MLOps at Scale

Naveen Edapurath Vijayan

Head of Engineering AWS Central Economics

DISCLAIMER

- All views and opinions shared in this presentation are **my own personal perspectives.**
- They do not represent or reflect the views of **Amazon or AWS.**
- I am sharing this purely out of **my personal passion** for **MLOps and AI.**

Scalable MLOps for Enterprise AI

A model in a Jupyter notebook is an experiment.

A model in production, monitored and governed at scale, is a business solution.

The Challenge

Brittle pipelines

Manual deployments

Lack of monitoring

Bigger risks with LLMs

What is MLOps?

Data pipelines

Models

Automated deployment

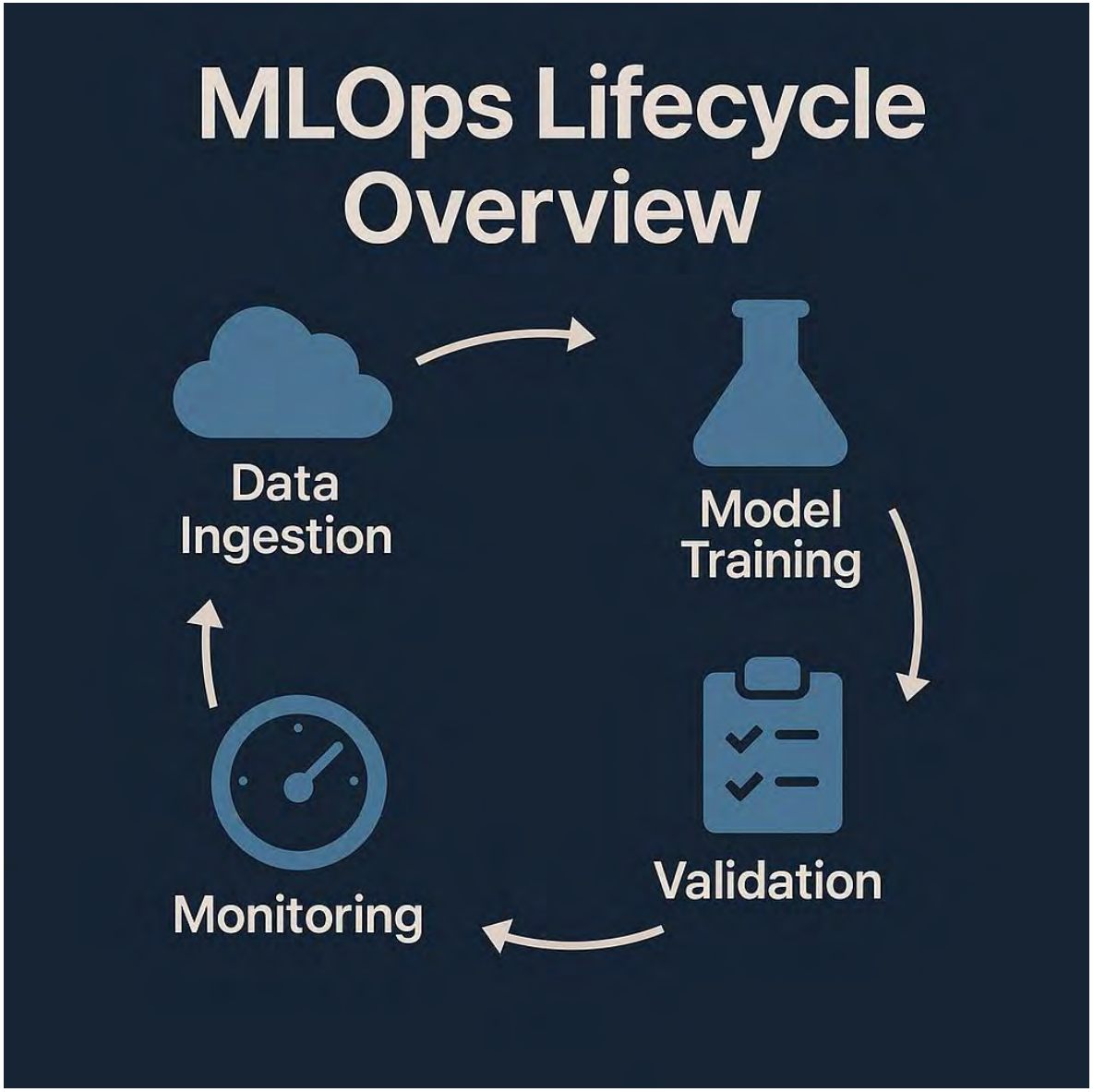
Monitoring

MLOps in the LLM Era

Manages prompt pipelines like feature pipelines

Ensures monitoring of hallucinations, toxicity, and grounding

Provides governance & compliance for high-stakes LLM use cases


MLOps Lifecycle Overview
Data Ingestion
Model Training
Validation
Monitoring

AUTOMATION

- Automates ML workflows
- Standardizes LLM processes
- Reduces manua intervention
- Keeps models up-to-date

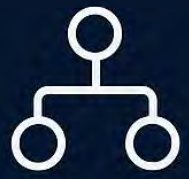
CI/CD for ML & LLMs

Version Control

Git for code,
DVC for datasets
MLflow for models

Continuous Evaluation

Tests for regression,
bias, and units



Deployment Strategies

Canary roliout,
shadow deployment

Governance & Responsible AI

Model Lineage

Git for code, DVC for datasets, MLflow

Data Provenance

Tests for regression, bias, and units

Audit Trails

Regulatory Compliance

Model & LLM Monitoring

Hallucination Rate

Toxicity

Audit Trails

Grounding

Best Practices in the LLM Era

Use infrastructure as code

Adopt feature stores & prompt stores

Automate bias testing, fairness checks & safety evaluations

Build cross-functional teams

CLOSING & CALL TO ACTION

MLOps is essential for classical ML and modern LLMs

Provides automation, CI/CD, monitoring and governance

→

Enables AI that is reliable, trustworthy, and scalable

Thank you.

Thank you.