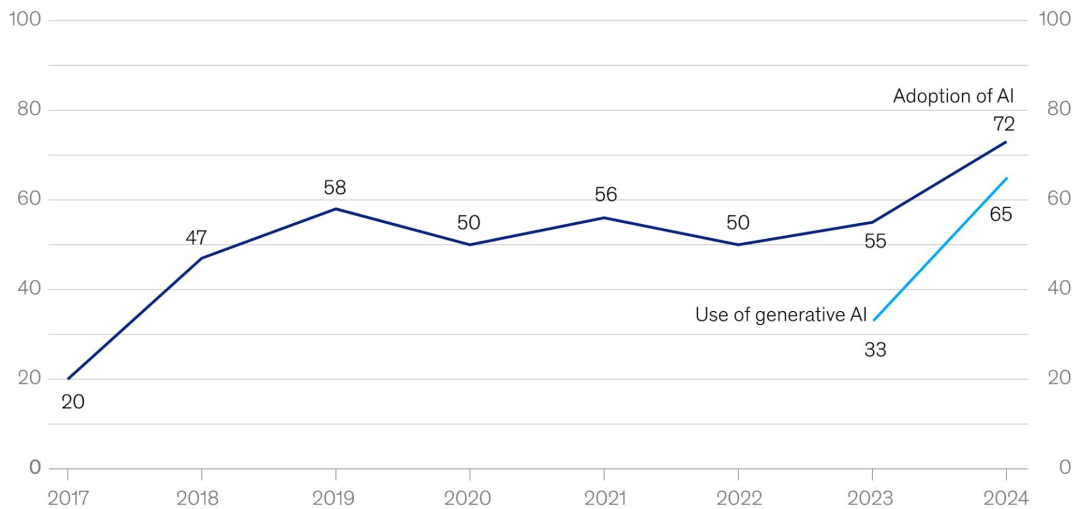
Rajesh Bishundeo  
Software Development Manager, AWS



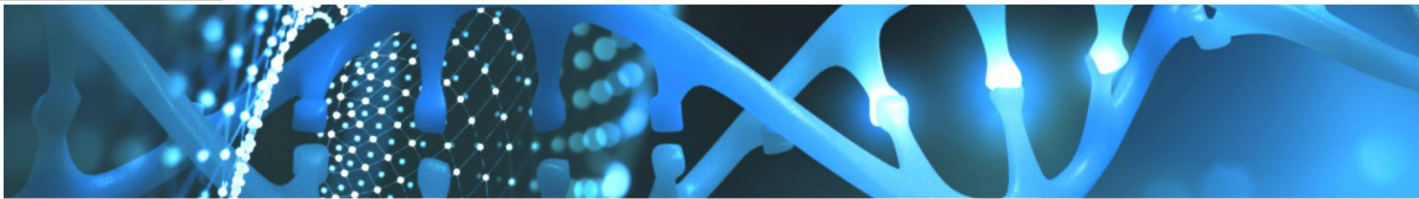
# The AI Revolution

AI Adoption worldwide has increased dramatically in the past year, after slow growth for years.

