Deep Science Publishing
https://doi.org/10.70593/978-93-49910-19-5_9

DeepScience
Open Access Books

Chapter 9: Ensuring data governance, lineage, explainability, and auditability in artificial intelligence-driven financial models

9.1. Introduction

Advances in AI have offered unprecedented capability for tackling complex decision environments that the financial services industry faces today. Breakthrough achievements in deep learning and natural language processing, among other areas, are enabling new solutions that could not previously have been imagined, let alone attempted. The impetus to grow innovative product and service offerings based on complex AI-driven models, however, underscores an equally pressing need to ensure regulatory safety and soundness, risk and control management, and, above all, the trust of stakeholders. And while traditional statistical and empirical methods lack sufficient predictive power for the complexity and volume of financial services activity, customers, investors, and regulators may not necessarily trust new AI-driven models as much as traditional credit scoring models. For them, there is a huge risk in failing to disclose decision-making processes in fintech platforms that use machine learning to assist in decision-making for areas such as algorithmic trading, credit-risk assessment, and fraud detection (Barredo Arrieta et al., 2020; Sicular & Beyer, 2020; Jain & Aggarwal, 2021).

The onus is on organizations using AI-driven models to approach product and service innovation with a commensurate level of rigour and caution. Data-driven decision-making in financial services is a two-sided coin. At one end, financial services-related academics, managers and practitioners have called for deeper integration of AI-driven models into existing competency domains such as model validation, model risk management, investment strategy, etc. On the other end, regulators have voiced concern that tech companies are, or soon will be, offering financial services activities without sufficient controls or expert regulatory oversight to safeguard investor trust. This paper aims to offer a unique contribution on the second, skeptical, side of the discussion coin. In doing so, it provides a much-needed foundation for the other, pro-innovation side, to vet the feasibility of such models for various fintech and financial market applications (Veale & Edwards, 2018; Weller, 2019).

9.1.1. Purpose and Objectives of the Study

The objective of this work is to explore the many aspects who are related to the conceived pillars of trustworthy AI, and to explain why they are so relevant, focusing particularly in AI-driven financial models. Data procedures for modeling, prediction, and decision-making in AI systems in finance allow human beings and technical systems to interact closely with each other, where output results from this interaction can have great implications. Thus, ways need to be established for having governance regarding the conducted tasks, facilitating ways to explain the applicability of model outcomes when making predictions, and available structures for auditing model behavior. Procedures regarding data underpinning the creation and construction of AI-driven financial models need to be developed and disseminated on how to govern the models' entire life-cycle, from planning, collecting, and labeling the data, during training and testing, and during execution – including supervision, monitoring, retraining, testing, and redeployment processes – maintaining explicit information regarding the models' data lineage, ensuring trust and reliability, and thus ensuring explainable and auditable AI.

