Chapter 8. Model Governance

Mark Treveil

We explored the idea of governance as a set of controls placed on a business in Chapter 3. These goals aim to ensure that the business delivers on its responsibilities to all stakeholders, from shareholders and employees to the public and national governments. The responsibilities include financial, legal, and ethical, and are all underpinned by the desire for fairness.

This chapter goes even more in depth on these topics, shifting from why they matter to how organizations can incorporate them as a part of their MLOps strategy.

Who Decides What Governance the Organization Needs?

National regulations are a key part of a society's framework for safeguarding fairness. But these take considerable time to be agreed upon and implemented; they always reflect a slightly historical understanding of fairness and the challenges to it. Just as with ML models, the past cannot always anticipate the evolving problems of the future.

What most businesses want from governance is to safeguard shareholder investment and to help ensure a suitable ROI, both now and in the future. That means the business has to perform effectively, profitability, and sustainably. The shareholders need clear visibility that customers, employees, and regulatory bodies are happy, and they want reassurances that appropriate measures are in place to detect and manage any difficulties that could occur in the future.

None of this is news, of course, nor specific to MLOps. What is different with ML is that it is a new and often opaque technology that carries many risks, but it is rapidly being embedded in decision-making systems that impact every aspect of our lives. ML systems invent their own statistically driven decision-making processes, often extremely difficult to understand, based on large volumes of data that is thought to represent the real world. It's not hard to see what could go wrong!

Perhaps the most surprising influence on the direction of ML governance is public opinion, which evolves much faster than formal regulation. It follows no formal process or etiquette. It doesn't have to be based on fact or reason. Public opinion determines what products people buy, where they invest their money, and what rules and regulations governments make. Public opinion decides what is fair and what is not.

For example, the agricultural biotechnology companies that developed genetically modified crops felt the power of public opinion painfully in the 1990s. While the arguments rage back and forth about whether there was, or was not, a risk to health, public opinion in Europe swung against genetic modification, and these crops were banned in many European countries. The parallels with ML are clear: ML offers benefits to all and yet brings risks that need to be managed if the public is to trust it. Without public trust, the benefits will not fully materialize.

The general public needs to be reassured that ML is fair. What is considered "fair" is not defined in a rule book, and it is not fixed; it will fluctuate based on events, and it will not always be the same across the world. Right now, opinion on ML is in the balance. Most people prefer getting sensibly targeted ads, they like their cars being able to read speed-limit signs, and improving fraud detection ultimately saves them money.

But there have also been well-publicized scandals that have rocked the public's acceptance of this technology. The Facebook-Cambridge Analytica affair, where the companies used the power of ML to manipulate public opinion on social media, shocked the world. This looked like ML with explicitly malicious intent. Equally worrying have been instances of entirely unintentional harm, where ML black box judgments proved to be unacceptably and illegally biased on criteria such as race or gender, for example in criminal assessment systems and in recruitment tools.

If businesses and governments want to reap the benefits of ML, they have to safeguard the public trust in it as well as proactively address the risks. For businesses, this means developing strong governance of their MLOps process. They must assess the risks, determine their own set of fairness values, and then implement the necessary process to manage them. Much of this is simply about good housekeeping with an added focus on mitigating the inherent risks of ML, addressing topics such as data provenance, transparency, bias, performance management, and reproducibility.

Matching Governance with Risk Level

Governance is not a free lunch; it takes effort, discipline, and time.

From the business stakeholders' perspective, governance is likely to slow down the delivery of new models, which may cost the business money. For data scientists, it can look like a lot of bureaucracy that erodes their ability to get things done. In contrast, those responsible for managing risk and the DevOps team managing deployment would argue that strict governance across the board should be mandatory.

Those responsible for MLOps must manage the inherent tension between different user profiles, striking a balance between getting the job done efficiently and protecting against all possible threats. This balance can be found by assessing the specific risk of each project and matching the governance process to that risk level. There are several dimensions to consider when assessing risk, including:

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

This assessment should not only determine the governance measures applied, but also drive the complete MLOps development and deployment tool chain.

For example, 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. Conversely, a model deployed to a public-facing website making decisions that impact people's lives or company finances requires a very thorough process. This process would 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.

But the business risk is not always so clear cut. An SSA project that makes a decision that has a long-term impact can also be high risk and can justify stronger governance measures. That's why across the board, teams need well thought out, regularly reviewed strategies for MLOps risk assessment (see Figure 8-1 for a breakdown of project criticality and operationalization approaches).