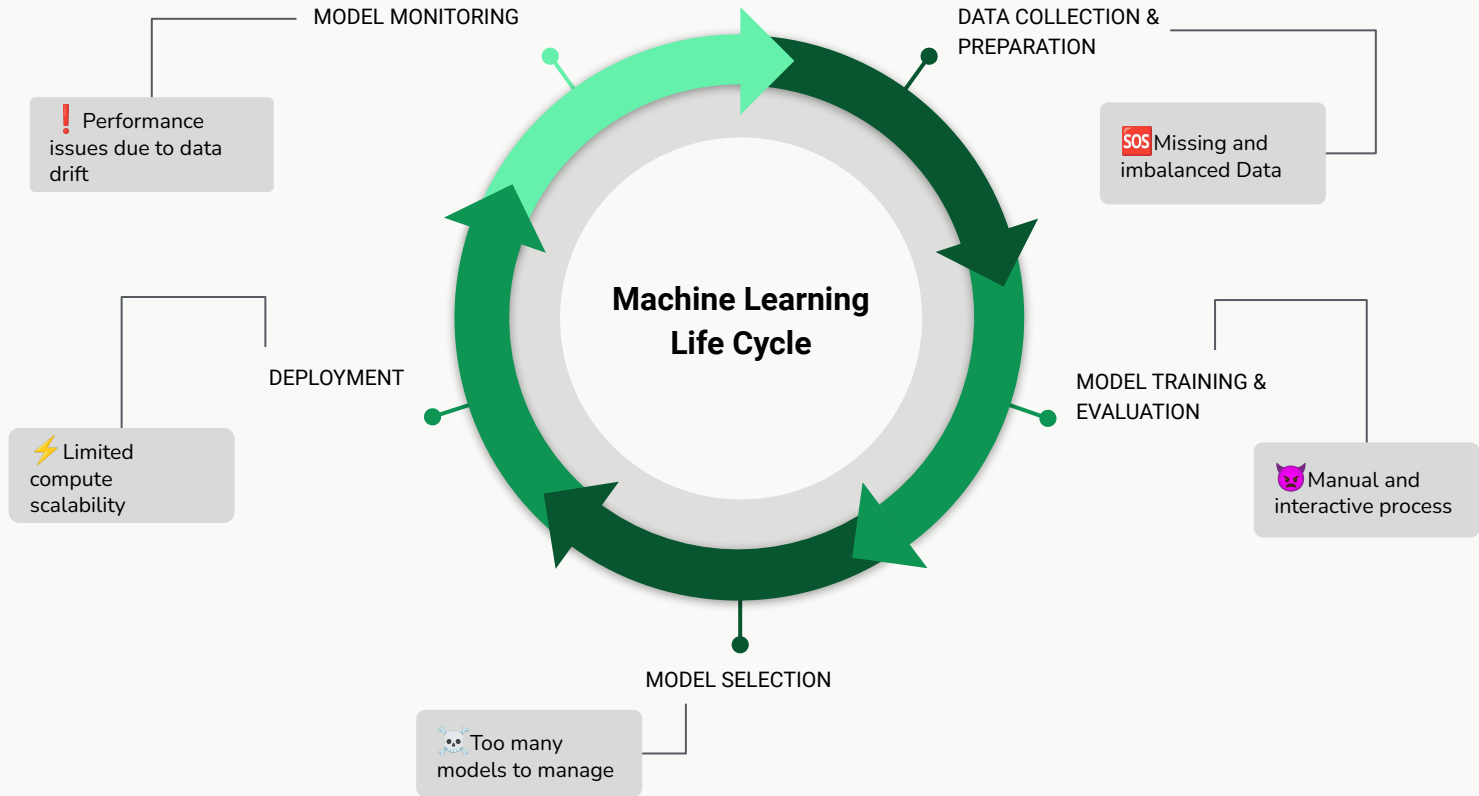
Organizations that have adopted AI in at least 1 business function,<sup>1</sup> % of respondents