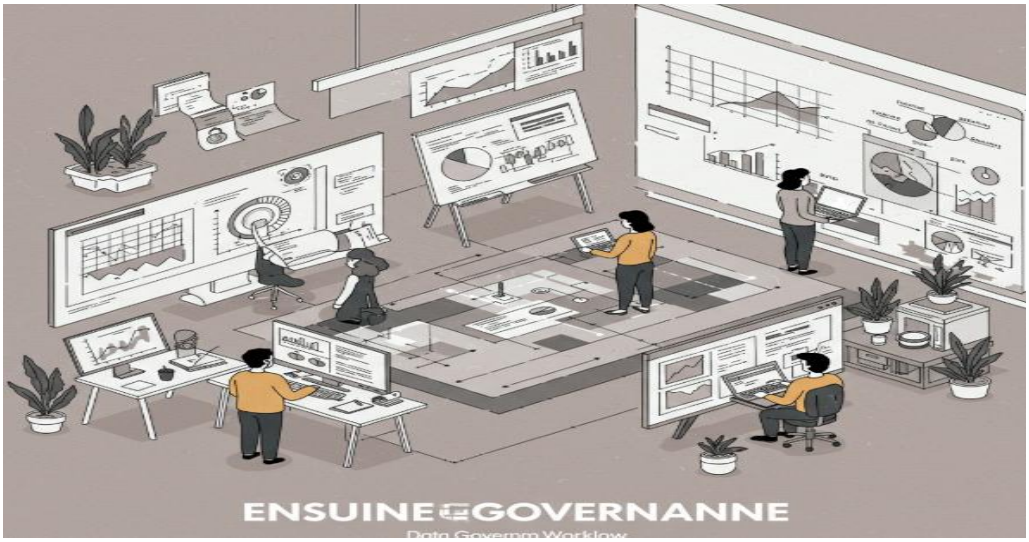

Fig 9 . 1 : Ensuring Data Governance

In what concerns the increasingly notoriety of AI in the financial services industry, we argue that superior performance from humans, systems or both can only be expected on harnessed AI by historically known principles of data governance, considering ethical and legal principles regarding data aspects, model design, and deployment. Adopted ideas can support financial institutions when handling the new risks and challenges they are facing due to the use of AI, being in accordance with discussions, tools, and frameworks provided by regulators around the globe. By briefly describing the pillars of trustworthy AI, we present a framework on how to ensure these principles when developing and deploying AI-driven financial services products. We illustrate our ideas studying a portfolio investment use case.

9.2. Overview of AI in Financial Models

We are living through a period of great change in financial services, wrought by a perfect storm of constituent factors including increased public scrutiny, technological advancement, new business entrants, and shifting geopolitical relations. The industry is transitioning to a new economic model that prioritizes growth policies including investing in resources to pursue long-term growth, a customer-centric approach innovating for enhanced customer experience, and productivity reform optimizing cost-to-income ratios. These factors, coupled with the industry's considerable investments in digitalization, are converging to uplift productivity growth.

AI is fundamental to these developments. Addressing political and social pressures requires a more determinative reliance on advanced quantitative models for risk management, governance, and regulation. Data-driven decision-making is at the foundation of virtually all the key trends across the financial industry, and fair, explainable, and robust AI is essential to effective and prudent data governance. The increasing irrelevance of past financial data, coupled with the use of digital media for financial services, requires the development of new approaches to modeling key dynamics. Here, AI's ability to efficiently extract and process information from multiple, multi-modal, and heterogeneous data sources is vital and urgently required.

9.2.1. Key Trends Shaping AI in Financial Modeling

Abstract. The popularity and accessibility of large generative foundation models are giving rise to a new wave of automation in knowledge work. At the same time, these models present a new plethora of capabilities, as well as risks tied to their misuse, that all organizations must evaluate and address. This chapter focuses on trends transforming financial modeling systems. We discuss the historical, technological, and market forces shaping the trajectory of AI in finance, including accessible foundation models, how the current wave of LLMs are different and advantageous, what high-priority use cases and capabilities organizations are pursuing with LLMs, workforce and job requirements

evolving with the responsible integration of LLMs, and the macro-financial landscape altering the context for AI initiatives. We conclude with actionable recommendations for leveraging LLMs in finance organizations. There are key macro trends reshaping how financial models are created, used, and maintained. The first is the growing number of junior workers in the finance and investing ecosystem tasked with the responsibility of modeling. Entry-level jobs on finance teams are being filled by increasingly younger talent with less experience as the finance industry faces widespread layoffs from mass market and cyclical shocks. Additionally, there is a growing shift of investment dollars towards smaller upstart hedge funds focused on niche asset classes and new ideas. Many of these new funds are small teams, often with a singular person tasked with responsibilities that entailed a team of modelers only a decade prior. AI can provide leverage to all these roles as an intelligent assistant in model build and model review. It can increase efficiency and productivity for the domain experts who have already spent years mastering their craft, while also providing guidance and structure to the new or less experienced workers who have less time in the trenches.

