NASSCOM®

K-tech CENTRE OF EXCELLENCE
Data Science and Artificial Intelligence
POWERED BY NASSCOM®

genpact

EY

MLOps

A KEY LEVER IN REVOLUTIONIZING AI/ML ADOPTION FOR INDUSTRIES

March 2022

Artificial Intelligence (AI) and Machine Learning (ML) are likely to usher in the greatest technological revolution of our lifetime. From organisations working in AI/ML having a mythical status to applications of AI/ML being present at every single channel, we have come a long way. Organisations are investing increasingly in ML as new use cases come to the forefront regularly. While immersing themselves in ML, organisations are also gradually understanding the importance of MLOps.

MLOps is all about how to effectively manage data scientists and operations resources to allow for the effective development, deployment, and monitoring of models. As Samir Tout, Professor of Cybersecurity at Eastern Michigan's School of Informed Security & Applied Computing puts it, "MLOps' real emphasis is on the consistent and smooth development of models and their scalability." It reduces the time to market and increases productivity; thus, helping in increasing revenues.

ML models are growing exponentially, but only 27% of the pilot projects are into production. The challenges are around operations engineering practices that cover reproducibility (similar results with same datasets), auditing the process compliance, validations of the model output, and monitoring for the integrated systems. This is a similar situation that the software world faced around 2007/2008 that led to the birth of DevOps, which combines best practices from IT operations and development.

This compendium intends to provide an MLOps adoption framework and methodology for enterprises that have started or are just starting their journey with Machine Learning. It is an attempt to encapsulate the best practices to guide organisations on effective set-up, management, and scaling of ML operations. We believe that this playbook will help the stakeholders gain a clear understanding of MLOps and embrace them in foreshadowing and mitigating the challenges that are associated with it at different stages. We also hope this playbook encourages countless others who are interested to embark on the MLOps journey.

We hope you find the playbook to be informative and useful and this acts as a guiding light in your journey of ML adoption.

Sangeeta Gupta

Senior Vice President
at NASSCOM

Table of Contents

1

Introduction

- Introduction
- What is MLOps?
- Identified Benefits of MLOps
- MLOps Adoption by Function
- Practices and Maturity of MLOps
- Dimensions of MLOps

2

MLOps Implementation Pillars

- MLOps Implementation Framework
- MLOps Implementation RACI Matrix
- Stage 1: Define and Design
- Stage 2: Data Pre-Processing
- Stage 3: Model Development
- Stage 4: Model Deployment and Monitoring
- Model Management
- Model Testing
- Core Technologies and Tools for MLOps

3

Industry Use Cases

- Industry Use Cases

4

MLOps Business and Operation Pillars – Investment and Change Management

- Planning Transformation
- Where to Invest?
- MLOps Platform Build Vs Buy?
- Change Management
- Evolution of MLOps Stakeholders

5

Control & Governance

- Data Governance
- Process Governance
- Current Regulations Driving MLOps Governance

6

Innovation and Future

- The Future of MLOps

7

Appendix

- Appendix I
- Appendix II
- Appendix III

01

Introduction

05 Introduction

06 What is MLOps?

07 Identified Benefits of MLOps

08 MLOps Adoption by Function

09 Practices and Maturity of MLOps

10 Dimensions of MLOps

Introduction

Background

Business applications that are driven by Artificial Intelligence (AI) and Machine Learning (ML) models support faster and more intelligent decision-making. However, from our experience working with enterprises across multiple industries, we have witnessed that roughly only half of all AI proof of concepts are ever scaled to production.

Organisations have realised that the data changes over time and can have different characteristics depending on the region. As a result, they are looking for ML models that can be developed and deployed automatically, with strict governance and monitoring. Adoption of Machine Learning operations is assured if firms want to continue with automation.

MLOps refers to DevOps as applied to Machine Learning and Artificial Intelligence. Short for “software development” and “IT operations,” DevOps is the application of software engineering practices to IT operations, such as packaging and deploying production software.

On the other hand, MLOps aims to shorten the analytics development life cycle and increase model stability by automating repeatable steps in the workflows of software practitioners (including data engineers and data scientists). While MLOps practices vary significantly, they typically involve automating integration (the frequent checking-in and testing of code) and deployment (packaging code and using it in a production setting).

Why MLOps?

1. Industrialization of ML models and drive to a DevOps culture for AI/ML implementations
2. Key challenges around data quality, model decay, and data locality are driving towards MLOps adoption
3. Need for transparent governance and compliance
4. Focus has shifted from building just the core data science capability to moving the model to production with an intent to maximize return on its investment and drive business value
5. The rise of auto-ML tools and platforms has democratized efforts of data mining and decision sciences with the constant growth of citizen developers

> “
>
> ***50% of AI experts use standard development tools and frameworks to create AI models.***
> *– Global survey: The state of AI in 2020 | McKinsey*

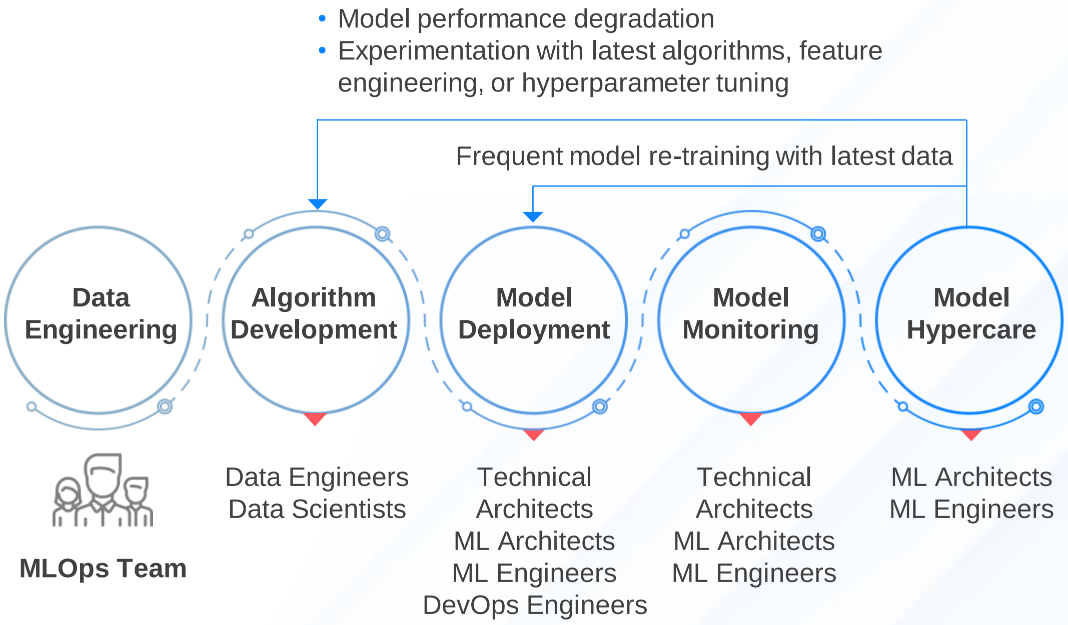
> “
>
> ***42% of AI leaders require to deploy a new model in 1 to 30 days.***
> *– Algorithmia_2020_State_ Of Enterprise ML*

> “
>
> ***73% of Business Leaders believe that MLOps adoption will keep them competitive.***
> *– Forrester: 7 Key Requirements For Successful MLOps Deployment*

What is MLOps?

Machine Learning Operations (MLOps) is a set of practices and methodology that helps organisations achieve automated and reliable ML model deployment, consistent model training, model monitoring, rapid experimentation, reproducible models, and accelerated models. Its key features are listed below:

Machine Learning Operations (MLOps)



Source: Genpact

Model Lifecycle Management

Similar to DevOps tools that need application development processes, MLOps tools are also required for the process management of model development, model deployment, model training, and mode of operation.

Model Versioning & Variation

MLOps have changed to provide the means that use various versions of the models, alerts, and notifications to inform users about the changes in model versions and support multiple versions in operations as the need arises.

Model Monitoring & Management

In recent years, the need for MLOps has been more relevant than ever. MLOps platforms are necessary to monitor the consumption and results of models to assess their accuracy, performance, and every other measure to ensure acceptable results.

Model Governance

MLOps have evolved to provide mediums for auditing, compliance with regulations and programs, governance and access control of models.

Identified Benefits of MLOps

Reduced time to market of ML-driven products

With automated model training and retraining processes and continuous integration and continuous delivery practices for deploying and updating Machine Learning pipelines, MLOps helps ML-based solutions get into production faster

Improvement in ROI on AI/ML & Analytics Initiatives

Optimized and lower investments: Lower Capex and Opex investment required due to managed nature of infrastructure

Advanced Data Management for ML-ready systems

MLOps practices and framework enable data engineers to design and build automated data pipelines, data ops platforms, and automated data feedback loops for model improvement, thereby solving >50 problems concerning the absence of clean, regulated, governed, and monitored data to build production-grade models

Improved interoperability & transparency of ML Models

Helps to enhance the explainability of the MLOps solution, making it white-box in nature

Increased speed of innovation with ML-driven products

Automated data and model management that can handle rapidly changing data privacy and compliance regulations across a vast portfolio of organisational units, thereby accelerating iteration and time to production

Enhanced quality & accuracy of predictions

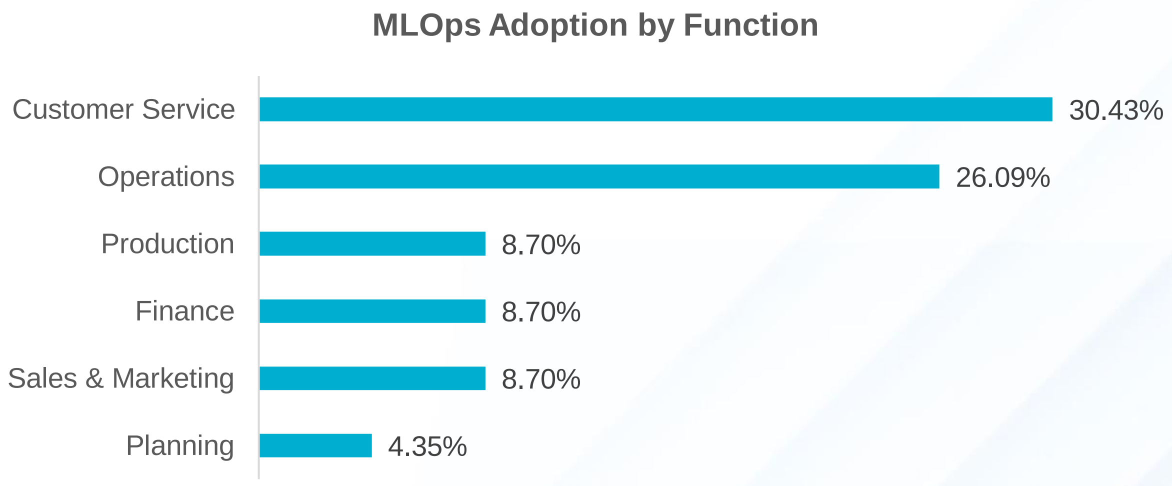
MLOps enables data and model validation, performance monitoring in production, and retraining against fresh datasets. These ensure that results produced by the algorithm when making important decisions are trustworthy and accurate

Helps in scaling up the ML solutions

Allows easy and limitless on-demand scalability to address seasonality and growth

MLOps Adoption by Function

According to the results of the focused group survey conducted for compilation of this compendium, MLOps adoption varies across different industry functions. Listed below are the percentage adoptions in different business functions.



