Understanding MLOps to Operationalize Machine Learning Projects

Published 26 April 2021 - ID G00750904 - 12 min read

By Analyst(s): Farhan Choudhary, Sumit Agarwal

Initiatives: Artificial Intelligence

Enterprises are increasingly challenged with operationalizing their artificial intelligence and machine learning initiatives. Data and analytics leaders should use the step-by-step approach set forth in this research to manage and expand their AI initiatives.

Additional Perspectives

- Summary Translation: Understanding MLOps to Operationalize Machine Learning Projects
  (27 September2021)

Overview

Key Findings

- Organizations struggle with scaling artificial intelligence (AI) and machine learning (ML) initiatives because of data, security, privacy and infrastructure concerns. This often makes it difficult for data and analytics leaders to measure value.
- The prime focus for organizations is to accelerate the speed at which the proofs of concept (POCs) can move into production. Hence, operationalization is toward moving minimum viable products (MVPs) into production and continuous enhancements of AI-based systems.
- Across all maturity levels, ML is the most leveraged AI technique, but not the only one. Analytical projects span from business intelligence to rule-based systems to even optimization and graph techniques.

Recommendations

Data and analytics leaders responsible for AI and ML initiatives should:

- Prepare for common ML operationalization pipeline pitfalls around data collection, model development, deployment and adoption in advance.

- Leverage DevOps best practices to ensure collaboration and open communication between the data, model and application engineering practices.
- Embrace AI engineering and ModelOps to operationalize all models (e.g., analytical, statistical, ML and graph).

Analysis

> **Organizations struggle with scaling AI initiatives for many reasons. According to the 2019 Gartner AI in organizations survey (see Note 1), security and privacy concerns, integration complexity, and potential risks and liabilities exist on top of data challenges. This is compounded by a lack of understanding of AI's benefits and uses and unavailability of technology knowledge to operationalize AI. As a result, organizations take close to nine months to move AI-based system prototypes into production.**
>
> — *Gartner (April 2021)*

The challenge of operationalizing data science and machine learning initiatives is that the requirements of tools, data and best practices are different than for traditional software engineering projects. The AI workflow requires the flexibility to accommodate multiple iterative cycles of model validation and tuning before it's ready for integration. Once this is done, the model must still be monitored and retrained in order to optimize inference. During all this, snapshots of the data must be captured for iterative model training and testing.

Table 1: AI Workflow

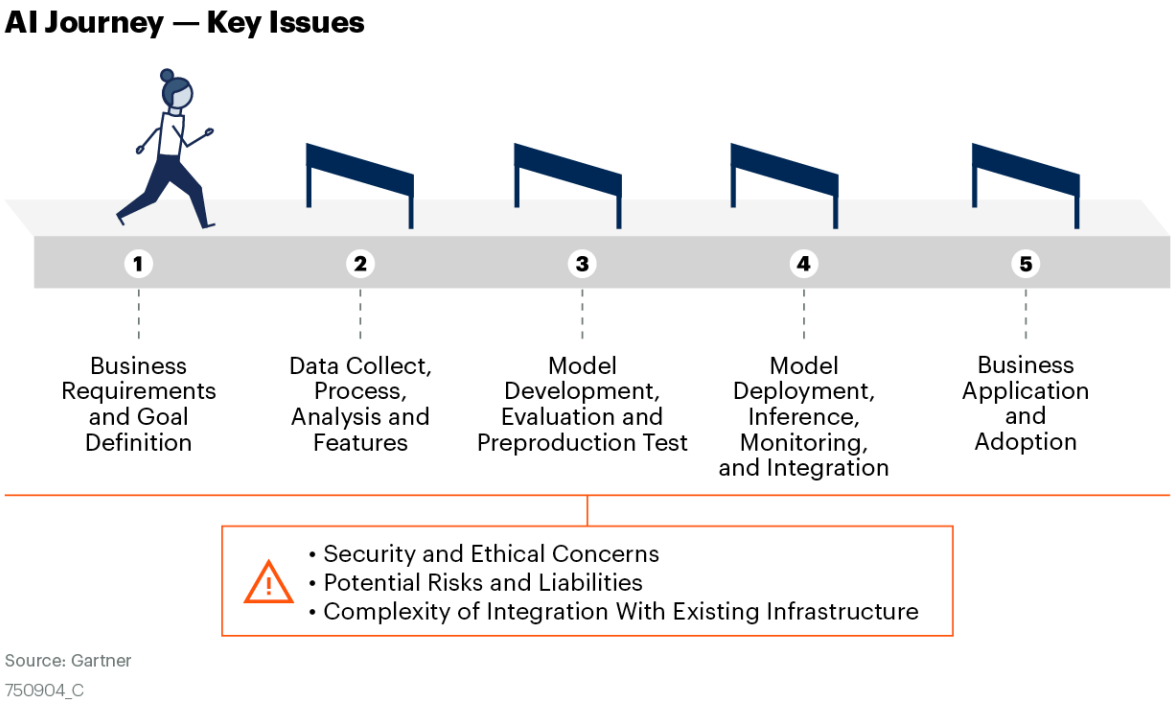
(Enlarged table in Appendix)