Project criticality	Operationalization	Builder autonomy	Versioning	Resources separation	SLA and 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 QA and scheduling	★	★★★	★★★	★	—
Operational projects that cannot suffer outages	Fully controlled deployment CI/CD	—	★★★	★★★	★★★	★★★

Figure 8-1. Choosing the right kind of operationalization model and MLOps features depending on the project's criticality

Current Regulations Driving MLOps Governance

There is little regulation around the world today specifically aimed at ML and AI. Many existing regulations do, however, have a significant impact on ML governance. These take two forms:

- Industry-specific regulation. This is particularly significant in the finance and pharmaceutical sectors.
- Broad-spectrum regulation, particularly addressing data privacy.

A few of the most pertinent regulations are outlined in the following sections. Their relevance to the challenges of MLOps governance is striking, and these regulations give a good indication of what governance measures will be needed broadly across the industry to establish and maintain trust in ML.

Even for those working in industries that don't have specific regulations, the following sections can give a brief idea of what organizations worldwide, regardless of industry, might face in the future in terms of the level of specificity of control with regards to machine learning.

Pharmaceutical Regulation in the US: GxP

GxP is a collection of quality guidelines (such as the Good Clinical Practice, or GCP, guidelines) and regulations established by the U.S. Food and Drug Administration (FDA), which aim to ensure that bio and pharmaceutical products are safe.

GxP's guidelines focus on:

- Traceability, or the ability to re-create the development history of a drug or medical device.
- Accountability, meaning who has contributed what to the development of a drug and when.
- Data Integrity (DI), or the reliability of data used in development and testing. This is based on the ALCOA principle: attributable, legible, contemporaneous, original, and accurate, and considerations include identifying risks and mitigation strategies.

Financial Model Risk Management Regulation

In finance, model risk is the risk of incurring losses when the models used for making decisions about tradable assets prove to be inaccurate.

These models, such as the Black–Scholes model, existed long before the arrival of ML.

Model risk management (MRM) regulation has been driven by the experience of the impact of extraordinary events, such as financial crashes, and the resulting harm to the public and the wider economy if severe losses are incurred. Since the financial crisis of 2007–2008, a large amount of additional regulation has been introduced to force good MRM practices (see Figure 8-2).

The UK Prudential Regulation Authority's (PRA) regulation, for example, defines four principles for good MRM:

Model definition

Define a model and record such models in inventory.

Risk governance

Establish model risk governance framework, policies, procedures, and controls.

Life cycle management

Create robust model development, implementation, and usage processes.

Effective challenge

Undertake appropriate model validation and independent review.