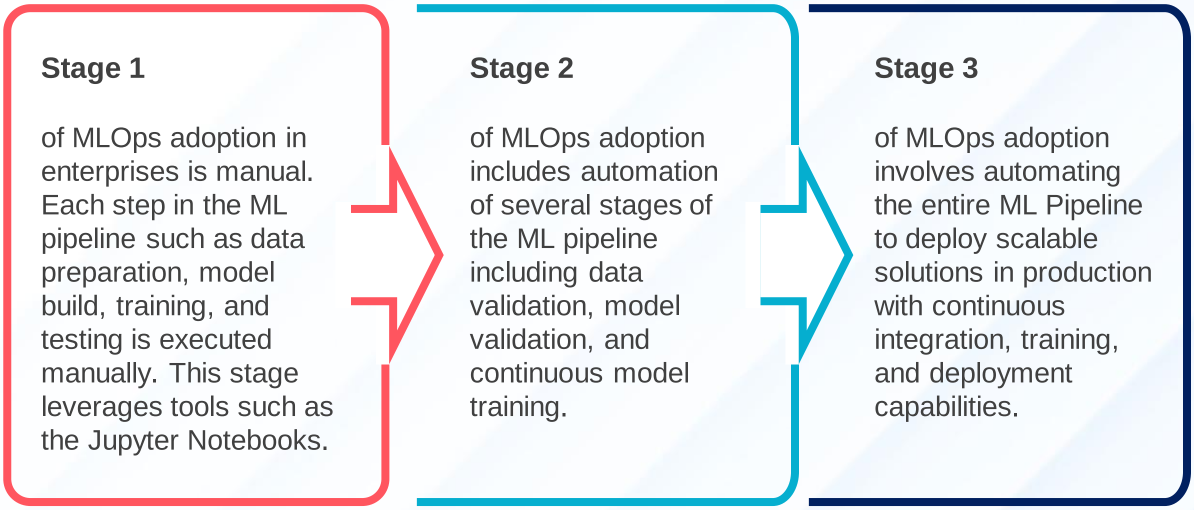
Source: NASSCOM and EY

Practices and Maturity of MLOps

MLOps Principles

1. **Automated CI** – Continuous Integration (CI) enables testing and validating code, components, data, data schemas, and models
2. **Automated CD** – Continuous Delivery (CD) enables an ML training pipeline that automatically deploys a model prediction service
3. **Automated CT and Model Performance Monitoring** – Enables automatically retraining and serving the ML models through ML pipelines, proactively triggering ML model re-runs on new training data and preventing staling of ML models due to training — serving skew challenge
4. **Automated Model and Data Versioning** – Enables faster reproducibility through consistent version tracking of training data, ML models metadata, and ML model features and codes
5. **Automated Model & Data Validation** – Enables comparing the performance metrics of a new ML model with previous ML models or threshold values and of new training data frequency distributions with previous training data's
6. **Model Verifiability, Traceability, and Reproducibility** – Enables tracing a model in production back to its roots with model repo, versioning, and metadata management
7. **Collaboration and Cross-Team Alignment** – Ensures standardized and repeatable architectural patterns, project templates, and machine infrastructure and configuration-as-code for seamless collaboration between teams

MLOps Maturity



Stage 1

of MLOps adoption in enterprises is manual. Each step in the ML pipeline such as data preparation, model build, training, and testing is executed manually. This stage leverages tools such as the Jupyter Notebooks.

Stage 2

of MLOps adoption includes automation of several stages of the ML pipeline including data validation, model validation, and continuous model training.

Stage 3

of MLOps adoption involves automating the entire ML Pipeline to deploy scalable solutions in production with continuous integration, training, and deployment capabilities.

Dimensions of MLOps

The survey results reveal that organisations must focus their efforts on six key dimensions to form a holistic MLOps framework from both, Implementation and Business & Operations standpoints.
The Implementation pillar gives details on how organisations can build and adopt MLOps and ensure smoother implementation while the Industrialisation pillar informs how the businesses need to manage and control the MLOps adoption in their organisations. Both are equally critical areas for MLOps adoption for the organisations.

Business & Operations Pillars

06 Innovation & Future

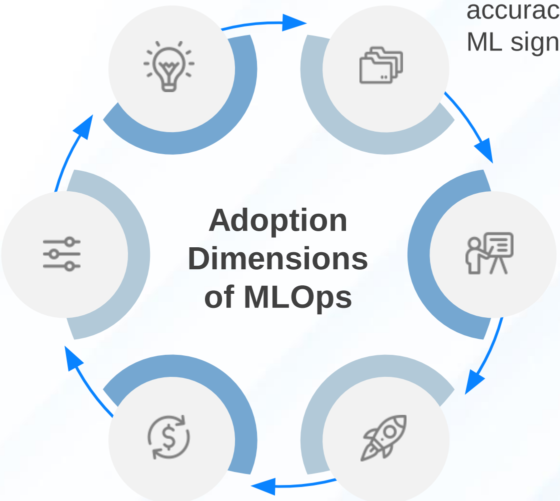
The future scope of Machine Learning is to make the most out of the advancing technologies and platforms such that it makes human life easy as well as assist in making profitable business decisions.

05 Control & Governance

Government regulations and laws that help to regulate and operate Machine Learning projects from global and India's point-of-view are important factors.

04 Investment & Change Management

Investments from technology, infrastructure, and business standpoints are required for change management to successfully implement MLOps.



Implementation Pillars

01 Data

Data is the fuel for many digital transformations; and hence most organisations, use it to gain better business insights. The quality of data used to build predictive models heavily influences the model's accuracy making the role of data in ML significant.

02 Training Model

An ML algorithm is trained by feeding datasets. This is the stage where the learning takes place. Consistent training can significantly improve the prediction rate of the ML model.

03 Deployment

Deployment is the method by which you integrate an ML model into an existing production environment to make practical business decisions based on data. The model deployment phase is the most riddled with challenges and relevant expertise is required to address the issues.

Source: NASSCOM, Genpact, and EY

02

MLOps Implementation Pillars

12 MLOps Implementation Framework

13 MLOps Implementation RACI Matrix

14 Stage 1: Define and Design

15 Stage 2: Data Pre-Processing

17 Stage 3: Model Development

19 Stage 4: Model Deployment and Monitoring

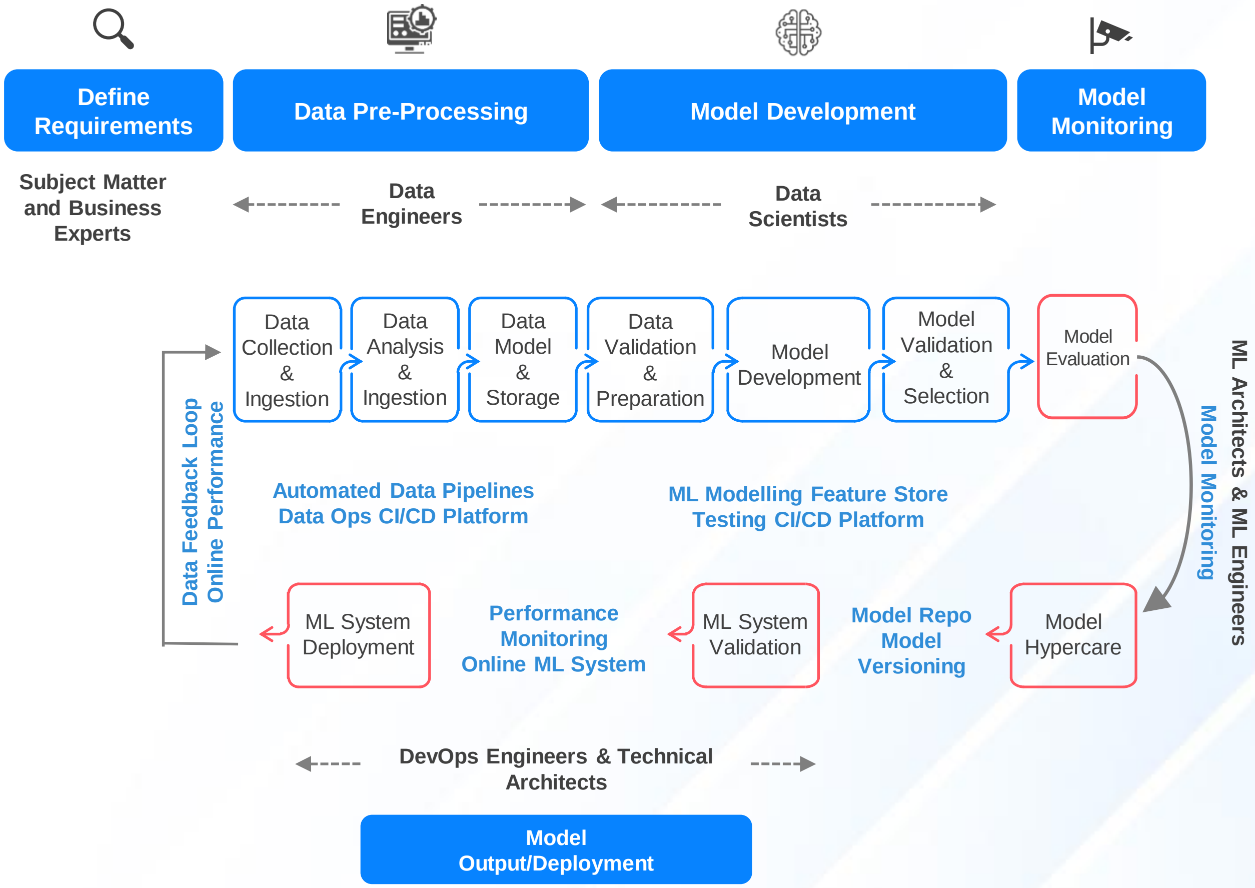
22 Model Management

25 Model Testing

27 Core Technologies and Tools for MLOps

MLOps Implementation Framework

A Machine Learning deployment practice with automated ML platforms, data pipelines, continuous model monitoring, automated deployment, and reorganized team structures is depicted below:



Define Requirements

Data Pre-Processing

Model Development

Model Monitoring

Subject Matter and Business Experts

Data Engineers

Data Scientists

Data Collection & Ingestion

Data Analysis & Ingestion

Data Model & Storage

Data Validation & Preparation

Model Development

Model Validation & Selection

Model Evaluation

Automated Data Pipelines
Data Ops CI/CD Platform

ML Modelling Feature Store
Testing CI/CD Platform

ML Architects & ML Engineers

Model Monitoring

Data Feedback Loop
Online Performance

ML System Deployment

Performance Monitoring
Online ML System

ML System Validation

Model Repo
Model Versioning

Model Hypercare

DevOps Engineers & Technical Architects

Model Output/Deployment

Key Focus Areas

1. Need for New Team Structure
2. End-to-End ML Platforms for Build, Deploy, & Monitor
3. Continuous Model Monitoring & Automated Deployment
4. IT Team for Infrastructure and Asset Provisioning
5. Responsible AI Practices at all stages

Source: Genpact

MLOps Implementation RACI Matrix

Task	Subtask	Stakeholder					
		Business Owner/ Leaders	Subject Matter Experts	Data Engineers	Data Scientists	ML Architects & ML Engineers	DevOps Engineers & Technical Architects
Business Understanding	Define Business Requirements	R	A	I	I	I	I
Data Pre-Processing	Data Collection & Ingestion	I	C	R A	C	I	I
	Data Analysis & Ingestion	I	C	R A	C	I	I
	Data Model & Storage	I	C	R A	C	I	I
Model Development	Data Validation & Preparation		I	C	R A	I	I
	Model Development		I	C	R A	I	I
	Model Validation & Selection		I	C	R A	I	I
Model Monitoring	Model Evaluation & Model Hypercare		I	I	C	R A	C
Model Output/ Deployment	ML System Validation & ML System Deployment		I	I	C	C	R A

R Responsible   A Accountable   C Consulted   I Informed

Source: Genpact

Stage 1: Define and Design

The development of a Machine Learning model often begins with a business objective, which can be as simple as minimizing fraudulent transactions to less than 0.1 percent or being able to recognise peoples' faces in a social network's photographs. Performance targets, technical infrastructure needs, and financial limits all come with business objectives; all of these aspects can be represented as key performance indicators, or KPIs, which allow the business performance of the ML models in production to be monitored. The part of setting the objectives includes change management, which may even provide some recommendations for how the organisation should proceed.

Key Activities in this stage

1 **Define Project Charter**

2 **Understand Business Problem & Industry Context**

3 **Design Solution Architecture**

4 **Design Data Pipeline**

Challenges in this stage

- **Unclear metrics of the project success**
- **Underestimating project costs**
- **Not examining if the ML problem is worth solving at all**
- **Overlooking ethical aspect of AI**
- **Understanding the information churned out of the model**
- **Not living up to stakeholder expectations**
- **Maximizing accuracy without trade-off context**
- **Improving the model endlessly**

Stage 2: Data Pre-Processing

The fundamental part of any Machine Learning workflow is data. An ML model is only as good as the data. Therefore, the data that has been used for training the ML model indirectly influences the overall performance of the production system.