1. Business Requirements and Goal Definition	2. Data Collect, Process, Analysis and Features	3. Model Development, Evaluation, Pre-Prod Test	4. Model Deployment, Inference, Monitoring, Integration	5. Business Application and Adoption
■ Lack of Understanding AI Benefits and Uses ■ Lack of Operationalization Skills ■ Unable/Hard to Measure Business Value	■ Data Volume or Complexity ■ Data Scope or Quality Problems ■ Data Accessibility Challenges	■ Lack of Technology Knowledge ■ Lack of Capability to Leverage AI Techniques ■ Little Improvement Over Existing Technologies	■ Lack of Formal Operationalization Methodology ■ Communication Gaps ■ Unable to Deal With ModelOps to Ensure Reproducibility and Reusability	■ Lack of Business Stakeholder Involvement ■ Inconsistent Provisioning of Models ■ Failure to Track and Monitor Models

Source: Gartner

Organizations often struggle with providing a continuous feedback loop for models in production. Hurdles exist at several critical touchpoints (see Figure 1).

Figure 1: AI Journey — Key Issues



Common Risks and Pitfalls While Operationalizing Machine Learning Projects and How to Overcome Them

Data and analytics leaders responsible for ML and AI initiatives may find themselves struggling (see Figure 1) with one or more of these pitfalls:

- **Incoherent data pipeline:** Devoting insufficient time and resources to data preparation and building a consistent data pipeline (primary or retraining) to support ML can lead to a mismanagement of ETL scripts. Too many data acquisition and transformation scripts with overlapping business rules lead to duplication of effort in IT operations.
- **Lack of collaboration between teams:** A machine learning project cannot be operationalized by a single person, and even in scenarios where it could be, friction occurs during handoffs between various teams such as data, ML and application engineering. To operationalize an ML project, diverse skills are required from all parts of the organization (see Staffing Data Science Teams: Mapping Capabilities to Key Roles).

- **Managing tools/platform integration complexity:** As organizations become more mature in their AI and ML implementations, they will likely have an arsenal of platforms, often overlapping in their capabilities, for different personas or tasks. Ensure that a common architecture pattern exists. This will help bring consistency to the delivery method and management of ML artifacts within different data engineering, data science and machine learning (DSML) tools (see also Magic Quadrant for Data Science and Machine Learning Platforms).

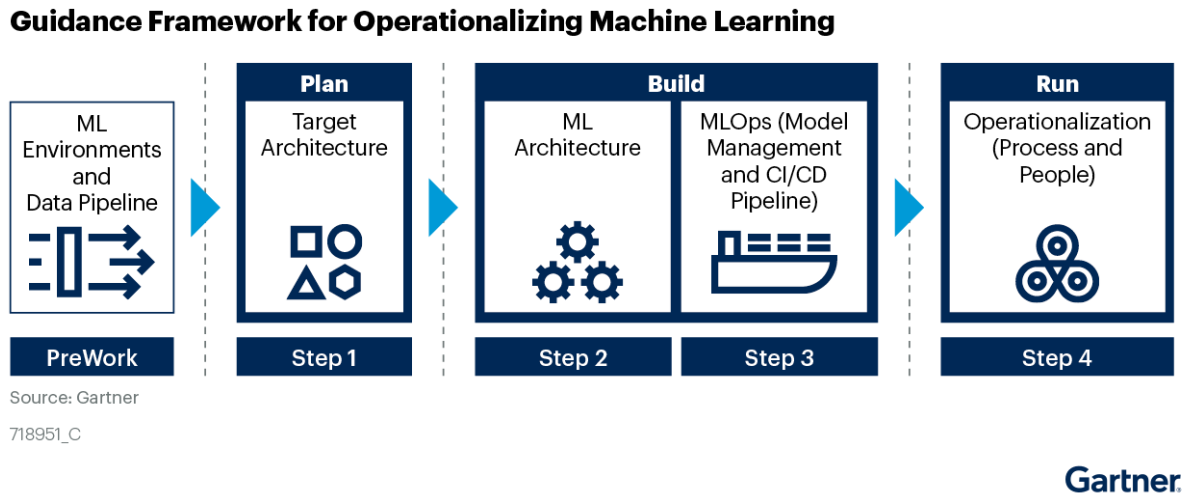
- **Lack of model monitoring and governance:** Failing to catalog models and interpret their outputs across multiple AI-based system implementations can lead to various complications to model audit, model reproducibility and model reusability. This results in delayed value realization or diminished returns from operationalized models.