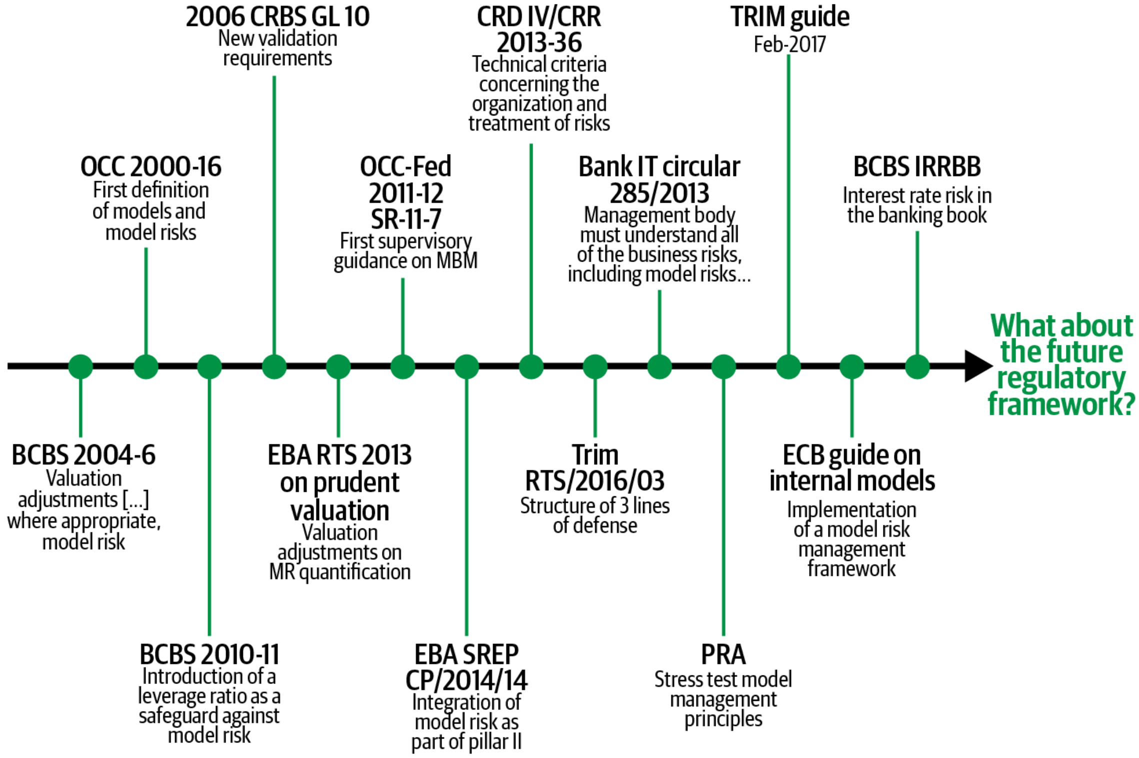

Figure 8-2. The history of model risk management (MRM) regulation

GDPR and CCPA Data Privacy Regulations

The EU General Data Protection Regulation (GDPR) was first implemented in 2018, setting guidelines for the collection and processing of personal information from individuals who live in the European Union. However, it was developed with the internet age in mind, so it actually applies for EU visitors to any website, regardless of where that website is based. Since few websites want to exclude EU visitors, sites across the world have been forced to meet the requirements, making GDPR a de facto standard for data protection. The regulations aim to give people control of their personal data that IT systems have collected, including the rights to:

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

The California Consumer Privacy Act (CCPA) is quite similar to GDPR in terms of who and what is protected, although the scope, territorial reach, and financial penalties are all more limited.

The New Wave of AI-Specific Regulations

Around the world, a new wave of regulations and guidelines specifically targeting AI applications (and thus all ML applications) is emerging. The European Union is leading the way with an attempt to establish a framework for trustworthy AI.

In a white paper on artificial intelligence, the EU emphasizes the potential benefits of AI for all walks of life. Equally, it highlights that scandals surrounding the misuse of AI and warnings of the dangers of potential advances in the power of AI have not gone unnoticed. The EU considers that regulatory framework based on its fundamental values "will enable it to become a global leader in innovation in the data economy and its applications."

The EU identifies seven key requirements that AI applications should respect to be considered trustworthy:

- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination, and fairness
- Societal and environmental well-being
- Accountability

The EU approach is not one-size-fits-all: it will primarily impact specific high-risk sectors, including healthcare, transportation, energy, and parts of the public sector. The regulations are expected to be optional for other sectors.

As with GDPR, the EU approach is likely to have a worldwide influence. It is also probable that many large organizations will decide to opt in considering the importance to their businesses of public trust in the use of AI. Even for those not opting in, the framework is likely to establish a way of thinking about governance in AI and will influence their approach.

Table 8-1 outlines some of the statuses of AI governance initiatives across the world. All are following an unmistakably similar route, even if the level of prescriptiveness reflects their traditionally distinct approaches to regulation.

Table 8-1. Status of AI governance initiatives across the world

Regions & organizations	Stage	Focus	Coming next
OECD	Guidance	• 42 signatories • 5 principles for responsible stewardship of trustworthy AI: inclusive growth, human-centered and fairness, transparency and explainability, robustness, and accountability • Recommendations for national policies	

Regions & organizations	Stage	Focus	Coming next
EU	Guidance, communication, direction, and regulation	• Binding for high-risk activities (Sector X impact), optional with possibility for label for others • Specifically targeting model fairness, robustness, and auditability, mixing policies and controls, integrating strong ethical considerations on environmental and social impacts	• Directive by end 2020/early 2021 • To be translated into national regime

Regions & organizations	Stage	Focus	Coming next
Singapore	Guidance	• Positive, nonsanctioned-based approach focusing on practical steps to implementation AI governance at an organization level • Best practice center, supporting AI governance work at Economic Forum level	• Regulation by end 2020/early 2021
US	Guidance, communication, and regulation	• Federal guidelines issued to prepare ground for industry-specific guidelines or regulation • Focus on public trust and fairness; no broader ethics considerations	
UK	Guidance	High-level guidelines only; nonbinding and broad in coverage	

Regions & organizations	Stage	Focus	Coming next
Australia	Guidance	Detailed guidelines issued, integrating ethical and a strong focus on end-consumer protection	

The Emergence of Responsible AI

As the adoption of data science, machine learning, and AI has accelerated worldwide, a loose consensus among AI thinkers has emerged. The most common banner for this consensus is Responsible AI: the idea of developing machine learning systems that are accountable, sustainable, and governable. In essence, AI systems should do what they are supposed to, remain reliable over time, and be well controlled as well as auditable.

There is no strict definition of Responsible AI or the terms used to frame it, but there is agreement about the overarching considerations and largely about what is needed to deliver it (see Table 8-2). Despite the lack of any single body driving the movement, Responsible AI has already had a significant influence on collective thinking, and especially on the EU's trustworthy AI regulators.

Table 8-2. Components of Responsible AI, an increasingly critical part of MLOps

Intentionality	**Accountability**
Must have: • Assurance that models are designed and behave in ways aligned with their purpose • Assurance that data used for AI projects comes from compliant and unbiased sources plus a collaborative approach to AI projects that ensures multiple checks and balances on potential model bias • Intentionality also includes explainability, meaning the result of AI systems should be explainable by humans (ideally not just the humans that created the system)	Must have: • Central control, management, and the ability to audit the enterprise AI effort (no shadow IT!) • An overall view of which teams are using what data, how, and in which models • Trust that data is reliable and being collected in accordance with regulation as well as a centralized understanding of which models are being used for which business process. This is closely tied to traceability—if something goes wrong, is it easy to find where in the pipeline it happened?