Key Activities in this stage

Program or project leader collaborates with client and secures IT support for pipeline provisioning

Collect Data Sources

Configure Data Ingestion

Perform Data Transformations

Create Data Storage

Challenges in this stage

Sub-Stage Name	Challenges
Data Relevance	Biased data
	Seasonal effect in data
	Data changes over time
	Relevance of the train/test subsets
Data Quality	Outliers
	Duplicates
	Missing values
	Noise in data
Data Scale	Amounts of data
	Randomized data subsets
	Data representation
	Data matching
Data Preparation Procedure	Including new data into processes
	Automated ETL/ELT

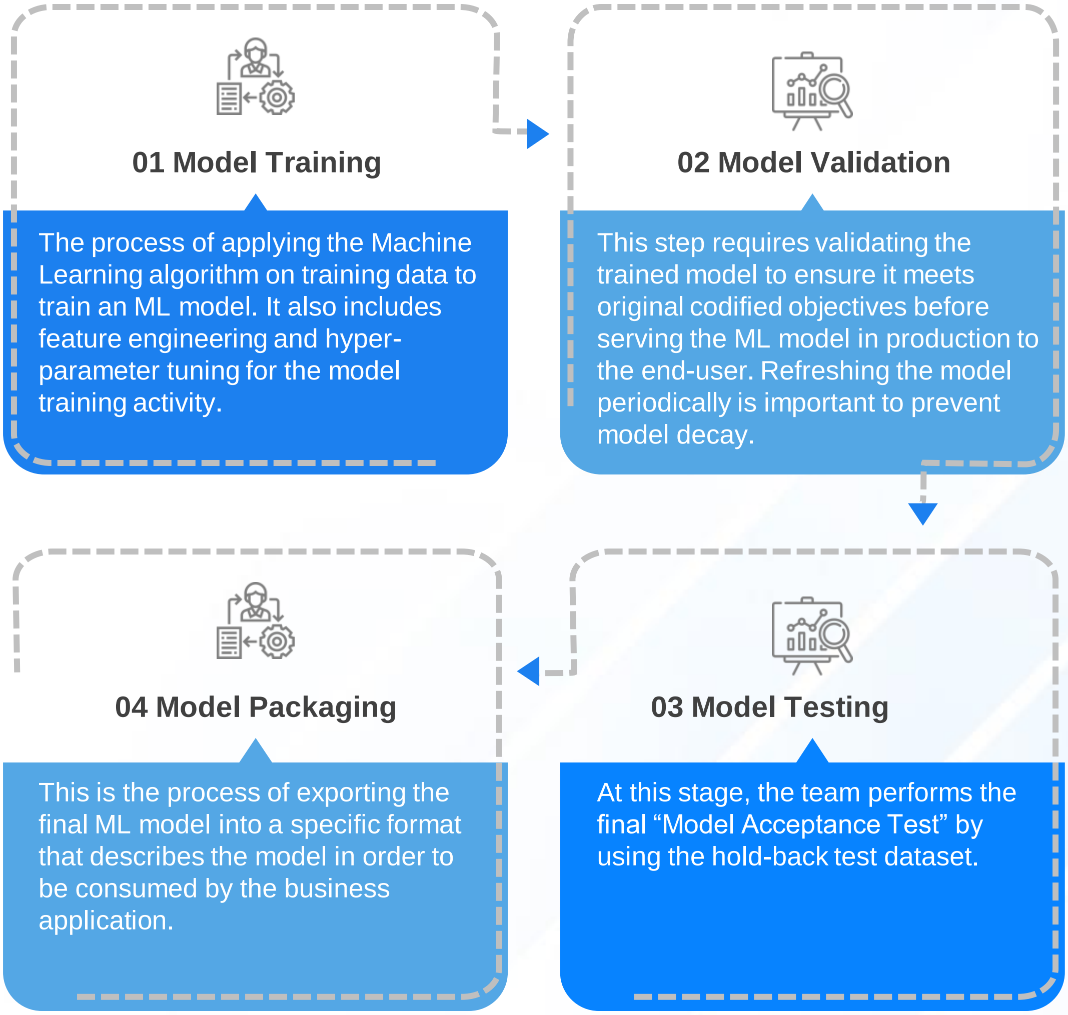
Stage 2: Data Pre-Processing

Best Practices in this stage

Activity	Description	Best Practices
Data Validation	There should be an automatic check for expected data and features	Statistics from the training data should be calculated, i.e., expected value and standard deviation to be used as an expected definition for input data during the training or prediction process
Feature Importance Test	This is a test to understand the predictive power of the feature	• Computation of correlation coefficient on features columns • Further training the model with one or more features • Measure the data dependencies, inference latency, and RAM usage for each new feature and compare it with the predictive power of the newly added features • Drop out unused/deprecated features from your infrastructure and document them
Feature and Data Policy Compliance	The features and data should be compliant with the General Data Protection Regulation (GDPR) and other regulatory bodies	These requirements should be programmatically checked in both development and production environments and documented
Feature Creation Code	The code developed to create new features should be tested	Unit tests should be written to thoroughly test and catch the bugs in features

Stage 3: Model Development

Key Activities in this stage



01 Model Training

The process of applying the Machine Learning algorithm on training data to train an ML model. It also includes feature engineering and hyper-parameter tuning for the model training activity.

02 Model Validation

This step requires validating the trained model to ensure it meets original codified objectives before serving the ML model in production to the end-user. Refreshing the model periodically is important to prevent model decay.

03 Model Testing

At this stage, the team performs the final "Model Acceptance Test" by using the hold-back test dataset.

04 Model Packaging

This is the process of exporting the final ML model into a specific format that describes the model in order to be consumed by the business application.

Source: EY

Challenges in this stage

Skillset - A wide range of skillsets at play from engineering, cloud, to Machine Learning form the base of an ML engineer.

Data Availability and Quality - Data availability has been a big bottleneck due to a lack of centralised data management practices; poor data quality affects the trust of stakeholders, which negatively impacts the ability to make decisions based on the data.

Evolving Business Process - With evolving versions, it is required to ensure that the features are evolving and are duly accommodated.

Stage 3: Model Development

Best Practices in this stage

Stage: Model Development and Model Validation

Test	Description	Action to be taken
Right Algorithm Metrics for Business Objective	This means that loss metrics from Machine Learning algorithms (Hinge Loss, MSE, log-loss, etc.) should be correlated with the business impact indicators of that particular engagement (cost management, revenue, client retention, etc.).	The loss metrics - impact metrics relationship can be measured in small-scale A/B testing using an intentionally degraded model. This process can be done in the experimentation step.
Model Staleness Test	The model is defined as stale if the trained model does not include up-to-date data and/or does not satisfy the business impact requirements. Fixing a stale model is always of utmost importance, especially in the level 2 automation level where there is a constant delivery of prediction, and wrong predictions might be costly to the project.	A/B experiments with older models including the range of ages to produce an Age vs. Prediction Quality curve to facilitate the understanding of how often the ML model should be trained. Another way to prevent model staleness is to perform error analysis or tune thresholds in the model.
Cost of More Sophisticated ML Models	This is the cost and benefit analysis of the utilization of more sophisticated ML models being used in production.	When in production, at times, state-of-the-art models are used to increase performance but these models' performances should be compared to the simple baseline ML model (e.g., linear model vs neural network) to see if the simpler models perform better.
Validating Performance of a Model	It is recommended to separate the teams and procedures collecting the training and test data to remove the dependencies and avoid false methodology propagating from the training set to the test set (source).	Use an additional test set, which is disjoint from the training and validation sets. Use this test set only for a final evaluation.
Tests for Fairness/Bias/ Inclusion in the ML Model Performance	This is the process of testing the model to show its predictions don't have any unfair biases implicit in the model.	• Examine input features if they correlate with protected user categories • Collect more data that includes potentially under-represented categories

Stage 4: Model Deployment and Monitoring

Model Deployment

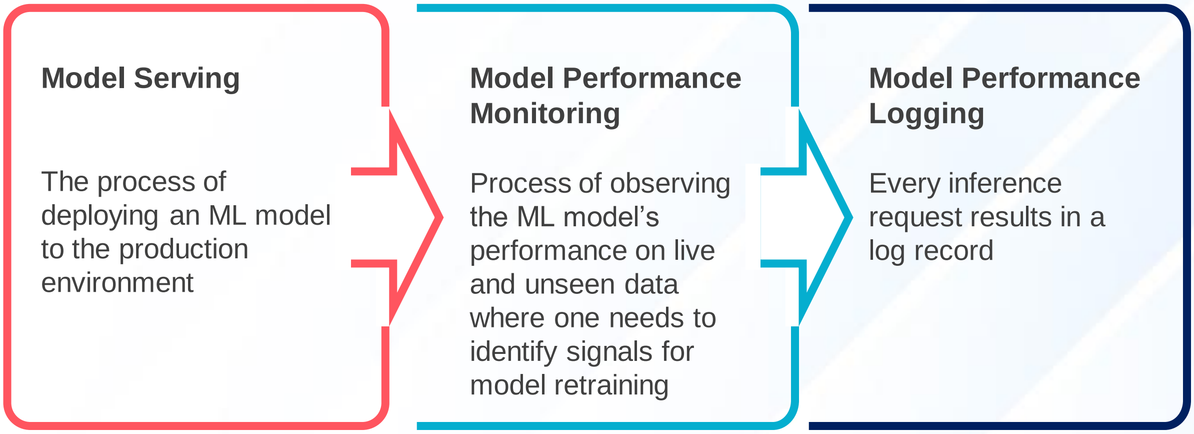
Productionalizing and deploying models are key components of MLOps that present an entirely different set of technical challenges compared to when developing the model. It is the domain of the software engineer and the DevOps team and the organisational challenges in managing the information exchange between the data scientists and these teams must not be underestimated.

Without effective collaboration between teams, delays, or failures to deploy are inevitable. Deploying an ML system to the production environment includes these two steps:

1. Deploying the pipeline for automated retraining and ML model deployment
2. Providing API for prediction on unseen data

Model Deployment Pipelines

The final stage of delivering an ML project includes the following three steps:



Model Serving

The process of deploying an ML model to the production environment

Model Performance Monitoring

Process of observing the ML model's performance on live and unseen data where one needs to identify signals for model retraining

Model Performance Logging

Every inference request results in a log record

Source: EY

Model serving is a way to integrate the ML model into a software system. We distinguish between five patterns to put the ML model in production: **Model-as-Service, Model-as-Dependency, Precompute, Model-on-Demand, and Hybrid-Serving.**

Stage 4: Model Deployment and Monitoring

Challenges in this stage

Stage Name	Challenges	Detailed Description
Operationalization Phase	Programming Languages Conflict	Joining ML models with production components may result is reduced performance of the model because they are written in different languages.
	Long Response Time	The time needed to generate an inference rises with the number of requests to the model.
	Quality Issues	Quality of data is most important in the post-deployment phase as the cost of mistakes will be too high.
	High Cloud Costs	High costs of cloud service is a major investment for organisations.
	Testing the ML Pipeline	Testing individual components and ignoring to test the ML pipeline can result in the system components not working together.
	Model Poisoning	Bad data may seem acceptable, but it gradually reduces the model's accuracy.
	Inconsistent Metric Results	At times, metrics pipelines have been completed successfully but the model's performance is insufficient.

Stage 4: Model Deployment and Monitoring

Best Practices in this stage

Test	Description	Action to be taken
Reproducibility on the Same Model	Differential testing, which relies on deterministic training, is performed on the MI model. This is hard to achieve due to the non-convexity of the ML algorithms, random seed generation, or distributed ML model training.	Reduce non-determinism by using the same random seed number when rerunning ML applications, etc.
Testing ML API Usage by Stress Testing	Testing the limits of the ML API to ensure it doesn't break down when deployed.	• Perform unit tests to randomly generate input data and train the model for a single optimization step (e.g., gradient descent). • Perform crash tests for model training. The ML model should restore from a checkpoint after a mid-training crash.
Test the Algorithmic Correctness	This is the process of testing if the selected algorithm is working as it should be.	Perform the Unit Test not intended to complete the ML model training but to train for a few iterations and ensure that loss decreases while training. Note: Avoid differential testing with previously built ML models because such tests are hard to maintain.
Integration Testing	This is the process where the steps in the ML pipeline are tested as a whole to determine all the steps and process works as expected.	Create a fully automated test that regularly triggers the entire ML pipeline. The test should validate that the data and code successfully finish each stage of training and the resulting ML model performs as expected. Note: Make sure all integration tests are run before the ML model reaches the production environment.
Validating the ML Model Before Serving It	This is the process of vetting that the model works correctly by training a certain percentage of the data.	• Setting a threshold and testing for slow degradation in model quality over many versions on a validation set. • Setting a threshold and testing for sudden performance drops in a new version of the ML model.
ML Models are Canaried Before Serving	Canary Testing is a way to reduce risk and validate your model by testing it on real-life data to see if the predictions are correct.	Testing that an ML model successfully loads into the production server and the prediction on real-life data is generated as expected.