# The AI Revolution



# The ML Lifecycle

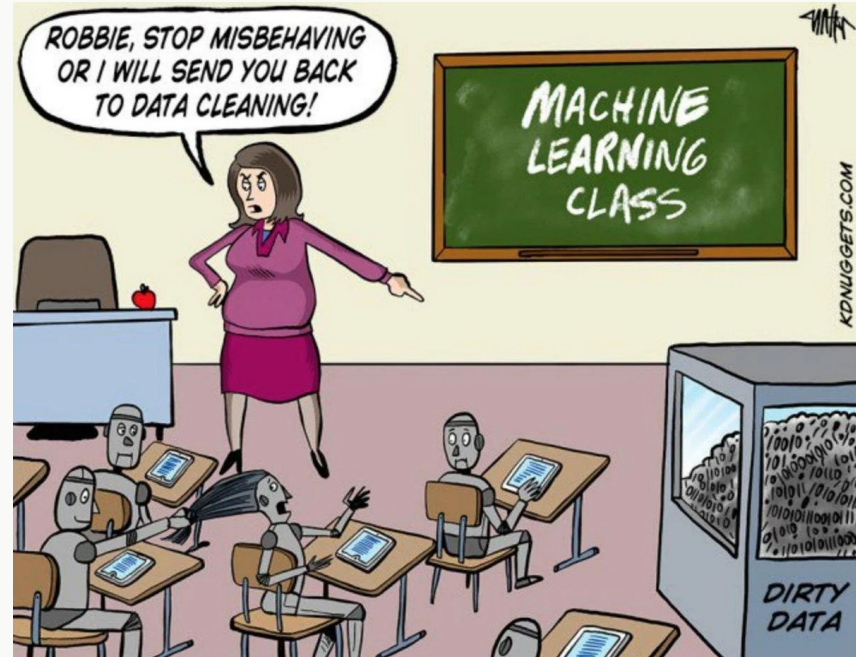


# Common Pitfalls in AI Operations



# Data Collection and Preparation Woes

Issues with data quality, consistency, and imbalance.



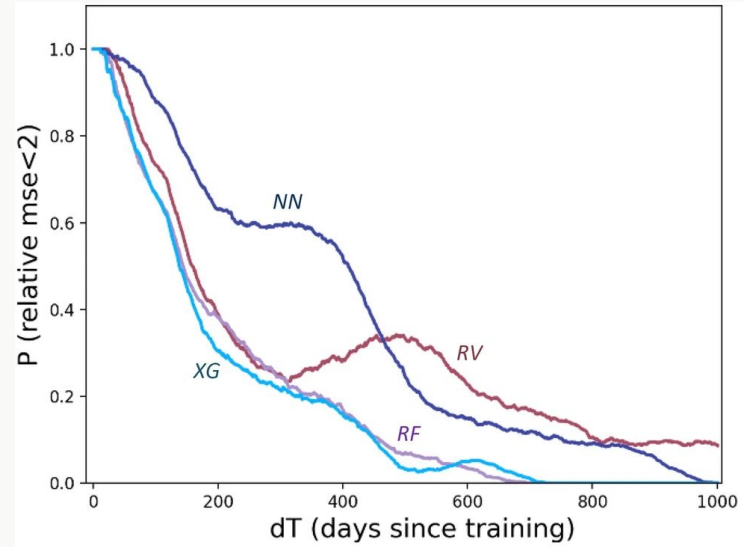
# The Manual Model Training Grind

Issues with manual and interactive model training processes



# Performance Degradation

Challenges in detecting and addressing data drift



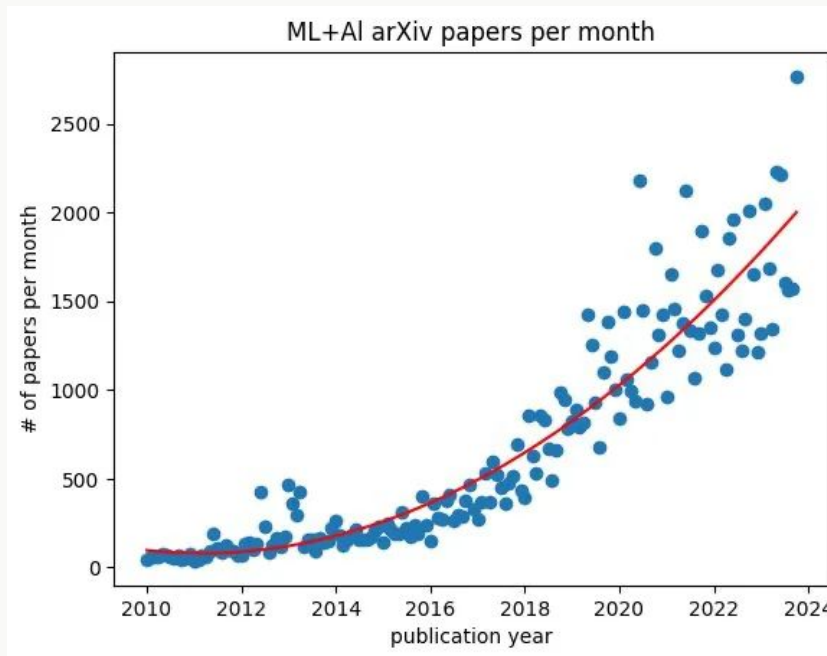
Probability of Models performing accurately