9.3. Importance of Data Governance

Data governance refers to the overall management of availability, usability, integrity, and security of data used in an organization. It aims to ensure that data is consistent and trustworthy and does not get misused. Data quality can affect model outcomes and business results in dramatic and unpredictable ways. Developing and implementing an appropriate data governance process is, therefore, essential to minimize failure risk. Data governance helps to define who is accountable for various aspects of data. This includes responsibility for making sure research data is accurate, accessible, reusable, and protected. This is especially true when products are intended for commercial use because an algorithm for commercial product use could cause loss of user’s financial assets if the output of such an algorithm is not quantitatively precise. Also, there could be legal, regulatory, social, or commercial repercussions for the company if data for commercial use is not well governed and managed appropriately. The aim of data governance is to ensure that the data can be trusted, is accurately represented, avoids duplication, is not stored when no longer required, is properly classified, is protected and kept secure, has a documented capability, and is consistent with the purpose of data capture.

For the financial sector, the data governance policy must align with the goals and objectives set by the Board of Directors. The policy must also align with regulatory laws that enforce data governance requirements. Implementing a data governance policy helps the organizations by safeguarding sensitive data, protecting revenue and shareholder value, protecting the organization against direct and indirect damages, granting maximum visibility, and reducing collateral risks. However, a data governance policy itself does not guarantee the success of the data governance program. To be successful, a policy must be implemented with the understanding and cooperation of the stakeholders. A strategic business approach must be put into practice.

9.3.1. Definition and Scope

Data governance primarily focuses on two issues: first, the availability of data for people to do their jobs properly, and second, the assurance of good data quality. These two issues are critical not only to success in technology-driven businesses, but also directly tied to the bottom line and shareholder value. Data governance frequently has risk mitigation at its center, especially with regard to data-related regulations. Over time, we have witnessed several financial services companies suffer vast penalties as a result of both data breaches and poor data quality. Enacting a data governance model, especially in a federated model, opens the door to avoiding such breaches and costs.

In its simplest understanding, data governance refers to the lines of accountability and decision-making rights over data. Well-governed data helps drive execution and decision-making at key levels-ensuring the right data is available, is of the right quality, and is being used properly. Data governance provides market regulators and the financial services industry with the assurance that risks are appropriately analyzed, capitalized, and controlled in the day-to-day running of financial services organizations. More technically detailed, data governance comprises a framework of decision-making rights and accountability for data-related decisions. It is supported by the structure of the enterprise, the enterprise policies, and the data standards. Data governance helps data owners and stewards throughout the enterprise fulfill their data-related responsibilities and elevate data as an enterprise asset.

9.3.2. Key Principles of Data Governance

The key principles of data governance are based on the validity of the conceptualized relationships and logical hierarchies that these structures outline for an organization. A dataset has logical relationships that are semantic in nature; key identifiers in record sets can be used to relate records about the same instance. Logical relationships are dictated by the rules of the domain of knowledge being modeled, such as rules about valid customer addresses in business operations. The illustration of logical relationships depicting their hierarchies is one of the most powerful uses of data modeling, helping both business and technical stakeholders understand the granularity to which data attributes are defined, duplicate attributes, and the relationships on which data validity is predicated. Without a clear logical map, which is detailed and regularly curated by experts, data is added willy-nilly, in isolation, and no one is the wiser until data use discovers a data issue, requiring significant business remediation to sort out what happened.

Business decisions about data should be driven by logical hierarchies that dictate what constitutes higher-level aggregates and lower-level breakdowns of data. Thus, operational data is only usable for strategic decisions if its logically defined hierarchies agree with how the business thinks about reporting and analysis. This same principle applies to metadata hierarchies, which dictate parent-child relationships regarding how metadata are defined and thus how they will be stored. Poorly defined or overlooked metadata stored at a higher level of detail than that at which a user expects to ingest, stored in a different way, or lacking significant attributes to identify user need can obfuscate data, rendering it difficult to use for subsequent users. Without a detailed map of metadata definitions, users often resort to help from data owners, who may no longer be available.