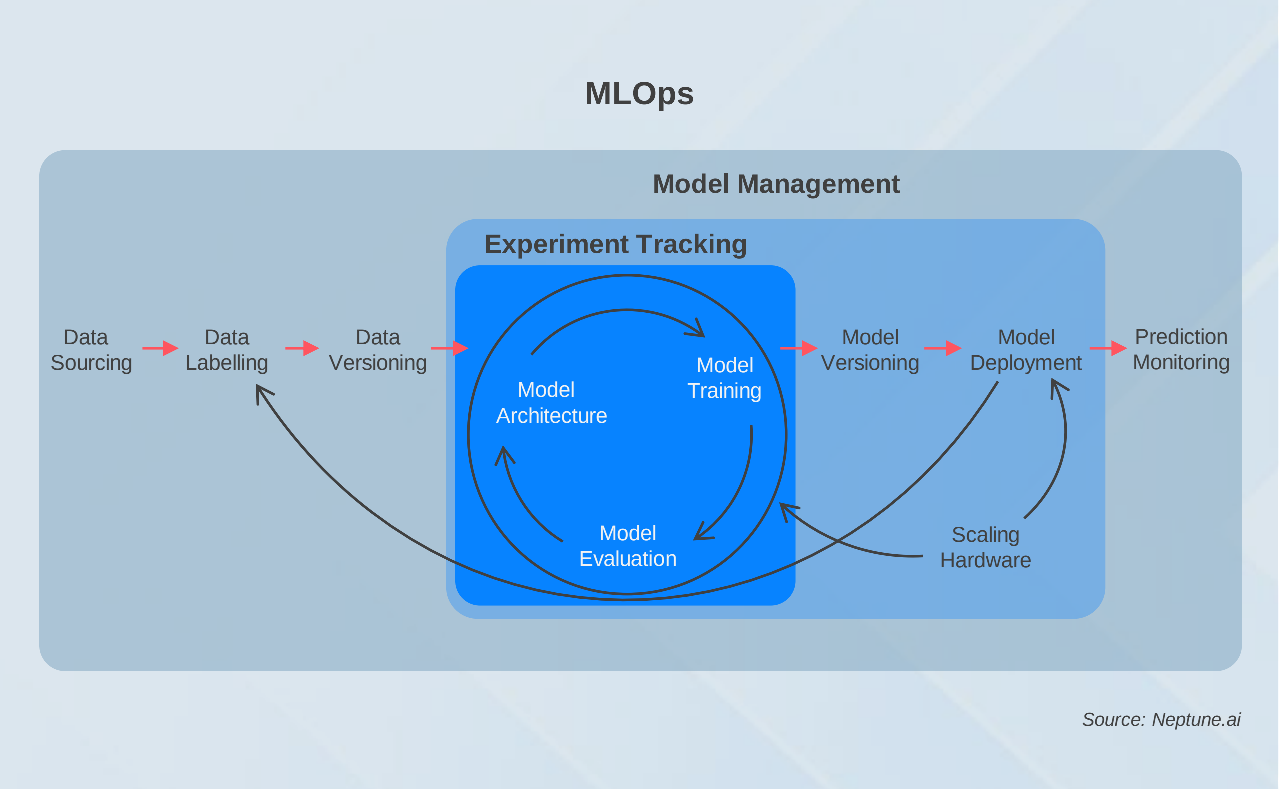
Model Management

What is ML Model Management?

ML Model Management is a critical component of the model lifecycle. ML model management ensures ML models are consistent and reach all business goals at scale. ML model management encompasses a total of model training, model maintenance, model deployment, model monitoring, and organisation and documentation of ML models. If incorrect model management is used, it will result in a significant performance reduction and non-utility. Thus, the main aim of ML Model Management is to store and monitor the model's performance.

Some questions that ML Model Management answers are:

- How good is the model performance today versus its performance on the previous day?
- What are the features that the model is trained on?
- What are the current hyperparameters? Do they change over time?
- Which model is in production/integration?
- How to bring about a major model change into this setup?



Source: Neptune.ai

Model Management

Steps of ML Model Management

ML Model Management is broadly divided into two phases, the Experimentation Phase — track training parameters, metrics, collaborations, and data and model versioning, and the Deployment Phase — packaging, deployment, monitoring & governance, and re-training.

1. **Model Architecture:** Identify the ideal model that will help solve the business case.
2. **Model Training:** Train the model with the train data and check for consistency in results. If the accuracy has been reached for the use-case scenario, then the model is ideal for moving to the stage of model evaluation.
3. **Model Evaluation:** Test the model with the test data, keeping in mind that the train and test data need to be different so that the model finalized is robust in its performance.
4. **Model Versioning:** Monitor the performance of multiple models and showcase their results to utilize in the model deployment phase as ML is an iterative process that needs to be monitored in detail.
5. **Model Deployment:** Deploy the model chosen, based on accuracy and success in mitigating business-challenge, in the client environment.

Challenges of ML Model Management

These are conceptual challenges in ML Model Management:

1. **Decisions on Model Retraining:** New events/incidents in a business environment force data scientists to retrain the model or in certain instances, remodel their ML model (for example, remodelling retail store demand when a lockdown is declared in the city). Thus, it is a challenge to predict the number of times that a model will need to be re-trained.
2. **Adversarial Settings:** Adversaries try to reverse engineer the ML model and find the limitations of the model. After feeding different data, they can game the model and make it practically ineffective.

Best Practices

1. MLOps constantly monitors and retrains the model to ensure it performs optimally in every business environment.
2. Consistent monitoring of any possible data drift or concept drift will prevent all forms of adversarial attacks.

Model Management

Version Management

The survey highlighted the significance of Version Management as a key practice for Model Management. Below are some factors that necessitate the versioning:

1. Data may reside across multiple systems in restricted jurisdictions
2. Data storage may not be immutable
3. Data ownership may be a factor
4. Sometimes, models need to be quickly rolled back to the previous serving version
5. Corporate or government compliance may require audit or investigation on both ML models and/or data, hence access to all versions of the productionized ML models is required

Best Practices Of Version Management

There are three aspects of best practices:

1. **Commit & Version Control**
   - Have all commits be atomic, complete, consistent, traceable, and with a single intent
   - Make changes visible through frequent commits
   - Consider how you would use the comments in the future
   - Review code before committing to the mainline
   - Make commits reversible

2. **Code Branching**
   - Optimize productivity
   - Enable parallel development
   - Allow for a set of planned, structured releases
   - Provide a clear promotion path for software changes through production
   - Evolve to accommodate changes that are delivered, perhaps daily
   - Support multiple versions of released software and patches

3. **Data & Model Security**
   - Data encryption at rest and in transit
   - Authentication and authorisation
   - Partitioning access control according to the intent of a change
   - Audit trails
   - Threat detection for unexpected incidences

Model Testing

1. The complete development pipeline includes three essential components, data pipeline, ML model pipeline, and application pipeline. In accordance with this separation, we distinguish three scopes for testing in ML systems: **tests for features and data, tests for model development, and tests for ML infrastructure.**
2. **Features and Data Tests**
   - Data validation comprises of automatic check for data and features schema/domain
   - A crucial test to understand whether new features add predictive power
   - Features and data pipelines should be policy-compliant (e.g., GDPR)
   - Feature creation code should be tested by unit tests (to capture bugs in features)
3. **Tests for Reliable Model Development**
   - Testing ML training should include routines that verify the algorithms make decisions aligned to business objectives or not
   - Model staleness test
   - Assessing the cost of more sophisticated ML models
   - Validating performance of a model
   - Fairness/Bias/Inclusion testing for the ML model performance
   - Conventional unit testing for any feature creation, ML model specification code (training), and testing
   - Model governance testing

Model Testing

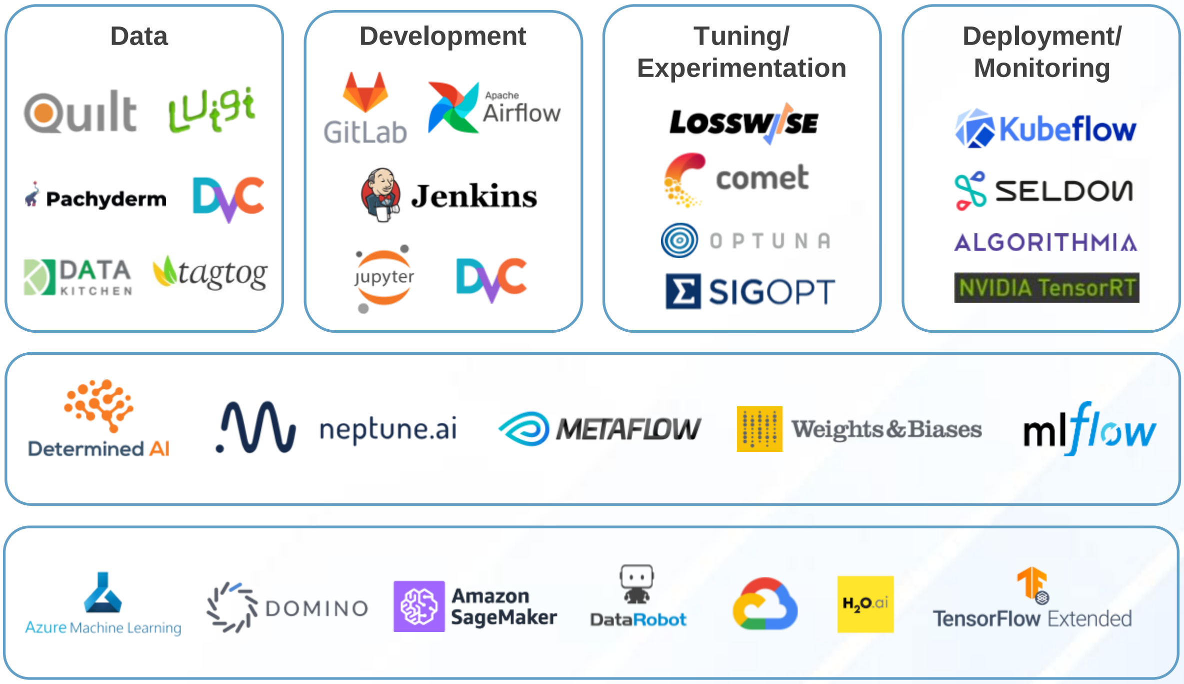
Challenges

1. A sync between IT, business requirements, and implementation teams can be a challenge.
2. Selection between a cloud solution and open-source solutions can be a bottleneck.
3. Streamlining data gathering and engineering to feed into downstream model development can be arduous. The challenges in MLOps vary with each use case. Common issues are — a data pipeline concern if the data is changing drastically at the source; a data security problem arising out of data localization and other customer data access choices; a complication due to multiple monitoring scenarios, etc.
4. Unavailability of skilled teams to handle the complexity of model building and model deployment.

Best Practices

1. Understanding the requirements of the project is key and can drastically change the MLOps effort and timelines.
2. Before beginning the design of the solution and architecture, an agreement on the tool and techniques to be used amongst the client and team is paramount.
3. Defining a clear MLOps framework and steps to be followed is necessary.
4. Opting for microservices-based development works well where multiple components are involved or when the model has to interact with various sections of the IT solution.
5. Data and Model Versioning is a must for governance.