Human-centered approach

Providing people with the tools and training to be aware of and then execute on both components

Key Elements of Responsible AI

Responsible AI is about the responsibility of data practitioners, not about AI itself being responsible: this is a very important distinction. Another important distinction is that, according to Kurt Muemel of Dataiku, “It is not necessarily about intentional harm, but accidental harm.”

This section presents five key elements that figure in Responsible AI thinking—data, bias, inclusiveness, model management at scale, and governance—as well as MLOps considerations for each element.

Element 1: Data

The dependence on data is a fundamental differentiator between ML and traditional software development. The quality of the data used will make the biggest impact on the accuracy of the model. Some real-world considerations are as follows:

- Provenance is king. Understand how the data was collected and its journey to the point of use.
- Get the data off of desktops. Data must be manageable, securable, and traceable. Personal data must be strictly managed.
- The quality of data over time: consistency, completeness, and ownership.
- Bias in, bias out. Biased input data can occur easily and unintentionally.

Element 2: Bias

ML predictive modeling is about building a system to recognize and exploit tendencies in the real world. Certain types of cars, driven by certain types of people, in certain places are more likely to be costlier to insurance companies than others. But is matching a pattern always considered ethical? When is such pattern-matching proportionate, and when is it an unfair bias?

Establishing what is fair is not clear-cut. Even using a churn model to give rebates to the customers who are more likely to leave might be considered as unfair against dormant customers who will pay more for the same product. Regulations are a place to start looking, but as already discussed, opinion is not universal and is not fixed. Even with a clear understanding of the fairness constraints to work toward, achieving them is not simple. When the developers of the recruitment system that was biased against women's schools adapted the model to ignore the words like "women's," they found that even the tone of the language in a resume reflected the gender of the author and created unwanted bias against women. Addressing these biases has deep implications on the ML model to be built (see "Impact of Responsible AI on Modeling" for a detailed example).

Taking a step back, these bias problems are not new; for example, hiring discrimination has always been an issue. What is new is that, thanks to the IT revolution, data to assess biases is more available. On top of that, thanks to the automation of decision making with machine learning, it is possible to change the behavior without having to go through the filter of individuals making subjective decisions.

The bottom line is that biases are not only statistical. Bias checks should be integrated in governance frameworks so that issues are identified as early as possible, since they do have the potential to derail data science and machine learning projects.

It's not all bad news: there are many potential sources of statistical bias (i.e., of the world as it was) that *can* be addressed by data scientists:

- Is bias encoded into the training data? Is the raw material biased? Has data preparation, sampling, or splitting introduced bias?
- Is the problem framed properly?
- Do we have the right target for all subpopulations? Beware that many variables may be highly correlated.
- Is feedback-loop data biased through factors such as the order in which choices are presented in the UI?

It is so complex to prevent the problems caused by bias that much of the current focus is on detecting bias before it causes harm. ML interpretability is the current mainstay of bias detection, bringing understanding to ML models through a set of technical tools to analyze models including:

- Prediction understanding: Why did a model make a specific prediction?
- Subpopulation analysis: Is there bias among subpopulations?
- Dependency understanding: What contributions are individual features making?

A very different, but complementary, approach to addressing bias is to leverage as broad a range of human expertise as possible in the development process. This is one aspect of the idea of inclusiveness in Responsible AI.

Element 3: Inclusiveness

The human-in-the-loop (HITL) approach aims to combine the best of human intelligence with the best of machine intelligence. Machines are great at making smart decisions from vast datasets, whereas people are much better at making decisions with less information. Human judgment is particularly effective for making ethical and harm-related judgments.

This concept can be applied to the way models are used in production, but it can be equally important in the way models are built. Formalizing human responsibility in the MLOps loop, for example through sign-off processes, can be simple to do, but highly effective.

The principle of inclusiveness takes the idea of human-AI collaboration further: bringing as diverse a set of human expertise to the ML life cycle as possible reduces the risk of serious blind spots and omissions. The less inclusive the group building the ML, the greater the risk.

The perspectives of the business analyst, the subject matter expert, the data scientist, the data engineer, the risk manager, and the technical architect are all different. All of these perspectives together bring far greater clarity to managing model development and deployment than relying on any single user profile, and enabling these user profiles to collaborate effectively is a key factor in reducing risk and increasing the performance of MLOps in any organization. Refer to Chapter 2 for clear examples of collaboration among different profiles for better MLOps performance.