- **Lack of foresight:** Operationalization of machine learning is an imperative step in aligning ML and AI investments with strategic business objectives — the "last mile" toward delivering business value. Think of your initiatives in terms of sustained operationalization, with capabilities that ensure model resilience such as model security, model explainability, model retuning and rollback, and model management, which ultimately allow for consuming insights within different business processes (see Innovation Insight for ModelOps).

- **Inconsistent workflow:** A coherent machine learning workflow provides a structured approach to dealing with the proliferation of new tools and the exchange of data and model assets across multiple environments and is key to successful operationalization. Data and models suffer from drifts that require teams to pay more attention to crafting strategies that guarantee the most uptime of models in production (see Use Gartner's 3-Stage MLOps Framework to Successfully Operationalize Machine Learning Projects).

The Guidance Framework for operationalizing machine learning, as shown in Figure 2, shows how to carry the process step by step.

Figure 2: Guidance Framework for Operationalizing Machine Learning



Step 1: Use the Right Environments and Pipelines

One of the best ways to successfully operationalize an ML project is to create different environments for users. These include development (dev) for experimentation, preproduction as a scaled-down version of production and production (prod) — the mission-critical inference production environment.

ML pipeline: In the case of the ML pipeline, each environment should have its own set of ML artifacts (model code, configuration file and metadata). These should come alongside a copy of the data used during the model building, training phase or the extraction, transformation and loading (ETL) scripts. This helps avoid any kind of configuration drift across environments and will support auditability and lineage when debugging an AI-based system.

Data pipeline: To support the ML pipeline, two data pipelines need to be created. One will support the build and train stages of the machine learning development life cycle (MLDLC). The second will support the retraining of the models based on the inference data and model monitoring data being collected from the deployed model in production.

Step 2: Designing Architecture to Support Machine Learning Development Life Cycle (MLDLC)

The logical architecture is divided into three sections:

1. **Acquire:** This process lays down the foundation to support discovery analytics, the ML model training and the retraining process

2. **Organize and analyze:** The logical data warehouse (LDW) helps with organizing data with varied formats and facilitates the discovery analytics, hypothesis validation, and building of the training and testing datasets for the ML model. The data science and ML tool or platform will facilitate the steps to identify the right algorithms and build the candidate model for production.

3. **Deliver and manage:** The model management system should support the capabilities of model registry (i.e., model versioning), maintaining lineage information and model manifestation, dependencies, inference scripts, sample data, testing and training data, and the schema.

Step 3: Machine Learning Architecture

A typical machine learning architecture includes five functional stages:

- **Data acquisition:** This brings data from various sources using data integration tools to establish a primary data pipeline.
- **Data processing:** Data processing also extends to feature engineering and includes steps for preprocessing the data for the training of the ML model.
- **Model engineering:** Model engineering refers to modeling the data in parallel to identifying the right machine algorithms.
- **Training:** The processed training and retraining data is forwarded for use in the execution of ML routines (such as A/B testing, model tuning and hyperparameter tuning).
- **Deployment:** The candidate model is integrated within an enterprise application or an analytics platform. The inferences from the model — model score, output data stream and insights — now affect the outcome of an ML-powered AI-based application

All of these steps can be performed by most DSML platforms and capabilities such as augmented data preparation, model training and tuning. These can be leveraged for faster delivery. For example, DataRobot, Dataiku, H2O, RapidMiner, KNIME, IBM and Alteryx, among others in the Magic Quadrant for Data Science and Machine Learning Platforms, have such capabilities. For additional information, review the Magic Quadrant and evaluate platforms and vendors using Critical Capabilities for Data Science and Machine Learning Platforms, Toolkit: RFP for Data Science and Machine Learning Platforms and Solution Criteria for Data Science and Machine Learning Platforms.