Core Technologies and Tools for MLOps


Data

Development

Tuning/ Experimentation

Deployment/ Monitoring

Source: NASSCOM, EY, and Genpact

03

Industry Use Cases

29 Industry Use Cases

Industry Use Cases

Global Healthcare Equipment Manufacturer

Problem Statement

To aid medical practitioners in detecting anomalies in X-Ray reports thus helping in early-stage detection of problems.

Solution

1. Gather multiple X-ray images and train the model to differentiate between an anomaly and a normal condition.
2. Train the model to remove any possible glitches in the classification of an anomaly.
3. Deploy the model to categorize the patient's medical condition as normal or an anomaly.

Business Impact

Reduce the turnaround time for generating X-ray reports. Thereby provide better care to the patients.

American Scientific Instrumentation Supplier

Problem Statement

Enable integrated visibility to order management & customer service information for CSR teams.

Solution

1. Extract the email and clean up the content for further processing.
2. Identify the various features of the email like the subject and the body of the email.
3. Categorize the email with deep learning classification tools like XgBoost, SVM, or LSTM.

Business Impact

Efficient routing of mails to appropriate teams for process automation and accelerating revenue growth.

Global Healthcare Organisation

Problem Statement

Predict invoices likely to be paid later than the entitled date by dynamically analysing the payment behaviours of the customers.

Solution

1. Gather raw data containing 3.6 million+ invoice lines.
2. Predict the probability of invoices likely to be paid late with 87% accuracy by using an ensemble of ML algorithms.
3. Develop agile customer segmentation based on the variable customer payment behaviour by using clustering techniques.

Business Impact

Achieved a reduction in past due-invoices from 20-25% to less than 12%.

Industry Use Cases

Automobile

Problem Statement

To introduce AI-based augmentation to the collections and cash application processes within the organisation to increase accelerated collections, improve days outstanding, and customer satisfaction.

Solution

A combination of NLP, IOCR, and ML to process collections communication, follow-up activities, and cash application to move from traditional reactive collections to a more engaged and customer-partnering oriented collections process.

Business Impact

1. Improved visibility and accountability into resolving collections delays and disputes
2. As an owner and driver of data-based operations, enabling finance to drive the cash flow as an organisation-wide agenda
3. Moving from static and reactive response to dynamic and proactive response and use of insights in hindsight

Automobile

Problem Statement

To create a radio-analytics platform that consumes unstructured radio chatter from the race-track.

Solution

Automatic transcription to text and applying AI techniques to identify relevant intelligence that is gathered to augment the race strategy in real-time.

Challenges

To conceptualize and realise the platform given the fast-paced nature of the sport leading to unique and novel algorithms custom-designed for the purpose on top of a complete open-source toolset.

Business Impact

The platform delivered immediate value with multiple podium finishes for the client.

Industry Use Cases

Food & Beverage

Problem Statement

Evolution of a supply chain inventory for an F&B giant to cater to the problem of manual deployments at an on-premise infrastructure.

Solution

The seasonality in demand for products led the model to be retrained and deployed involving a massive joint effort whenever there was a data drift, which lead to redeployment. Thus, this resulted in reviewing the entire approach from an automation standpoint and rearchitecting the entire platform on Azure with automated tollgates for training and versioning, and the champion model selection and lifecycle management.

Business Impact

Led to cost-savings in resourcing and infrastructure costs and the elimination of manual tollgates compared to manual deployment.

Food & Beverage

Problem Statement

Implementing shelf watch for a beverage company to understand the visibility of their brand and competition analysis.

Solution

Using AI/ML for advanced image recognition technology platform to track retail execution can mitigate the issue.

Business Impact

- Increased accuracy in results of the brand visibility exercise
- Reduced human error
- Reduction of fraud by third-party for verification
- Inventory Management helps in identifying goods that are out of stock

Industry Use Cases

Banking

Problem Statement

Maintaining security and trust in transactions without intermediaries in the fintech industry, e.g., crypto-currencies is a challenge.

Solution

An automated Machine Learning platform is well-suited to identify and help prevent identity theft, fraud, and illicit transactions in the blockchain by developing and deploying algorithms that can detect anomalous behaviour anywhere along the chain.

Banking

Problem Statement

Capturing a greater share of existing client assets and attracting new clients by developing new investment products is a huge requirement.

Solution

Robo-advisors have expanded the market for portfolio management by providing new clients with the right opportunities to match their risk tolerance and financial profile. By rapidly testing and deploying predictive algorithms, they can automatically rebalance portfolios.

Insurance

Problem Statement

The ability to predict the final claim amount has a significant impact on financial statements. Additionally, loss cost modelling relies on incurred loss amounts that are underdeveloped.

Solution

With MLOps, extremely accurate predictive models are built that lead to a better understanding of how much a claim will ultimately cost, giving confidence on how much to reserve on Incurred but ‘Not Reported Loss’ amounts. Thus, this will help to build robust and accurate pricing models.

04

MLOps Business and Operation Pillars – Investment and Change Management

34 Planning Transformation

35 Where to Invest?

36 MLOps Platform Build Vs Buy?

38 Change Management

39 Evolution of MLOps Stakeholders

Planning Transformation

Why?

A. MLOps to Mitigate Risk

MLOps is relevant to any team with even a single model in production. Depending on the model, continuous performance monitoring is required. For carrying safe and authentic operations, MLOps is key in mitigating the risks induced by the usage of ML models.

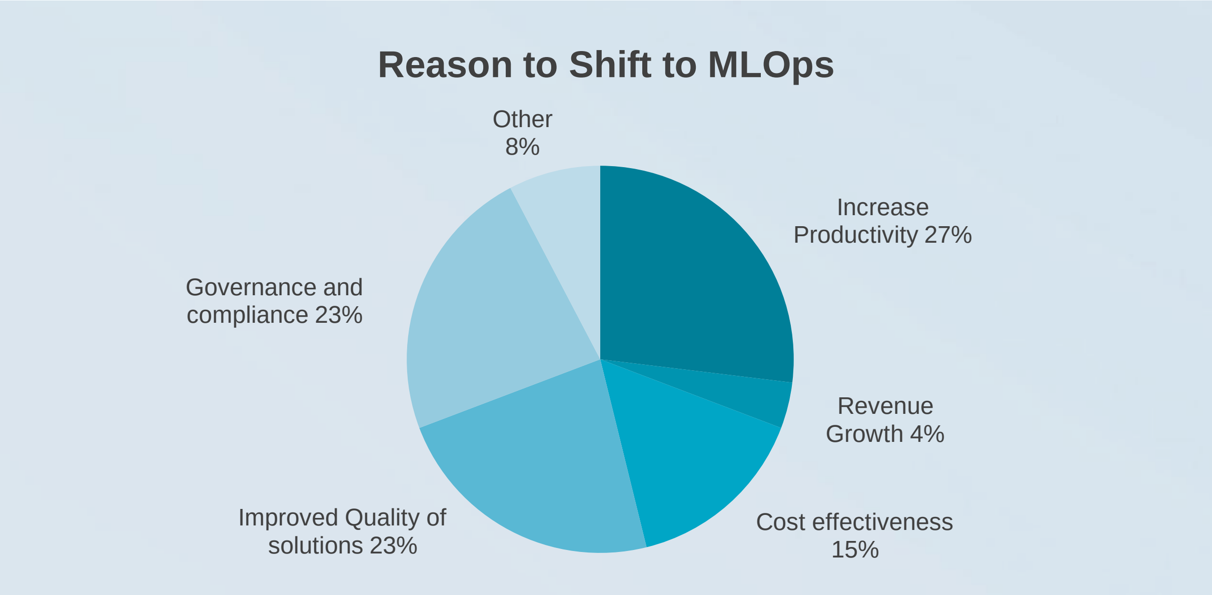
B. For Responsible AI

- **Intentionality:** To ensure models are designed and are behaving according to the business objectives
- **Accountability:** Having a clear picture of which team uses which data, as well as the confidence that the data is accurate and compliant with regulations

C. MLOps for Scale

MLOps is a critical component in deploying Machine Learning efforts at a large scale (and in turn, benefiting from the corresponding economies of scale). It helps to scale from one or a handful of models in production to tens, hundreds, or even thousands, producing a beneficial business impact.

With organisations developing a better understanding of ML and the nature of data challenges through their experience with ML solutions, MLOps is becoming a default requirement for any deployment with governance and monitoring. Our study suggests that organisations primarily move towards MLOps to increase their productivity and improve the quality of their solutions.



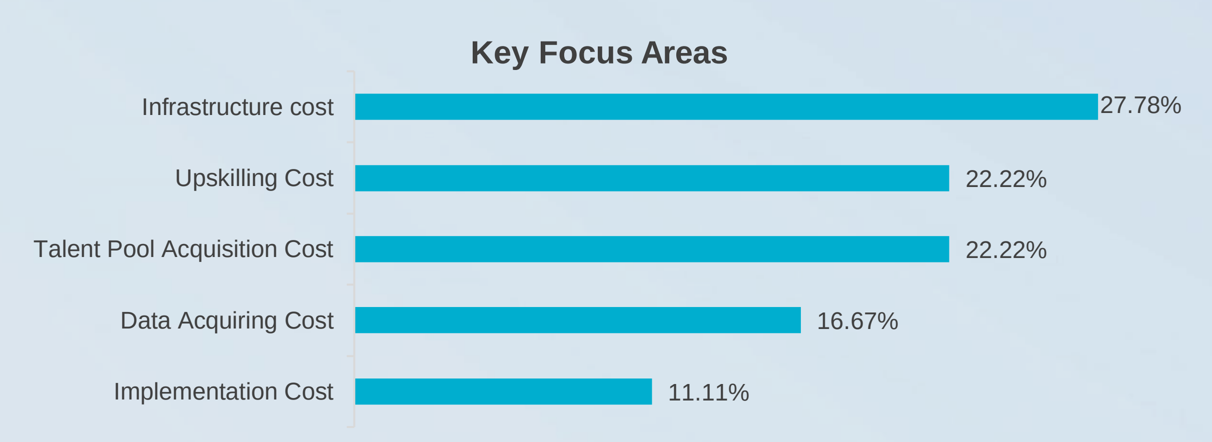
Source: NASSCOM and EY

Where to Invest?

The key focus areas organisations need to keep in mind depending on their appetite for investments are:

1. **Infrastructure Cost:** The major investment goes into building the infrastructure for MLOps. The CAPEX is generally less than OPEX because of the cloud adoption and on-demand nature of usage. The various managed services across cloud platforms have taken over the biggest work of infrastructure management, in effect helping the organisations pay for the solutions that they use.
2. **Skill Augmentation of the Workforce:** The workforce needs to be upskilled readily in order to be hands-on with the constantly evolving solutions.
3. **Talent Acquisition Cost:** The amount spend in acquiring a well-trained and skilled resource pool acts as one of the key focus areas.
4. **Data Acquiring, Management, and Governance:** Consistent acquisition of data to train and re-train models that adhere to the laws and compliances forms a formidable part of the OPEX of an organisation.

Based on the survey conducted, we can see **27.78%** of the respondents have considered Infrastructure Cost to be the key focus area for investment, followed by Upskilling Cost and the Cost to Acquire Talent.



Source: NASSCOM and EY

> **1. 2020 spend on MLOps: 525.29 Mn USD**
> **2. Projected Spend on MLOps on 2025: 4 bn USD**
> **3. Projected Spend on MLOps on 2025: 5.734 bn USD**
> - Historical Global AI and ML Ops Spend | Wonder

MLOps Platform Build Vs Buy?

Build or Buy?

The Build versus Buy decision is an extremely important one, even at the nascent stages of a company's AI journey.

It plays an important role in how AI can be leveraged, and thus, contributes to the ROI that the company can achieve. The main scenarios for which Build or Buy decisions need to be made are listed below:

Scenarios	Build	Buy
Long-term vs Short-term	1. If the solution is required for a longer period of time 2. If the project delivery time is high or the organisation is large-scaled 3. There is enough time, investment, and resources to monitor and update the platform	1. If the solution is required for a shorter period of time 2. If the project delivery time is low and the organisation is small-scale 3. Not enough time or investment to maintain and monitor the platform
DIY or SLA	Cost of maintenance is high	Cost of licensing is high
Opportunity Cost vs Licensing Cost	Cost of building time and the opportunity cost of delaying the AI project offset licensing cost	Licensing cost is considerably lower than the opportunity cost incurred while licensing
Current Solution vs the Solution in the Market	A trade-off between the solution provided by the organisation and the solutions available in the market	
Right Technical Knowledge	Ensuring that the organisation possesses the right technical know-how and resources to complete the Build project. Else, the organisation should opt for buying the platform	

MLOps Platform Build Vs Buy?