Full inclusiveness may even bring the consumer into the process, perhaps through focus group testing. The objective of inclusiveness is to bring the appropriate human expertise into the process, regardless of source. Leaving ML to data scientists is not the answer to managing risk.

Element 4: Model Management at Scale

Managing the risk associated with ML when there are a handful of models in production can afford to be largely manual. But as the volume of deployments grows, the challenges multiply rapidly. Here are some key considerations for managing ML at scale:

- A scalable model life cycle needs to be largely automated as well as streamlined.
- Errors, for example in a subset of a dataset, will propagate out rapidly and widely.
- Existing software engineering techniques can assist ML at scale.
- Decisions must be explainable, auditable, and traceable.
- Reproducibility is key to understanding what went wrong, who or what was responsible, and who should ensure it is corrected.
- Model performance will degrade over time: monitoring, drift management, retraining, and remodeling must be built into the process.
- Technology is evolving rapidly; an approach to integrating new technologies is required.

Element 5: Governance

Responsible AI sees strong governance as the key to achieving fairness and trustworthiness. The approach builds on traditional governance techniques:

- Determine intentions at the beginning of the process
- Formalize bringing humans in the loop
- Clearly identify responsibilities (Figure 8-3)

- Integrate goals that define and structure the process
- Establish and communicate a process and rules
- Define measurable metrics and monitor for deviation
- Build multiple checks into the MLOps pipeline aligned with overall goals
- Empower people through education
- Teach builders as well as decision makers how to prevent harm

Governance is, therefore, both the foundation and the glue of MLOps initiatives. However, it's important to recognize that it goes beyond the borders of traditional data governance.

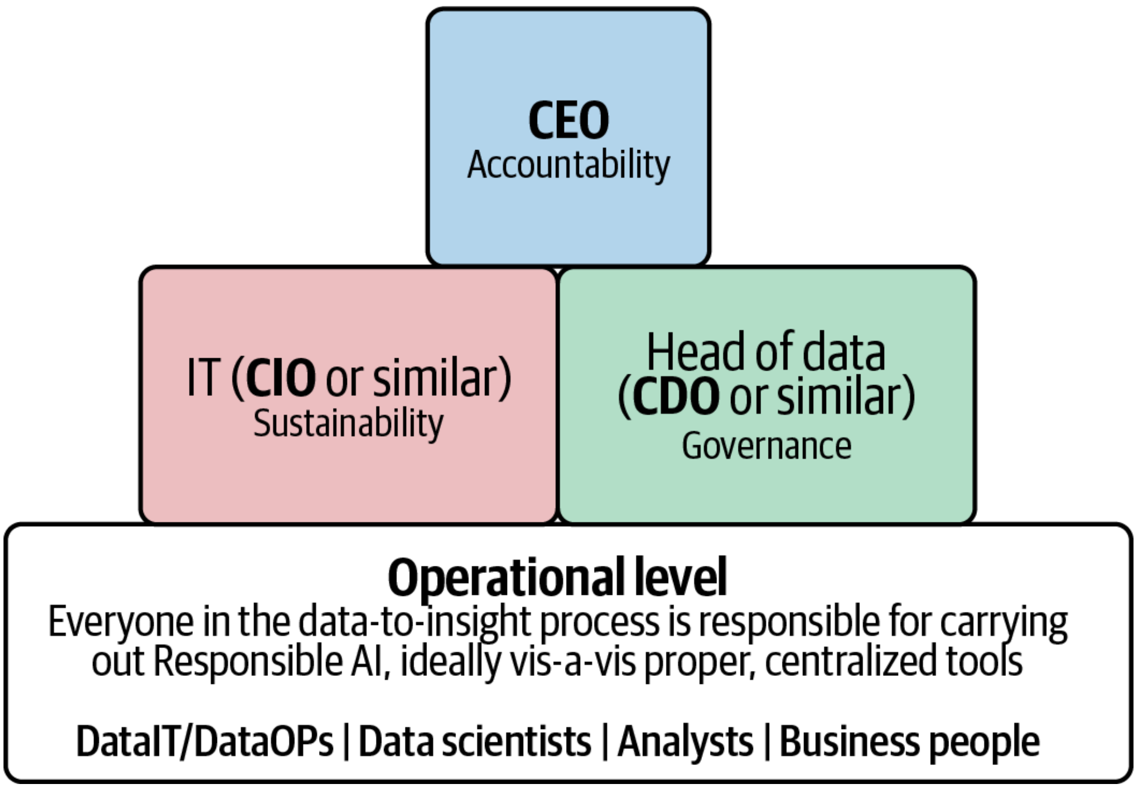

Figure 8-3. A representation of who is responsible at different levels of the organization for different parts of the Responsible AI process

A Template for MLOps Governance

Having explored the key themes to be addressed by an MLOps governance, both through regulatory measures and the Responsible AI movement, it is time to map out how to implement a robust governance framework of MLOps.

