

BEYOND THE ALGORITHM

When Models Met Governance — Why Machine Learning in Production Exposed Every Fault Line in Organisational Oversight

Giovanni Leonardi

October 2019

A Practitioner's Perspective on the Collision Between Model Deployment and Enterprise Governance in Financial Services and Beyond



When Models Met Governance — Why Machine Learning in Production Exposed Every Fault Line in Organisational Oversight

Transformation · No. 287

Originally written October 2019 | Re-edited and re-rendered for the WhereBeyond Edition, 2025

“The organisation that could train a model in a week discovered it could not explain one in a year.”

1. Executive Summary

The rush to deploy machine learning models into production environments has exposed a structural failure that few organisations anticipated: the governance frameworks built over decades to oversee technology change were never designed for systems that learn. In my experience across financial services, insurance, and telecommunications during 2018 and 2019, the pattern is remarkably consistent. Organisations that had invested heavily in model development capability found themselves unable to move those models into production — not because the models did not work, but because nobody could answer the questions that production deployment demanded. Who owns a model that changes its own parameters? How do you audit a decision that no single person made? What does change management mean when the system's behaviour shifts without anyone deploying new code?

This essay examines why that collision between machine learning ambition and governance reality has become one of the defining challenges of this period, and what it reveals about the deeper structural tensions that organisations must resolve if they are to move from experimentation to genuine operational capability.

2. The Promise That Outran the Plumbing

The trajectory of machine learning adoption over the past two years has followed a pattern familiar to anyone who has watched technology waves move through large organisations. The initial excitement — fuelled by visible successes in image recognition, natural language processing, and recommendation engines — created a surge of investment in data science teams, tooling, and infrastructure. By mid-2018, most large financial institutions had established some form of data science or advanced analytics function. Many had dozens of models in various stages of development.

What they did not have was any coherent way to move those models from a data scientist's notebook into a production system that served real customers and made real

decisions.

The gap was not primarily technical, although technical challenges certainly existed. The gap was organisational. The entire apparatus of change management, risk assessment, model validation, and operational oversight had been built for a world in which software behaved deterministically. A traditional application does the same thing every time it encounters the same input. Its behaviour can be specified, tested, documented, and audited against that specification. When it changes, someone changes it — deliberately, through a controlled release process.

Machine learning models violate every one of those assumptions. They are probabilistic rather than deterministic. Their behaviour is derived from data rather than specified by a developer. They can drift — changing their effective behaviour as the data they encounter in production diverges from the data on which they were trained. And the relationship between their inputs and outputs is, in many architectures, opaque even to the people who built them.

3. The Governance Vacuum

The pattern I have observed across multiple organisations is strikingly consistent. A data science team builds a model that demonstrates genuine predictive power. It outperforms the existing rule-based system in testing. The business sponsor is enthusiastic. The technology team is ready to integrate it. And then the model enters the governance process — and stops.

The first question that surfaces is ownership. In a traditional application, ownership is relatively clear: a product owner specifies the requirements, a development team builds to those requirements, and an operations team runs the result. But a machine learning model confounds this structure. The data science team trained it, but they do not own the data it learned from. The business owns the decision it supports, but they did not specify how it should make that decision — the model learned that from historical patterns. The technology team will host it, but they cannot explain what it does in the way they can explain a rules engine.

The fundamental governance challenge of machine learning is not that models are complex. It is that they blur every line of accountability that traditional governance depends on — between specification and learning, between deployment and drift, between a decision and the data that shaped it.

The second question is validation. Traditional model risk management, particularly in financial services, rests on independent validation — a second team reviews the model's methodology, tests its assumptions, and confirms it is fit for purpose. But the validators were trained to assess statistical models with explicit assumptions and transparent parameters. Confronted with a gradient-boosted ensemble or a neural network, the existing validation frameworks strain. The validators can test the model's outputs, but they struggle to assess its methodology in the way they would assess a logistic regression or a Monte Carlo simulation.

The third question is monitoring. Even if a model passes validation at a point in time, its behaviour in production is not static. Data drift, concept drift, and distributional shift can degrade a model's performance without any change to its code. The traditional approach to production monitoring — alerting on system health metrics like uptime, latency, and error rates — tells you nothing about whether a model's predictions are still accurate. And yet most organisations in 2019 have no operational framework for monitoring model performance in production.

4. Three Structural Forces That Sustain the Problem

The persistence of this governance vacuum is not accidental. It is sustained by at least three structural forces that operate across organisations regardless of sector.