Build

The option of building an MLOps platform usually means an open-source platform will be set up and customized to suit the needs of the organisation. The most popular MLOps platforms are KubeFlow and MLFlow.

Pros	Cons
Customizability	Speed of Adoption
Extendibility	Maintenance
Community	

Buy

Many start-ups offer managed MLOps platforms that are niche and focus on a single aspect of MLOps (for example — Seldon for deployment) and can be integrated with other solutions.

Pros	Cons
Speed of Adoption	Vendor Selection
Features Without Extra Investments	Vendor Lock-in
Strategic Partner	

Open-Source Platforms

Kubeflow METAFLOW

Flyte mlflow

MLOps Platforms

Valohai iguazio allegro AI

cnvrg.io ALGORITHMIA dataiku

Change Management

Change management is an organisation's method for describing and implementing change in both internal and external activities. While adopting MLOps, an organisation goes through both technical as well as organisational changes. They are described as below:

Technical Change Management

Organisations need to adopt technical change management in terms of both, infrastructure and solutions:

1. An infrastructure needs to be built that allows smooth integration to ever-versioning models and existing systems.
2. Solutions need to be developed that have a vision for the future. The solution needs to be monitored and versioned in production so that a considerable change is not required later on.
3. Configuration model hyper-parameter, requirements, and data sources need to be monitored so that they can be changed via configuration.
4. Model quality should be checked constantly before serving.

Organisational Change Management

As firms move towards the adoption of MLOps, a plan is required for their efficient implementation and managing the changes. This is paramount to the success of MLOps implementation. Organisations need to conduct a gap analysis to find the shortfalls in the combination of existing processes and roles.

A. **Choosing limited focus areas to start MLOps Implementation:** If the organisation is just beginning the MLOps journey, it is best to have limited focus areas.

B. **Leveraging the right mix of skillset:** Many organisations dive straight into implementing MLOps, while omitting to conduct a gap analysis of the skillsets present in their existing structure. For e.g., it is a misconception that Data Scientists should shoulder the responsibility of managing the models throughout the model lifecycle. Instead, Data Scientists should focus entirely on building accurate ML models while the roles dealing with production machinery should be taken up by ML Engineers.

C. **Providing adequate resources:** As new processes and tools come into the picture, the availability of adequate trainings and resources is necessary for Data Scientists and ML Engineers to gain appropriate resources and skills to implement MLOps. The technology vendors can provide trainings to upskill them whereas an appointed team of practitioners can help them with their questions about the processes.

The subsequent section gives us a window into how the organisational roles have evolved post-MLOps adoption.

Evolution of MLOps Stakeholders

Role	Role in ML Model Life Cycle	MLOps Requirements
SME	- Provide business questions, goals, or KPIs around where the ML models should be framed - Continuously evaluate and ensure that model performance aligns with or resolves the initial need	- Easy way to understand deployed model performance in business terms - Mechanism or feedback loop for flagging model results that don't align with business expectations
Data Scientists	- Build models to address business questions - Deliver models to be used properly In production - Assess model quality	- Automated model packaging and delivery - Ability to develop tests to determine the quality of deployed models - Visibility into the performance of all deployed models from one centralized location - Ability to investigate data pipelines of each model to make quick assessments and adjustments irrespective of who originally built the model
Data Engineers	Optimize the retrieval and use of data to power ML models	- Visibility into performance of all deployed models - Ability to see the full details of individual data pipelines to address any underlying data plumbing issues
ML Engineers	- Integrate ML models in the company's applications and systems - Ensure that ML models work seamlessly with other non-Machine Learning-based applications	- Versioning and automatic tests - The ability to work in parallel on the same application
DevOps	- Conduct and build operational systems and test for security, performance, and availability - CI/CD pipeline management	- Seamless integration of MLOps into the larger DevOps strategy of the enterprise - Seamless deployment pipeline
Model Risk Managers	- Minimize overall risk to the company as a result of ML models in production - Ensure compliance with internal and external requirements before pushing ML models to production	Robust, likely automated, reporting tools on all models, including data lineage
ML Architects	- Ensure a scalable and flexible environment for ML model pipelines, from design to development to monitoring - Introduce new technologies, when appropriate, that improve ML model performance to production	- High-level overview of models and their resources consumed - Ability to drill down into data pipelines to assess and adjust infrastructure needs

05

Control & Governance

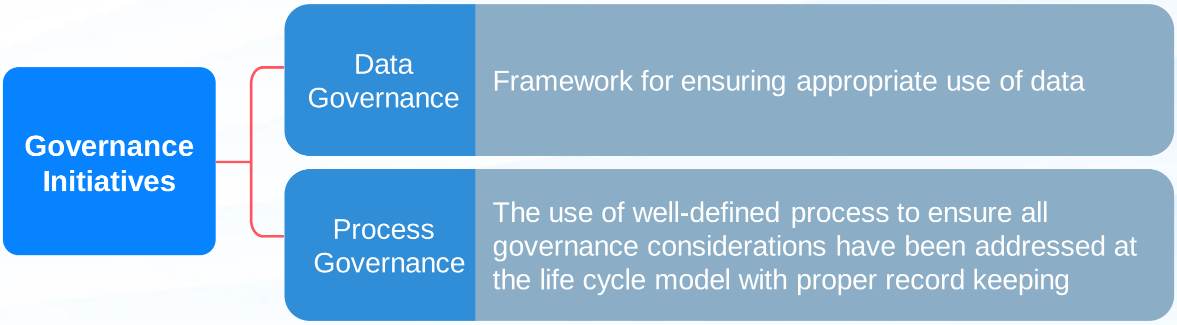
41 Data Governance

42 Process Governance

44 Current Regulations Driving MLOps Governance

Data Governance

Governments all over the world have recently enacted regulations to safeguard the public from the effects of the use of personal data by businesses. The EU General Data Protection Regulation (GDPR) of 2016, Health Insurance Portability and Accountability Act (HIPAA) of 1996, and the California Privacy Act (CCPA) of 2018 demonstrate this trend and its impact on Machine Learning. GDPR, for example, aims to protect personal data from industrial misuse with the goal of preventing discrimination against individuals. Governments are beginning to regulate Machine Learning in the hopes of reducing the negative effects of its use, with the EU leading the way with planned legislation outlining the acceptable use of various forms of AI. This is not always about limiting use as well; for example, there are certain useful applications of facial recognition technology that are now banned by data privacy laws.



Governance Initiatives

Data Governance: Framework for ensuring appropriate use of data

Process Governance: The use of well-defined process to ensure all governance considerations have been addressed at the life cycle model with proper record keeping

Data Governance

Data Governance concerns itself with the data being used and addresses questions pertaining to data provenance, data originality, data quality, and data sensitivity, which are all factors to consider. Understanding data lineage can be difficult, and anonymizing/pseudonymizing data isn't always the best method to manage personal data. Inappropriate biases in models can sometimes occur unintentionally.

Example: An ML recruitment model discriminated against women by identifying that all-female schools were less represented in Amazon’s upper management, reflecting the historical dominance of men in the organisation. Data governance tools can, thus, address these problems in their infancy.

> “ We work with a lot of PII data and the first thing to be done is to anonymize the data while working with the SMEs to identify various attributes that can be biased. Every model is sent to the validator who makes sure that the model is fair and ready for different data sets. – Survey respondent

Process Governance

Process Governance focuses on formalising and correlating actions with the steps in the MLOps process. These actions usually include reviewing, signing off, and capturing supporting items such as paperwork. The motivations for this are:

A. Ensuring that all governance-related decisions are made at the appropriate time and are implemented correctly. For instance, models should not be pushed to production until they have passed all validation checks.

B. Allowing external oversight of the MLOps process is necessary. The ability to track progress and review choices is important to auditors, risk managers, compliance officers, and the firm.

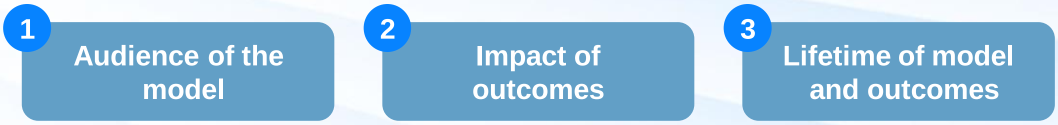
C. Audit Trail and Logging is required for MLOps to track and retain the whole lineage of prediction activities and of any model upgrades for regulatory and compliance purposes.

Effective implementation of process governance is difficult because:

1. Formal processes for the ML lifecycle are rarely straightforward to define. The knowledge of the entire process is frequently dispersed among the several teams engaged, with no single person having a thorough understanding of it.
2. Every team must be willing to embrace the process entirely for it to be successful. Teams will undoubtedly sabotage the process if it is simply too heavy-weight for specific use cases and much of the benefit will be lost.

Matching Governance with Risk Level

Governance, in the eyes of business stakeholders, is likely to slow down the introduction of new models, resulting in a loss of profitability. However, strict governance is required across the board. Those in charge of MLOps must strike a balance between getting the task done quickly and safeguarding against all hazards. This balance can be reached by assessing each project's specific risk and aligning the governance approach to that level of risk. When measuring risk, there are three dimensions to consider:



1 Audience of the model

2 Impact of outcomes

3 Lifetime of model and outcomes

Process Governance

Matching Governance with Risk Level

The risk assessment should not only determine the governance measures applied but also drive the complete MLOps deployment toolchain.

Example 1: A self-service analytics (SSA) project (one consumed by a small internal-only audience and often built by business analysts) calls for relatively lightweight governance.

Example 2: A model deployed to a public-facing website that makes decisions that impact people's lives/company finances needs an extremely thorough process. This process should essentially consider the type of KPIs chosen by the business, the type of model-building algorithm used for the required level of explainability, the coding tools used, the level of documentation and reproducibility, the level of automated testing, the resilience of the hardware platform, and the type of monitoring implemented.

Source: What is MLOps by O'Reilly

Risk in business isn't always apparent. An SSA project that takes a long-term decision might also be high risk, necessitating a robust governance measure. That is why, across the board, teams must be well-planned, with plans for MLOps risk assessment being reviewed on a regular basis.

MLOps managers must manage the inherent friction between various user profiles, establishing a balance between getting the job done quickly and defending against all dangers. This balance can be reached by assessing each project's specific risk and aligning the governance approach to that level of risk. When measuring risk, there are various factors to consider, including:

- The audience for the model
- The lifetime of the model and its outcomes
- The impact of the outcomes

The figure on the left defines the breakdown of the project criticality and operationalization approaches.

Project Criticality	Operationalization	Builder Autonomy	Versioning	Resources Separation	SLA & Support by IT	Integration to Ext. Systems
Irregular Ad-hoc usage	SSA with run on design node	★★★	—	—	—	—
Scheduled but can be inoperative for a small amount of time	Self-service development and scheduling	★★★	★★★	★★	—	—
Scheduled and requires specific monitoring	Light deployment process with rough OA and scheduling	★	★★★	★★★	★	—
Operational projects that cannot suffer outages	Fully controlled deployment CI/CD	—	★★★	★★★	★★★	★★★

Choosing the right kind of operationalization model and MLOps depending on the project's criticality.

Source: https://www.oreilly.com/library/view/what-is-mlops/9781492093626/ch04.html

Current Regulations Driving MLOps Governance

Current Regulations Driving MLOps Governance

Industry-specific regulation, which is particularly important in the financial and pharma sectors, and broad-spectrum regulation, which addresses data protection, are the two existing regulations that have a substantial influence on ML governance.