Step 4: ML Operationalization (MLOps)

MLOps aims to standardize the deployment and management of ML models alongside the operationalization of the ML pipeline. It supports the release, activation, monitoring, performance tracking, management, reuse, maintenance and governance of ML artifacts (see Use 3 MLOps Organizational Practices to Successfully Deliver Machine Learning Results).

The MLOps framework has been derived using the cross-industry standard process for data mining (CRISP-DM) methodology (see Note 2) and is composed of three main cycles: development, release and activation (production).

- **Development:** Users start with discovery analytics, hypothesis validation, data preprocessing, selecting the ML algorithm, training, building the candidate model and retraining the model.
- **Release:** The test/release cycle is where the candidate model is verified once integrated within the enterprise application of the analytics platform it was intended for. The validation is to ensure the inference and output key performance indicators (KPIs) of the ML model still hold true in the real world of the business workflow.
- **Activation:** Once the model has been tested and is performing as intended in a test environment after integration with its intended endpoint AI-based system service, it is ready to be moved into a production environment. The goal is to activate that model within existing business processes across the organization at the endpoints identified in the development process.

See Use Gartner's 3-Stage MLOps Framework to Successfully Operationalize Machine Learning Projects for a detailed, step-by-step process to operationalize ML projects.

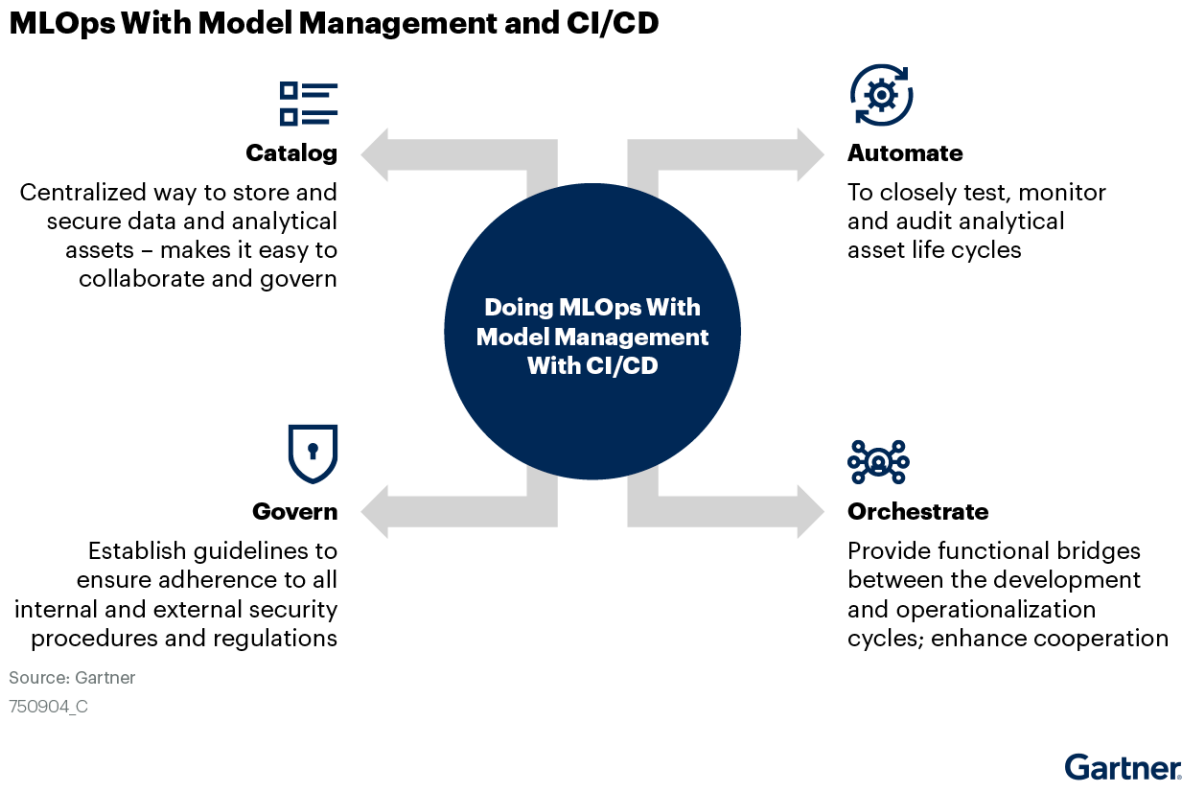
The MLOps process also supports the CI/CD framework, and it derives its core principles from the best practices of DevOps.