There is no one-size-fits-all solution across businesses, and different use cases within a business justify different levels of management, but the step-by-step approach outlined can be applied in any organization to guide the implementation process.

The process has eight steps:

1. Understand and classify the analytics use cases.
2. Establish an ethical position.
3. Establish responsibilities.
4. Determine governance policies.
5. Integrate policies into the MLOps process.
6. Select the tools for centralized governance management.
7. Engage and educate.
8. Monitor and refine.

This section will go through each of the steps in detail, including a simple definition and the "how" of actually implementing the step.

Step 1: Understand and Classify the Analytics Use Cases

This step entails defining what the different classes of analytics use cases are and, subsequently, what the governance needs are for each.

Consider the answers to the following questions for a representative cross-section of analytics use cases. Identify the key distinguishing features of the different use cases and categorize these features. Conflate categories where appropriate. Typically, it will be necessary to associate several categories to each use case to fully describe it.

- What regulations is each use case subject to, and what are the implications? Sector-specific regulations, regional, PII?
- Who consumes the results of the model? The public? One of many internal users?
- What are the availability requirements for the deployed model? 24/7 real-time scoring, scheduled batch scoring, ad-hoc runs (self-service

analytics)?
- What is the impact of any errors and deficiencies? Legal, financial, personal, public trust?
- What is the cadence and urgency of releases?
- What is the lifetime of the model and the lifetime of the impact of its decision?
- What is the likely rate of model quality decay?
- What is the need for explainability and transparency?

Step 2: Establish an Ethical Position

We established that fairness and ethical considerations are important motivating factors for effective governance, that businesses have a choice on their ethical stance, and that this impacts public perception and trust. The position a business takes is a trade-off between the cost to implement the position and public perception. Responsible stances rarely come at zero short-term financial cost even if the long-term ROI may be positive.

Any MLOps governance framework needs to reflect the ethical position of the company. While the position typically impacts what a model does and how it does it, the MLOps governance process needs to ensure that deployed models match the chosen ethical stance. This stance is likely to influence the governance process more widely, including the selection and verification of new models and the acceptable likelihood of accidental harm.

Consider the following ethical questions:

- What aspects of well-being in society matter? E.g., equality, privacy, human rights and dignity, employment, democracy, bias
- Is the potential impact on human psychology to be considered? E.g., human-human or human-AI relationships, deception, manipulation, exploitation
- Is a stance on the financial impact required? E.g., market manipulation
- How transparent should the decision making be?

- What level of accountability for AI-driven mistakes does the business want to accept?

Step 3: Establish Responsibilities

Identify the groups of people responsible for overseeing MLOps governance as well as their roles.

- Engage the whole organization, across departments, from top to bottom of the management hierarchy.
- Peter Drucker's famous line "Culture eats strategy for breakfast" highlights the power of broad engagement and shared beliefs.
- Avoid creating all-new governance structures. Look at what structures exist already and try to incorporate MLOps governance into them.
- Get senior management sponsorship for the governance process.
- Think in terms of separate levels of responsibility:
  - Strategic: set out the vision
  - Tactical: implement and enforce the vision
  - Operational: execute on a daily basis
- Consider building a RACI matrix for the complete MLOps process (see Figure 8-4). RACI stands for *responsible, accountable, consulted, informed*, and it highlights the roles of different stakeholders in the overall MLOps process. It is quite likely that any matrix you create at this stage will need to be refined later on in the process.

Tasks	Business stakeholders	Business analysis/ citizen DS	Data scientists	Risk/ audit	Data ops	Production/ exploitation	Resources admin/ architect
Identification	A/R	C		I			
Data preparation	C	A/R	C				
Data modeling	C	A	R				
Model acceptance	I	C	C	A/R			
Productionalization		C	A/R	I	C		
Capitalization			R		R		A
Integration to external systems					A/R		
Global orchestration		C			R	A	
User acceptance tests	A/R	R	C		I		
Deployments					R	A	I
Monitoring	I	C				A/R	I

A: accountable **R: responsible** **C: consulted** **I: informed**

Figure 8-4. A typical RACI matrix for MLOps

Step 4: Determine Governance Policies

With an understanding of the scope and objectives for governance now established, and the engagement of the responsible governance leaders, it is time to consider the core policies for the MLOps process. This is no small task, and it is unlikely to be achieved in one iteration. Focus on establishing the broad areas of policy and accept that experience will help to evolve the details.

Consider the classification of initiatives from Step 1. What governance measures do the team or organization need in each case?

In initiatives where there is less concern about the risk or regulatory compliance, lighter-weight, cheaper measures may be appropriate. For example, "what if" calculations to determine the number of in-flight meals of different types has relatively little impact—after all, the mix was never right even before the introduction of machine learning. Even such a seemingly insignificant use case may have ethical implications as meal choices are likely to be correlated to religion or gender, which are protected attributes in many countries. On the other hand, the implications of calculations to determine the level of fueling of planes carry substantially greater risk.