Pharmaceutical Regulations in US: GxP

The U.S. Food and Drug Administration (FDA) has created a set of quality guidelines (such as the Good Clinical Practice, or GCP) and regulations to guarantee that bio and pharmaceutical products are safe. The GxP guidelines are focused on:

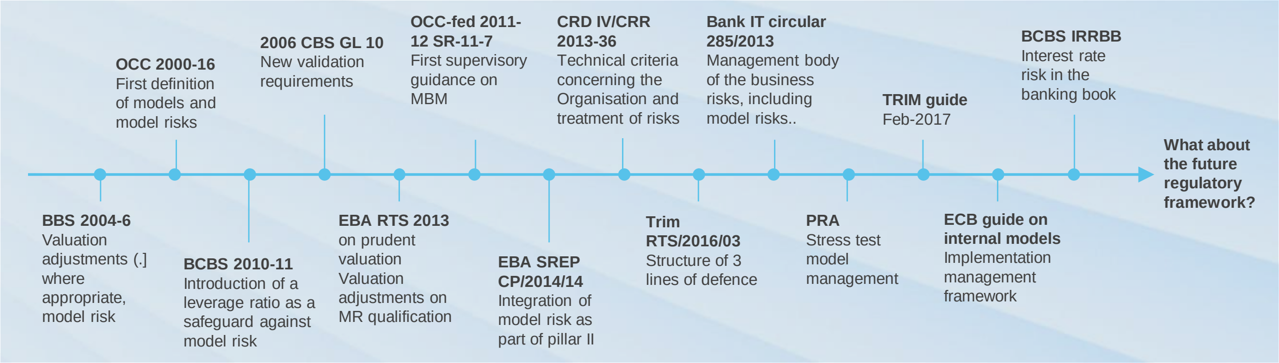
1. Traceability, which is the ability to reconstruct a drug's or medical device's development history
2. Accountability, which refers to who contributed what and when to the creation of a medicine
3. Data Integrity (DI), which is the dependability of data used in development and testing; this is based on the ALCOA principle: traceable, legible, contemporaneous, original, and accurate, with risk identification and mitigation techniques being considered

Financial Model Risk Management Regulations

The effect of unusual occurrences such as financial catastrophes, and the resultant harm to the public and the wider economy, if catastrophic losses are sustained, has prompted model risk management (MRM) regulations. Following the financial crisis of 2007–2008, a slew of new regulations has been enacted to compel appropriate MRM practices. The UK Prudential Rule Authority's regulation, for example, establishes these four principles:
For example, The UK Prudential Regulation Authority's regulation defines four principles

1. Model definition is necessary, which is the process of defining a model and storing it in inventory
2. Model risk governance structure, policies, procedures, and controls are to be established by risk governance
3. Model creation, implementation, and use are all made easier by lifecycle management
4. A successful challenge includes model validation and independent assessment

History of MRM Regulation



BBS 2004-6
Valuation adjustments (.] where appropriate, model risk

OCC 2000-16
First definition of models and model risks

BCBS 2010-11
Introduction of a leverage ratio as a safeguard against model risk

2006 CBS GL 10
New validation requirements

EBA RTS 2013
on prudent valuation Valuation adjustments on MR qualification

OCC-fed 2011-12 SR-11-7
First supervisory guidance on MBM

EBA SREP CP/2014/14
Integration of model risk as part of pillar II

CRD IV/CRR 2013-36
Technical criteria concerning the Organisation and treatment of risks

Trim RTS/2016/03
Structure of 3 lines of defence

Bank IT circular 285/2013
Management body of the business risks, including model risks..

PRA
Stress test model management

TRIM guide
Feb-2017

ECB guide on internal models
Implementation management framework

BCBS IRRBB
Interest rate risk in the banking book

What about the future regulatory framework?

Source: [*https://www.oreilly.com/library/view/what-is-mlops/9781492093626/ch04.html*](https://www.oreilly.com/library/view/what-is-mlops/9781492093626/ch04.html)

Current Regulations Driving MLOps Governance

GDPR and CCPA Data Privacy Regulations

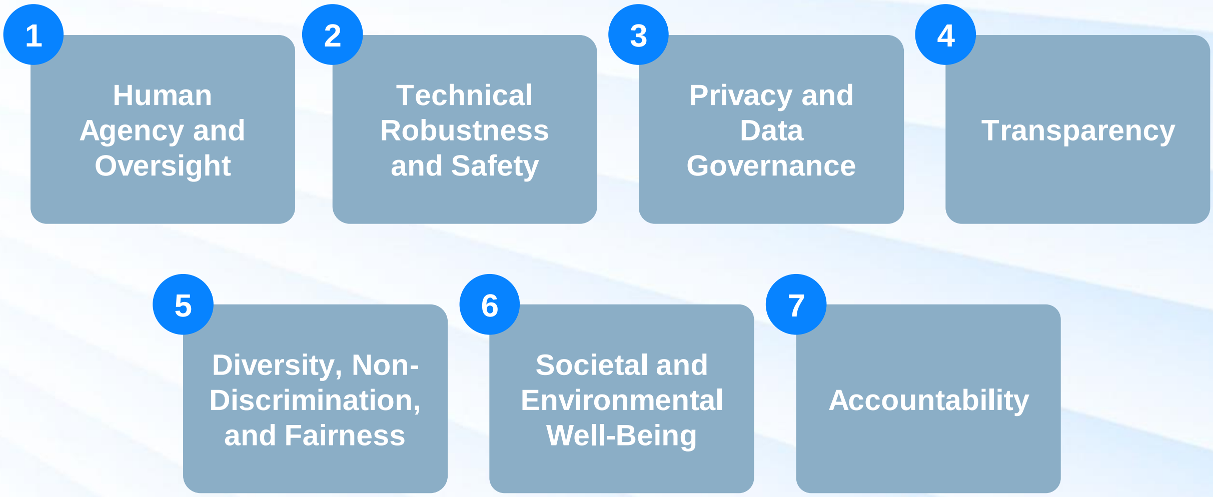
The EU General Data Protection Regulation (GDPR), which went into effect in 2018, establishes criteria for the gathering and processing of personal data from EU residents as well as EU visitors to any website, regardless of its location. The goal of the regulation is to provide people control over their personal data obtained by IT, including their rights to:

- Be informed about data collected or processed
- Access collected data and understand its processing
- Correct inaccurate data and the Right To Be Forgotten (i.e., to have data removed)
- Restrict the processing of personal data
- Obtain collected data and reuse it elsewhere object to automated decision-making

The California Consumer Privacy Act (CCPA) is similar to GDPR in terms of who and what is protected.

New Wave of AI-Specific Regulations

A new wave of legislation and standards aimed at AI and Machine Learning applications is gaining traction throughout the world. The EU is leading the way by attempting to build a framework for trustworthy AI, which identifies seven fundamental conditions that AI systems should meet in order to be considered trustworthy:



1 Human Agency and Oversight

2 Technical Robustness and Safety

3 Privacy and Data Governance

4 Transparency

5 Diversity, Non-Discrimination, and Fairness

6 Societal and Environmental Well-Being

7 Accountability

Current Regulations Driving MLOps Governance

Initiatives by Government of India to Regulate AI

The government has been compelled to examine AI's progress and ramifications. For the time being, the government intends to expand AI applications in the Indian context. The Ministry of Electronics and Information Technology (MeitY), the Ministry of Commerce and Industry, the Department of Telecommunications, and NITI Aayog are among the ministries that have put their foot forward to take the lead on AI legislation. According to NITI Aayog's ("National Strategy for Artificial Intelligence") reports, the government plans to take advantage of the "late mover's advantage" in the AI sector by "consistently delivering homegrown pioneering technology solutions" in AI to leapfrog and catch up with the rest of the world. Some of the government's initiatives are:

Committee and Taskforce

- In 2017, the Ministry of Commerce and Industry sets up an AI task force highlighting the sectorial importance of an AI regime and the challenges for the adoption of AI in India
- In 2018, NITI Aayog was directed to initiate AI programmes and applications. MeitY constituted four committees in order to develop a policy framework and analyse issues like leveraging AI, legal and ethical issues to AI, etc.
- In January 2020, NITI Aayog recommended an AI-explicit computer framework, 'AIRAWAT' to be set up to satisfy the processing needs of innovation hubs, start-ups, AI researchers, and students

Laws and Regulations

- Copyright Act: Source code and object code of AI are protected as 'literary works'
- Patents Act: An AI application may be patented if it is attached to an invention along with the hardware and proves that the hardware is essential along with the software
- Big Data: It is classified under Personal Data, which is protected under 'Right to Life'. Any other data is not governed by a legislation
- Competition Act: Seeks to prevent adverse effects on the competition

Competition Act – Example

Amazon and Flipkart have large repositories of data due to their unparalleled market base and market power. They analyse the data for targeted advertisements based on consumer preferences and marginalise other competitors who are unable to capture the market due to their lack of access to data. Lack of such access to data and cost associated with the development of complex self-learning computing algorithms has resulted in the creation of high entry barriers on account of network effects.

This has been noted by CCI, which has ordered an investigation focused on deep discounting, preferential listing, and market power. The case "Delhi Vyapar Mahasangh v. Flipkart and Amazon' is still ongoing.

Source: https://www.globallegalinsights.com/practice-areas/ai-machine-learning-and-big-data-laws-and-regulations/india

06

Innovation and Future

48 The Future of MLOps

The Future of MLOps

Design, deployment, and monitoring processes have all been standardised, courtesy of MLOps, and data science and Machine Learning platforms have been built to help with this. Designers of consumption forecasting systems can use these standard procedures and platforms to boost their systems' efficiencies in terms of cost, quality, and time to value.

Taking a step back, it is evident that different industries have diverse Machine Learning use cases, each with its own set of complexities when it comes to defining the problem, constructing models, and pushing to production. However, regardless of the industry or use case, MLOps processes are the common thread that enables data teams (and more broadly, entire companies) to scale.

MLOps has benefitted companies by manifolds, in terms of cost-effectiveness, increasing productivity, improving quality of solutions, etc.

Quote 1

“MLOps has benefitted our organisation in Code optimization, and creation of ML toolkit algorithms from our re-usable/modular ML Ops framework (10-15% reduction in cost and 25% improvement in accuracy)”

Quote 2

“It has helped in faster management deployment of ML model, creating transparency in monitoring model performance”

Future of MLOps

Considering the future, based on our study, many organisations are looking forward to having **centralized ML Operations** in the future, moving away from the current de-centralised approach. The advantage of this type of centralized learning is that the model can generalize based on data from a group of devices, and thus, instantly work with other compatible devices. Centralized learning also entails that data can explain all the variations in the devices and their environments.

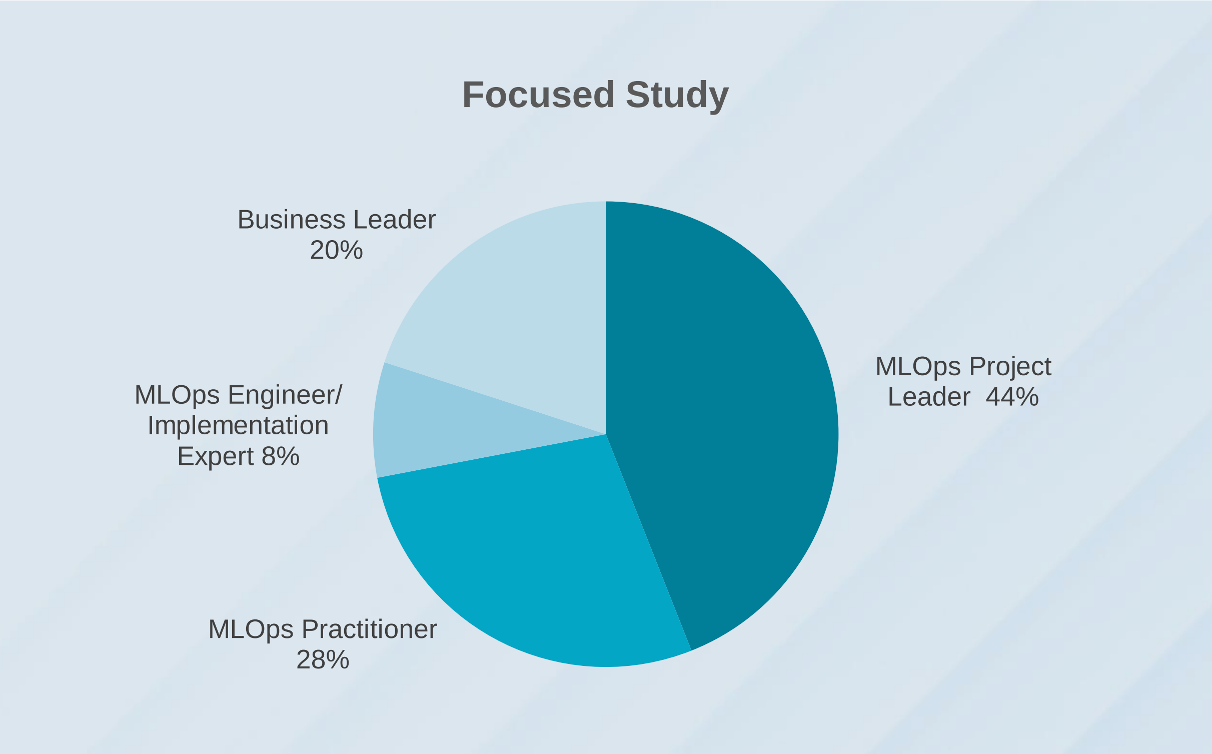
MLOps, still being unchartered territory for many, is becoming a need of the hour for almost all the organisations across different sectors. Today, we expect the software on which the business is built to be scalable, reliable, and efficient. And if the benefits of AI are to be reaped, the same must be true of the models that increasingly drive business decisions. For years, AI was optimized the way the software was built, run, and maintained through DevOps and now, it is time to do the same for Machine Learning. It is extremely important to make AI work at scale with MLOps to ensure that the business derives the most value from investments in Machine Learning.