Penalized linear regression RidgeRegressor model (RV) Gradient boosting XGBoost model (XG)  
Random forest RandomForestRegressor model (RF) Neural network MLPPerceptronRegressor model (NN)



# Model Proliferation Problems

Challenges in managing multiple models



# Enter Apache Airflow



# Automating Data Preparation

Using Airflow for data quality checks and balancing



## Data Wrangling

- Amazon SageMaker Data Wrangler
- Databricks (Spark)
- Amazon EMR & Glue (Spark)

## Data Quality

- Great Expectations
- AWS Glue DQ (PyDeequ)
- Soda

# Automating Data Preparation

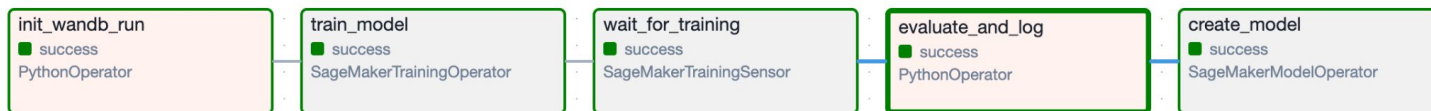
## Rule Sets

- Record Counts  
RowCount > 100
- Validness  
"A" in ['N', 'Y']
- Completeness (non-null)  
IsComplete "A"
- Standard Deviation  
StdDev "A" between 3 and 5

```
create_rule_set = GlueDataQualityOperator(  
    task_id="create_rule_set",  
    name=rule_set_name,  
    ruleset=RULE_SET,  
    data_quality_ruleset_kwargs={  
        "TargetTable": {  
            "TableName": athena_table,  
            "DatabaseName": athena_database,  
        }  
    },  
)  
  
start_evaluation_run = GlueDataQualityRuleSetEvaluationRunOperator(  
    task_id="start_evaluation_run",  
    datasource={  
        "GlueTable": {  
            "TableName": athena_table,  
            "DatabaseName": athena_database,  
        }  
    },  
    rule_set_names=[rule_set_name],  
)
```

# Integration with Experimentation Platform

Using Airflow for model versioning, deployment, and monitoring



## Model Training

- Amazon SageMaker
- Databricks
- Amazon Bedrock

## Model Cataloging

- Weights & Biases
- MLFlow (SageMaker & Databricks)
- Neptune

# Integration with Experimentation Platform

## SageMaker MLFlow

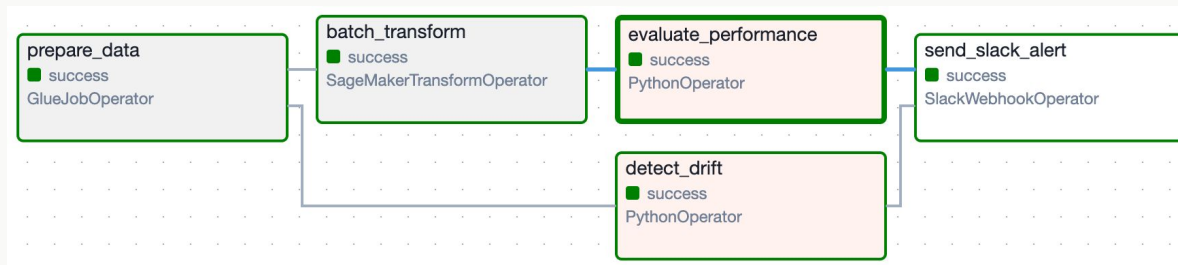
- Experiment Tracking
- Model Evaluation
- Model Registry