The first is the separation of data science from engineering and operations. In most organisations, data science teams were established as centres of excellence or innovation labs, deliberately positioned outside the mainstream technology function. This made sense as an incubation strategy — it gave data scientists freedom to experiment without the constraints of enterprise architecture and release management. But it also meant that data science developed its own tools, its own workflows, and its own culture, largely disconnected from the engineering practices that govern production systems. When models are ready for production, this separation becomes a chasm. The data scientist's Jupyter notebook does not translate into a production-grade service. The experiment-tracking tools the team uses are invisible to the enterprise monitoring stack. The feature engineering pipelines that feed the model are not integrated into the organisation's data platform.

The second is the regulatory lag. Regulators in financial services and other heavily supervised sectors have been aware of the growing use of machine learning, but regulatory guidance has consistently trailed practice. The Prudential Regulation Authority's expectations around model risk management, for example, were written for an era of simpler statistical models. The Senior Managers and Certification Regime establishes personal accountability for decisions, but does not address the specific challenges of decisions mediated by opaque models. Organisations find themselves in an uncomfortable position: they know they need governance, but the regulatory frameworks they would normally lean on for structure have not yet caught up.

The third is the incentive asymmetry between building and governing. Data science teams are measured and rewarded for building models — for demonstrating predictive power, for delivering proof-of-concept results, for winning internal competitions for funding. There is no equivalent incentive for making a model governable. Documentation, explainability, monitoring hooks, and validation-ready packaging are unglamorous work. In the absence of explicit organisational demand for these things, they are systematically deprioritised.

5. What This Reveals About Transformation

The machine learning governance challenge is, at its root, a transformation problem — and it is a revealing one, because it exposes a pattern that recurs far beyond data science.

Organisations consistently invest in capability before they invest in the conditions that allow that capability to operate. They build data science teams before they build the data platforms those teams need. They train models before they build the infrastructure to serve them. They pursue algorithmic decision-making before they resolve the accountability frameworks that algorithmic decisions require.

This is not a failure of planning so much as a failure of imagination. The leaders who sponsor these programmes can see the value of the capability — the better predictions, the automated decisions, the competitive advantage. What they cannot easily see is the invisible infrastructure of governance, integration, and operational readiness that turns a capability into a reliable, accountable, scalable part of the business.

*“The organisation that could train a model in a week
discovered it could not explain one in a year.”*

The pattern is identical to what many organisations experienced with earlier waves of technology adoption — the shift to service-oriented architecture, the adoption of cloud infrastructure, the move to agile delivery. In each case, the new capability arrived faster than the organisational structures needed to absorb it. And in each case, the resolution required not just new tools or new processes, but a fundamental renegotiation of roles, responsibilities, and ways of working.

6. The Emerging Response

Across the organisations I have observed, a set of responses is beginning to crystallise, though none has yet reached maturity.

The most promising development is the emergence of what some are calling MLOps — an attempt to apply the principles of DevOps to machine learning workflows. The core insight of MLOps is that a model is not a one-time artefact but a continuously maintained system. It needs version control — not just of code, but of data, features, and parameters. It needs automated testing — not just unit tests, but tests of model performance against held-out data. It needs continuous monitoring — not just of system health, but of prediction quality, data drift, and fairness metrics. And it needs a deployment pipeline that treats model updates with the same rigour as code releases.

MLOps is still nascent. The tooling is fragmented, the practices are not yet standardised, and most organisations are building bespoke solutions rather than adopting established frameworks. But the direction is clear, and it represents a genuine attempt to bridge the gap between data science experimentation and production engineering.

Alongside MLOps, some organisations are beginning to rethink their model risk management frameworks. The most thoughtful approaches I have seen do not simply

extend existing validation processes to cover machine learning. Instead, they ask a more fundamental question: what does it mean to validate a system whose behaviour is derived from data rather than specified by a human? The answers tend to involve a shift from point-in-time validation to continuous monitoring, from methodology review to outcome testing, and from individual model assessment to portfolio-level risk management.

7. The Deeper Lesson

The collision between machine learning and governance is not, ultimately, a story about technology. It is a story about how organisations absorb novelty — and how consistently they underestimate the organisational change that technological capability demands.

The organisations that will navigate this transition most effectively will not be those with the most sophisticated models or the largest data science teams. They will be those that recognise, early enough, that deploying a model into production is an organisational act, not a technical one. It requires clear ownership, robust validation, continuous monitoring, and — perhaps most importantly — a willingness to slow down the rush to production long enough to build the governance infrastructure that sustainable deployment demands.

In my experience, this is the hardest lesson for transformation leaders to absorb. The pressure to demonstrate value, to justify investment, to keep pace with competitors who claim to be further ahead — all of these push towards speed. But speed without governance is not progress. It is risk accumulation.

The organisations that get this right will be those that treat governance not as a constraint on innovation, but as the infrastructure that makes innovation sustainable. That shift — from governance as burden to governance as enabler — is the real transformation that machine learning demands.