- **Model management system:** This helps with versioning of the models, tracking models in production, providing an AI or a containerization framework and monitoring the models in production (see Getting Started with Machine Learning Monitoring in Production).

- **CI/CD pipeline:** CI/CD pipelines are key components for any mature software development environment. Similarly, they are an essential part of the MLDLC to ensure continuous delivery of ML models. The continuous delivery (CD) phase is where the updates with the application are automatically deployed to production.

Doing MLOps with a robust model management system along with a CI/CD pipeline has multiple benefits (see Figure 3).

Figure 3: MLOps with Model Management and CI/CD



Enable Collaboration Between Data Engineering, Model Engineering and Application Engineering

Data engineering: In order to fully support the ML pipeline, it is essential that you build a robust, reproducible and reusable data pipeline to ingest and persist data from various sources with varying structures (see Operational AI Requires Data Engineering, DataOps and Data-AI Role Alignment).

Data engineers should be leveraged to build and maintain the data pipelines (primary or retraining). At the same time, data scientists and citizen data scientists should be provided with the appropriate wrangling and data science tools, giving them the flexibility to transform the data, but in a governed self-service environment.

Model engineering: The highest level of friction occurs during the handoffs of data assets to model engineering teams and from model engineering to application engineering teams. Establishing a consistent framework for model development is more about orchestrating the tasks and processes between the integration points.

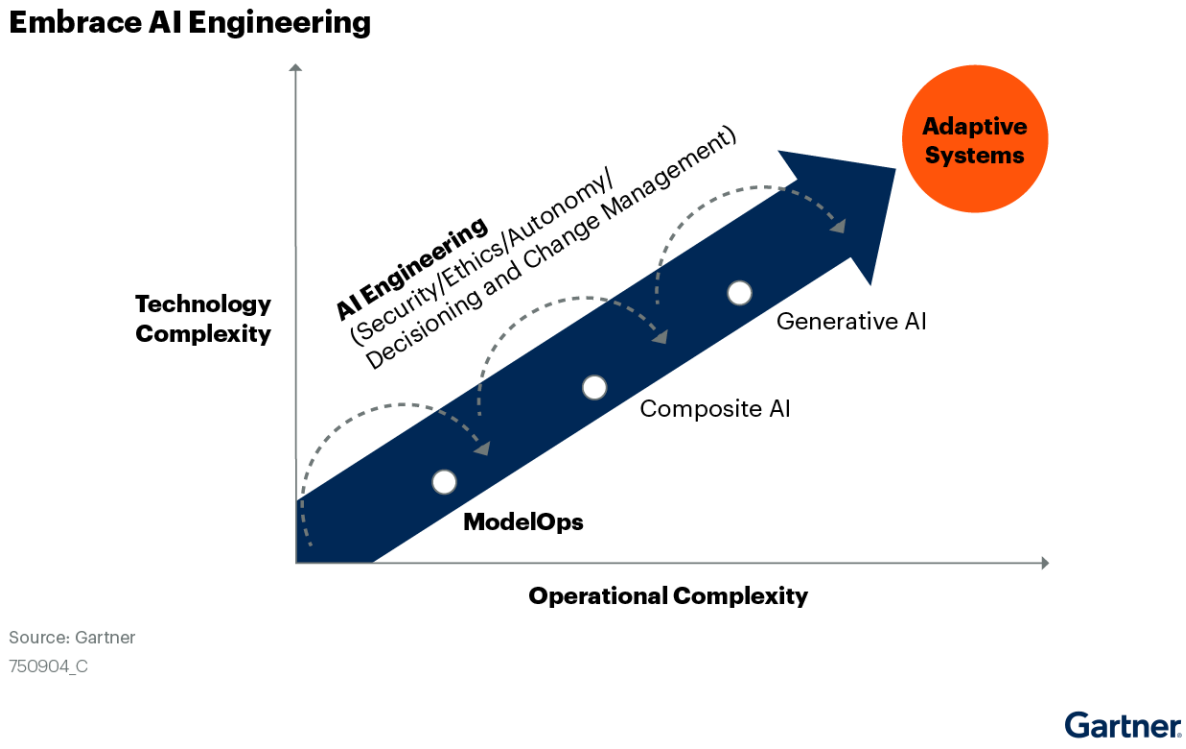
A plan to supplement the data science team with support from DevOps, ML engineers and architects, as well as with exposure to workflow management tools, develops a consistent process for delivering models.