The screenshot displays the SageMaker Studio interface for a specific model version. The breadcrumb navigation at the top reads: SageMaker Studio > Models > Registered Models > Iris Random Forest Model 37705e > Versions > Version 10 > Overview. The main content area is titled "Version 10 (Model Version)" and includes tabs for Overview, Activity, and Details. Below the tabs are four action buttons: Train (Complete), Evaluate (Undefined), Audit (Draft), and Deploy (Pending Approval). A "Metrics" section is expanded, showing a table with columns for Name, Value, and Notes. The table contains four rows of metrics: accuracy, precision, recall, and f1\_score. The bottom of the metrics section shows "4 results", "Metrics per page 10", "Go to page 1", and "Page 1 of 1".

Name	Value	Notes
accuracy	0.9555555555555556	--
precision	0.9573302469135803	--
recall	0.9555555555555556	--
f1_score	0.9557368557368557	--

# Automated Evaluation and Monitoring

## Using Airflow for continuous monitoring



### Evaluation Frameworks

- Amazon SageMaker
- Evidently (Data drift Evaluation)
- Ragas (LLM RAG Evaluation)

### Notification Systems

- Slack Notification
- Amazon Simple Notification Service
- Email

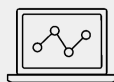
# Scaling AI Adoption with Airflow



Dag Factory

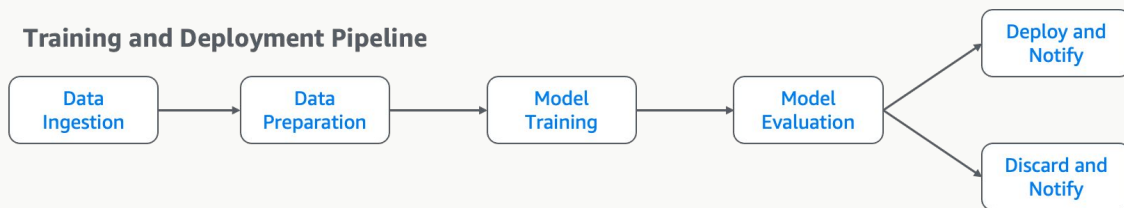


Dag Templating



Data-driven Scheduling

## Training and Deployment Pipeline



## Monitoring and Evaluation Pipeline



## Inference Pipeline





# Real-World Success Stories



# Airflow Success Story



## What do they do?

- Identity Verification - Financial, Healthcare, Gambling (including biometric solutions)

## Challenges

- People
- Tech
- Governance

## Solutions

- Airflow - Data prep & ETL
- SageMaker - model training
- Templates
- Datahub



# Airflow Success Story



## What do they do?

- Industry leader in GenAI for contact centers: GenerativeAgent, AutoTranscribe, and AutoSummary

## Challenges

- Workflow management: ML Lifecycle, Tech

## Solutions

- Airflow: Data ingestion & preparation; model training/tuning/evaluation; parallel workloads; scalable AI workload orchestration

For more details: [Airflow, Spark, and LLMs: Turbocharging MLOps at ASAPP](#) by Udit Saxena (Thursday @11am)



# Questions?

**Shubham Mehta**

LinkedIn: [linkedin.com/in/shubh-mehta](https://www.linkedin.com/in/shubh-mehta)

Email: [shubhx@amazon.com](mailto:shubhx@amazon.com)

**Rajesh Bishundeo**

LinkedIn: [linkedin.com/in/rajeshwar-bishundeo/](https://www.linkedin.com/in/rajeshwar-bishundeo/)

Email: [rbishund@amazon.com](mailto:rbishund@amazon.com)