Governance considerations can be broadly grouped under the headings in Table 8-3. For each heading, there are a range of measures to consider for each class.

Table 8-3. MLOps governance considerations

Governance consideration	Example measures
Reproducibility and traceability	Full VM and data snapshot for precise and rapid model re-instantiation, *or* ability to re-create the environment and retrain with a data sample, *or* only record metrics of models deployed?
Audit and documentation	Full log of all changes during development including experiments run and reasons for choices made *or* automated documentation of deployed model only *or* no documentation at all
Human-in-the-loop sign-off	Multiple sign-offs for every environment move (development, QA, preproduction, production)
Preproduction verification	Verify model documentation by hand-coding the model and comparing results *or* full automated test pipeline re-creating in production-like environment with extensive unit and end-to-end test cases *or* automated checks on database, software version, and naming standards only
Transparency and explainability	Use manually-coded decision tree for maximum explainability *or* use regression algorithms' explainability tools such as Shapely values *or* accept opaque algorithms such as neural networks
Bias and harm testing	"Red team" adversarial manual testing using multiple tools and attack vectors *or* automated bias checking on specific subpopulations

Governance consideration	Example measures
Production deployment modes	Containerized deployment to elastic scalable high-availability, multi-node configuration with automated stress/load testing prior to deployment *or* a single production server
Production monitoring	Real-time alerting of errors, dynamic multi-armed bandit model balancing, automated nightly retraining, model evaluation, and redeployment *or* weekly input drift monitoring and manual retraining *or* basic infrastructure alerts, no monitoring, no feedback-based retraining
Data quality and compliance	PII considerations including anonymization *and* documented and reviewed column-level lineage to understand the source, quality, and appropriateness of the data *and* automated data quality checks for anomalies

The finalized governance policies should provide:

- A process for determining the classification of any analytics initiative. This could be implemented as a checklist or a risk assessment application.
- A matrix of initiative classification against governance consideration, where each cell identifies the measures required.

Step 5: Integrate Policies into the MLOps Process

Once the governance policies for the different classes of initiatives have been identified, measures to implement them need to be incorporated into the MLOps process and responsibilities for actioning the measures assigned.

While most businesses will have an existing MLOps process, it is quite likely that this has not been defined explicitly, but rather has evolved in response to individual needs. Now is the time to revisit, enhance, and document the process. Successful adoption of the governance process can only happen if it is communicated clearly and buy-in is sought from each stakeholder group.

Understand all of the steps in the existing process by interviewing those responsible. Where there is no existing formal process, this is often harder than it sounds because the process steps are often not explicitly defined, and ownership is unclear.

Attempting to map the policy-driven governance measures into the understanding of the process will identify issues in the process very quickly. Within one business there may be a range of different styles of project and governance needs, such as:

- One-off self-service analytics
- Internally consumed models
- Models embedded in public websites
- Models deployed to Internet of Things devices

In these cases, the differences between some processes may be so great it is best to think in terms of several parallel processes. Ultimately, every governance measure for each use case should be associated with a process step and with a team that is ultimately responsible, as presented here:

Process step	Example activities and governance considerations
Business scoping	Record objectives, define KPIs, and record sign-off: for internal governance considerations
Ideation	Data discovery: data quality and regulatory compliance constraints Algorithm choice: impacted by explainability requirements
Development	Data preparation: consider PII compliance, separation of legal regional scopes, avoid input bias Model development: consider model reproducibility and auditabilityModel testing and verification: bias and harm testing, explainability
Preproduction	Verify performance/bias with production data Production-ready testing: verify scalability
Deployment	Deployment strategy: driven by the level of operationalization Deployment verification testsUse of shadow challenger or A/B test techniques for in-production verification
Monitoring and feedback	Performance metrics and alerting Prediction log analysis for input drift with alerting

Step 6: Select the Tools for Centralized Governance Management

The MLOps governance process impacts both the complete ML life cycle and many teams across the organization. Each step requires a specific sequence of actions and checks to be executed. Traceability of both the development of the model and the execution of governance activities is a complex challenge.

Most organizations still have a "paper form" mindset for process management, where forms are filled in, circulated, signed, and filed. The forms may be text documents, circulated via email, and filed electronically, but the limitations of paper remain. It is hard to track progress, associate artifacts, review many projects at once, prompt for action, and remind teams of responsibilities. The complete record of events is typically spread across multiple systems and owned by individual teams, making a simple overview of analytics projects effectively impossible.