Application engineering: Once the models have been developed, it is the responsibility of ML engineers and software engineers/developers to integrate the solution into a business application or package it into an API for further consumption. DevOps tools that can aid in tight application packaging, model service granularity, consistent code versioning and automated software delivery should be considered as part of the AI development life cycle.

Embrace AI Engineering

AI engineering brings together various disciplines from across the organization to tame the AI hype while providing a clearer path to value when operationalizing the combination of multiple AI techniques (including ModelOps; see Figure 4). AI engineering also includes responsible AI, dealing with risk, trust, transparency, ethics, fairness and accountability. A sound AI engineering practice will increase an organization's ability to move AI projects beyond pilots and prototypes into production, thus overcoming fatigue and skepticism and repositioning AI as a transformative and differentiating technology (see Top Strategic Technology Trends for 2021: AI Engineering).

Figure 4: Embrace AI Engineering



AI engineering also deals with the combination of AI techniques (or composite AI). Organizations are combining different AI techniques to:

- Improve the efficiency of learning
- Broaden the level of knowledge representations
- Solve a wider range of business problems more efficiently

A robust AI engineering strategy facilitates the performance, scalability, interpretability and reliability of AI models while delivering the full value of AI investments. Therefore, you should look at model operationalization (i.e., ModelOps), which includes MLOps as one of the fundamental steps toward establishing AI engineering.

Evidence

Note 1: Gartner's 2019 AI in Organizations Study

Gartner's 2019 AI in Organizations Study was conducted online during November and December 2019 among 607 respondents from organizations in the U.S., Germany and U.K. Quotas were established for company size and for industries, to ensure the sample was good representation across industries and company sizes. Organizations were required to have developed AI or to intend to deploy AI within the next three years.

- Respondents were screened to:
- Be part of the organization's corporate leadership or report into corporate leadership roles
- Have a high level of involvement with at least one AI initiative
- Have one of the following roles when related to AI in their organizations:
  - Determine AI business objectives
  - Measure the value derived from AI initiatives
  - Manage AI initiative development and implementation

Results of this study do not represent global findings or the market as a whole but reflect sentiment of the respondents and companies surveyed.

Note 2: CRISP-DM

Cross-industry standard process for data mining, known as CRISP-DM, is an open standard process model that describes a structured approach to planning a data mining project. The model is an idealized sequence of events. Many of the tasks may be performed in a different order or repeated, and they might often require a feedback loop, but the tasks provide all possible routes to structure the data mining process.

Recommended by the Authors

A Guidance Framework for Operationalizing Machine Learning

Use Gartner's 3-Stage MLOps Framework to Successfully Operationalize Machine Learning Projects

Innovation Insight for ModelOps

Demystifying XOps: DataOps, ModelOps, Platform Ops for AI, AIOps

Table 1: AI Workflow

1. Business Requirements and Goal Definition	2. Data Collect, Process, Analysis and Features	3. Model Development, Evaluation, Pre-Prod Test	4. Model Deployment, Inference, Monitoring, Integration	5. Business Application and Adoption
■ Lack of Understanding AI Benefits and Uses ■ Lack of Operationalization Skills ■ Unable/Hard to Measure Business Value	■ Data Volume or Complexity ■ Data Scope or Quality Problems ■ Data Accessibility Challenges	■ Lack of Technology Knowledge ■ Lack of Capability to Leverage AI Techniques ■ Little Improvement Over Existing Technologies	■ Lack of Formal Operationalization Methodology ■ Communication Gaps ■ Unable to Deal With ModelOps to Ensure Reproducibility and Reusability	■ Lack of Business Stakeholder Involvement ■ Inconsistent Provisioning of Models ■ Failure to Track and Monitor Models

Source: Gartner