About the Survey

The focus behind this study was to understand the fairly new concept of MLOps and to come up with a detailed playbook providing a holistic view of the journey of MLOps adoption by organisations.

For preparing this report, we conducted interviews with a focussed group of experts in the field of MLOps, who are renowned players and service providers in the market. We also conducted secondary research to understand the different nuances, global practices, and challenges in the field of MLOps to arrive at a holistic view of the best practices for the organisations adopting MLOps and how to mitigate the challenges that arise in the journey.

In our focussed study, we had a good mixture of MLOps Project Leaders, Business Leaders, Implementation Experts, and MLOps Practitioners to gain a comprehensive perspective on the report, awarding the MLOps Project Leader's perspective the most significance to understand the MLOps Adoption Journey by various organisations in detail.



Source: NASSCOM and EY

07

Appendix

51 Appendix I

52 Appendix II

53 Appendix III

Appendix I

ML Evolution and Why MLOps?

Figure 1. Evolving expectations from enterprise AI initiatives

2000 - 2021	VS	2021 and beyond
Create analytical assets - ML and mathematical models, prototypes, and predictive scorings	Investments	Develop production-ready, scalable AI solutions
Explore innovative technologies	Focus	Deliver business value by realizing AI/ML capabilities
Absence of data management and governance	Strategy	Organisation-wide cohesive data management, ML model governance, and intellectual property platforms
Lack of operationalization foresight and AI integration complexities	Execution	Prioritize high-impact solutions, streamline ML operationalization, and scale integration processes

Source: Genpact

ML models that go into production need to handle a large volume of data, often in real-time. Unlike traditional technologies, AI and ML deal with probabilistic outcomes — what is the most likely or unlikely result.

Therefore, the moving parts of the ML model, when deployed, need close monitoring and swift action to ensure accuracy, performance, and user satisfaction. There are three key factors that influence the proper development of ML models:

Data Quality: Since ML models are built on data, the quality, completeness, and semantics of data are critical in production environments.

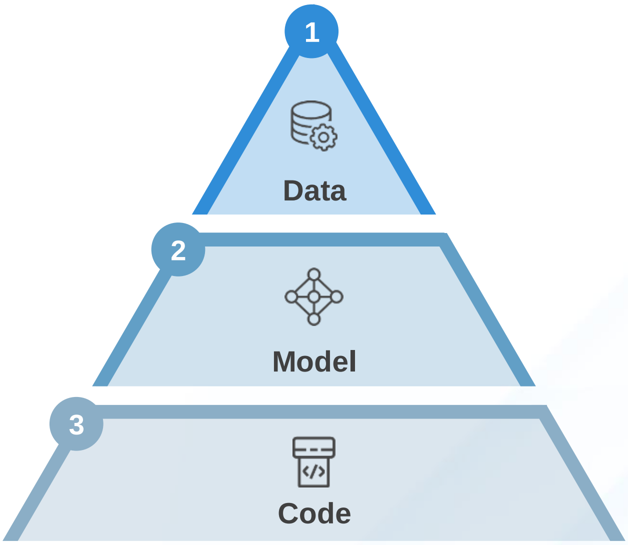
Model Decay: In ML models, data patterns change as the business environment evolves. This evolution leads to a lower prediction accuracy of models trained and validated on outdated data.

Data Locality: Data locality and access patterns are used to improve the performance of a given algorithm. However, such ML models might not work correctly in production due to the difference in the quality metrics.

Thus ML models need to evolve to tackle these challenges. Data and analytics leaders have to look for repeatable and scalable standalone software applications. In other words, they must rely on Machine Learning Operations or MLOps.

Appendix II

Three Layers of ML Model



1

The fundamental part of any Machine Learning workflow is Data. Collecting good data sets has a huge impact on the quality and performance of the ML model. The famous citation “Garbage In, Garbage Out” in the Machine Learning context means that the ML model is only as good as your data. Therefore, the data, which has been used for training the ML model, indirectly influences the overall performance of the production system. The amount and quality of the dataset are usually problem-specific and can be empirically discovered.

Owing to its significance, data engineering is substantially time-consuming. Most of the time on a Machine Learning project shall be spent on constructing data sets, cleaning, and transforming data. The data engineering pipeline runs a sequence of operations on the available data. The final goal of these operations is to create training and testing datasets for the ML algorithms. These operations consist of **Data Ingestion, Exploration and Validation, Data Wrangling (Cleaning), and Data Splitting.**

2

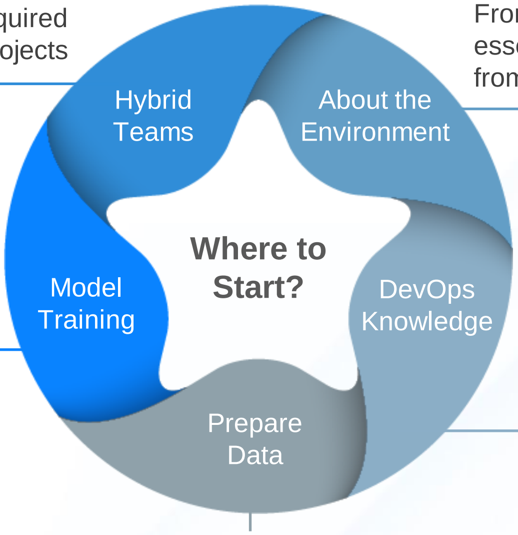
This is a critical stage of the ML workflow — the phase of writing and executing Machine Learning algorithms to obtain an ML model. The model engineering pipeline is usually utilized by a data science team and includes several operations that lead to a final model. These operations include **Model Training, Model Evaluation, Model Testing, and Model Packaging.**

3

The final stage of delivering an ML project includes the following three steps:
Model Serving: The process of deploying the ML model in a production environment
Model Performance Monitoring: The process of observing the ML model performance based on live and previously unseen data, such as prediction or recommendation
Model Performance Logging: The process of every inference request resulting in a log record

Appendix III

Change Management



A hybrid team with a multidisciplinary skillset is required for the success of projects

From hardware to software, it is essential to check everything from the start

Define and understand what KPIs and metrics are useful for successful model

Having a fair understanding of DevOps is quintessential to start with MLOps

It is essential to gather data regularly and keep it up to date

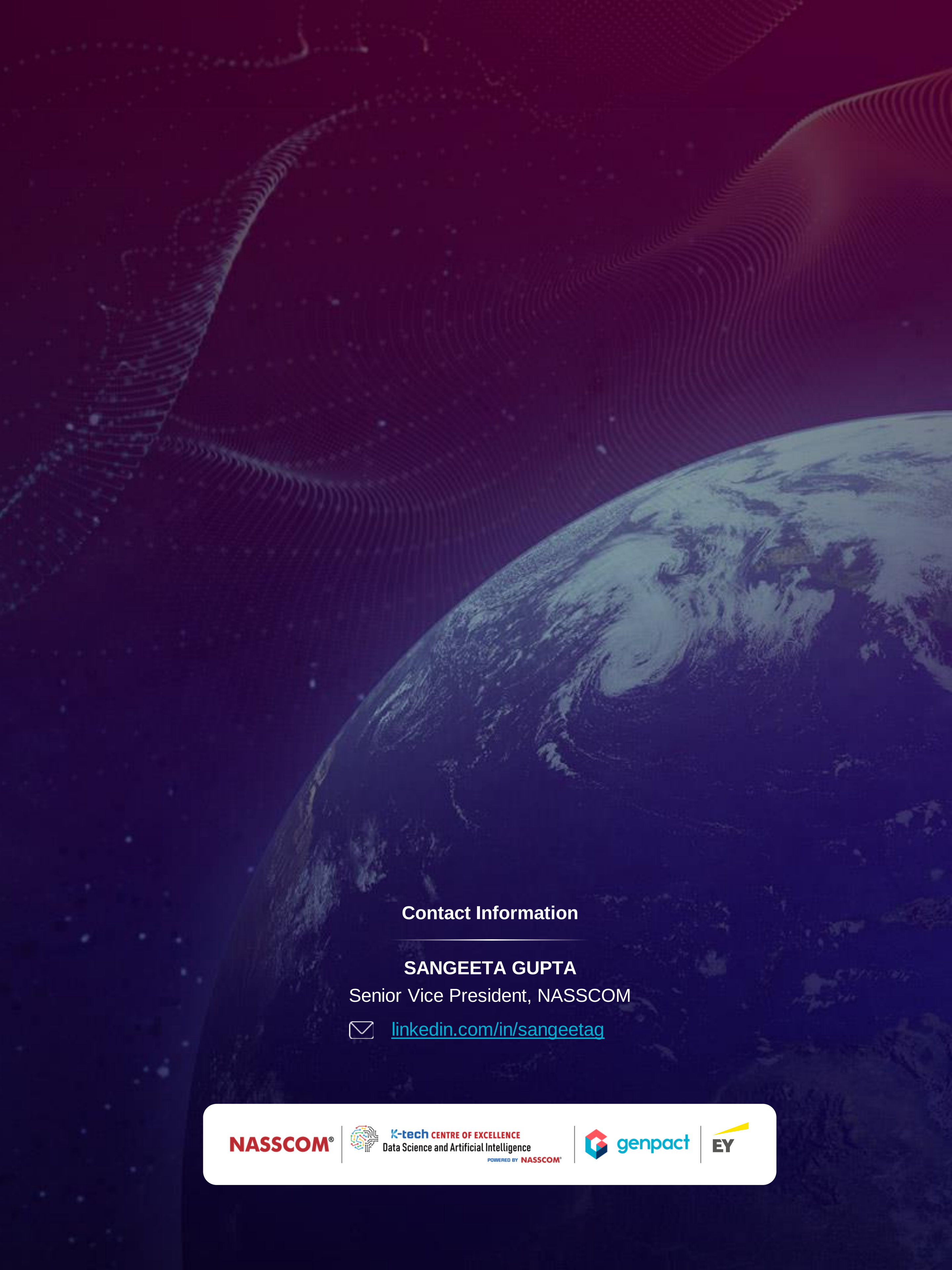
Source: EY

This compendium showcases the rapid progress in the adoption of MLOps across industries and functions. The focused group study and interactions with AI/ML leaders, practitioners, business leaders, and core engineering teams revealed the challenges and best practices that organisations follow. The next decade of AI/ML initiatives in the industry will witness rapid growth and therefore, require concerted efforts to define strategies for industrialization vs experimentation alone.

Together with our partners, vendor network, customers, and subject matter expertise, we can reimagine the adoption of AI initiatives at an enterprise level. The report seeks to enable an AI/ML and MLOps strategy-to-execution journey and we will be glad to join this endeavour of your enterprise.

Sreekanth Menon

AI/ML Practice Leader, Genpact



Contact Information

SANGEETA GUPTA

Senior Vice President, NASSCOM

✉ linkedin.com/in/sangeetag

NASSCOM® | K-tech CENTRE OF EXCELLENCE Data Science and Artificial Intelligence POWERED BY NASSCOM® | genpact | EY