While teams will always have tools specific to their roles, MLOps governance is much more effective if the overarching process is managed and tracked from one system. This system should:

- Centralize the definition of the governance process flows for each class of analytics use cases
- Enable tracking and enforcement of the complete governance process
- Provide a single point of reference for the discovery of analytics projects
- Enable collaboration between teams, in particular, the transfer of work between teams
- Integrate with existing tools used for project execution

The current workflow, project management, and MLOps tools can only partially support these objectives. A new category of ML governance tools is emerging to support this need directly and more fully. These new tools focus on the specific challenges of ML governance, including:

- A single view of the status of all models (otherwise known as a model registry)
- Process gates with a sign-off mechanism to allow ready traceability of the history of decision making
- Ability to track all versions of a model
- Ability to link to artifact stores, metrics snapshots, and documentation
- Ability to tailor processes specifically for each class of analytics use cases

- Ability to integrate health monitoring from production systems and to track the performance of models against the original business KPIs

Step 7: Engage and Educate

Without a program of engagement and training for the groups involved in overseeing and executing the governance process, the chances of it being even partially adopted are slim. It is essential that the importance of MLOps governance to the business, and the necessity of each team's contribution, is communicated. Building on this understanding, every individual needs to learn what they must do, when, and how. This exercise will require considerable documentation, training, and, most of all, time.

Start by communicating the broad vision for MLOps governance in the business. Highlight the dangers of the status quo, outline a process, and detail how it is tailored to the range of use cases.

Engage directly with each team involved and build a training program directly with them. Do not be afraid to leverage their experience to shape not only the training, but also the detailed implementation of their governance responsibilities. The result will be much stronger buy-in and more effective governance.

Step 8: Monitor and Refine

Is the newly implemented governance working? Are the prescribed steps being implemented, and are the objectives being met? What actions should be taken if things are going poorly? How do we measure the gap between today's reality and where the business needs to be?

Measuring success requires metrics and checks. It requires people to be tasked with monitoring and a way to address problems. The governance process and the way it is implemented will need to be refined over time, based both on lessons learned and on evolving requirements (including, as discussed earlier in this chapter, evolving regulatory requirements).

A big factor in the success of the process will be the diligence of the individuals responsible for the individual measures in the process, and incentivizing them is key.

Monitoring the governance process starts with a clear understanding of the key performance metrics and targets—KPIs for governance. These should aim to measure both whether the process is being enacted and whether objectives are being achieved. Monitoring and auditing can be time consuming, so look to automate metrics as far as possible and encourage individual teams to own the monitoring of metrics that relate to their area of responsibility.

It’s hard to make people carry out tasks that seemingly deliver nothing concrete to those doing the work. One popular tactic to address this is gamification. This is not about making everything look like a video game, but about introducing incentives for people to carry out tasks where the main benefit is derived by others.

Look to gamify the governance process in simple ways: publishing KPI results widely is the simplest place to start. Just being able to see targets being met is a source of satisfaction and motivation. Leaderboards, whether at the team or individual level, can add some constructive element of competition. For example, people whose work consistently passes compliance checks the first time, or meets deadlines, should be able to feel their efforts are visible.

However, excessive competition can be disruptive and demotivating. A balance must be struck, and this is best achieved by building up gamification elements slowly over time. Start with the least competition-oriented and add new elements one by one, measuring their effectiveness before adding the next.

Monitoring changes in the governance landscape is essential. This might be regulatory, or it might be about public opinion. Those with responsibility for the strategic vision must continue to monitor it as well as have a process to evaluate potential changes.

Finally, all monitoring of the process is only worthwhile if issues are acted upon. Establish a process for agreeing on change and for enacting it. This may result in revisiting policies, processes, tools, responsibilities, education, and monitoring! It's necessary to iterate and refine, but the balance between efficiency and effectiveness is hard to find; many lessons can only be learned the hard way. Build a culture where people see iteration and refinement as a measure of a successful process, not a failed one.

Closing Thoughts

It's hard to separate MLOps from its governance. It's not possible to successfully manage the model life cycle, mitigate the risks, and deliver value at scale without governance. Governance impacts everything from how the business can acceptably exploit ML, to the data and algorithms that can be used, to the style of operationalization, monitoring, and retraining.

MLOps at scale is in its infancy. Few businesses are doing it, and even fewer are doing it well. While governance is the key to improving the effectiveness of MLOps, there are few tools today that directly address this challenge, and there is only piecemeal advice.

Public trust in ML is at risk. Even slow-moving organizations like the EU understand this. If trust is lost, then so too will be many of the benefits to be derived from ML. Additional legislation is being prepared, but even without this, businesses need to worry about the potential damage to their public image that can be caused by an inadvertently harmful model.

When planning to scale MLOps, start with governance and use it to drive the process. Don't bolt it on at the end. Think through the policies; think about using tooling to give a centralized view; engage across the organization. It will take time and iteration, but ultimately the business will be able to look back and be proud that it took its responsibilities seriously.