9.4. Data Lineage in AI Models

The ability to understand the evolution of data transformations allows model users to ascertain how input transformations impact outputs, especially when inputs or environmental settings for the model are changed. Implementing model capabilities in a way that accurately tracks what changes are made, and by whom, to what data values enables auditability of decision-making. This collaborative and audited versioning of data transformations is sometimes referred to as data lineage. Data lineage is a critical feature for both governance and operation. Specifically, it enables models deployed in critical settings to be held accountable for their decision-making.

Transparency about data evolution lowers the chances that models, particularly those that are meant to work for the public good, are intentionally or unintentionally manipulated to provide biased or inaccurate outputs. It allows users to see which changes to data values and provenance are correlated with changes to the model outputs. This is especially important in fields such as credit risk and assessment, where model outputs can carry significant ramifications for people. Users can monitor the versioning of reference datasets, along with the specific field changes made to input datasets, and associate these changes with transformations of model outputs. They can evaluate whether the changes to data correspond to decision-critical changes in output.

While data lineage of input data is critical, it is only part of the picture. Users must also be able to view dependencies from the output data back to the input data, the decision logic, and reference data. Tracking upstream dependencies will fill out the lineage for input data to include the parameter values used in algorithmic decision-making conditioned on input data, as well as any parameter dependencies on decision-critical input features. Tracking downstream dependencies of those decisions will allow users to associate business processes with those decisions.

9.4.1. Understanding Data Lineage

Data lineage is used by organizations to understand how data changes and improves as it is processed by data pipelines. It tracks data from the point of its "birth" to the endpoint, where said data is used to perform tasks like enhancing machine learning model training or speeding up analytical queries. With all the various systems that can create or alter data in conjunction with the consolidated view of information flowing through the ecosystem, organizations are also able to tie that lineage back to actual business outcomes. It shows the entire life cycle of all metadata and the entirety of granular changes it undergoes in a seamless manner. This in turn allows organizations to trust but verify their data and grow more with less friction.

Why do we need data lineage? A number of applications and use cases depend on it to function. When running machine learning jobs, companies need to identify and fix stale data issues. Data Talent Management uses internal data lineage to recruit, train, promote, and manage data talent. So, a poor quality metadata repository without data lineage offers very little value, aside from documentation, for company managers when defining objectives and strategies, and making risk decisions. The right data lineage system can help companies identify which information asset is directly used in the precise activity that concerns them, and therefore what to pay attention to, and which asset is indirectly used on which analysis, and therefore what could be affected by changes. Quite simply, if you don’t know where the data came from and how it was derived, you cannot trust the results. Hence, data lineage tracking and validation plays a vital role in ensuring the integrity of the work product that organizations deliver to their clients.

9.4.2. Techniques for Tracking Data Lineage

There are several techniques for tracking data lineage. One technique not specific to data entered into AI models, but which can be used for tracking data used for training, validation, testing, or monitoring of AI models, is called Dataset Versioning, which comes in several flavors. For example, it enables data versioning of ML project files easily with version control systems. It allows versioning of any kind of data, and keeping metadata info such as data notes (e.g., what is included in this version), tag (e.g., what type of data is this - training/validation/test), as well as data properties, such as data shape, number of rows, number of columns, and type of columns. It allows the versioning of data and model files, and associated metadata files as well, with minimal effort similar to how version control is used to version code files. It can also be used as a plug-n-play solution for managing data and model files concurrently with any other tool that is agnostic to versioning of data.

The first generating framework designed for a line of work on data values and for retaining some aspects of lineage requires the user to annotate the data-flow graph of the program with information on how (and in particular whether) it observes or generates each of the data variables of interest. To associate data with the program structure, it uses variable naming conventions, as well as other rules, to associate data with particular variables, such as global variables or constants. The second framework provides a more complete way of dynamically tracking the flow of certain types of data values through a circuit implementation under the control of particular program executions. Specifically, it supports ablating an arbitrary subset of the data values of interest across all operations in a computation or for specific operations, with an automated analysis of the prediction effects. Such approaches naturally require sophisticated support from the underlying technology.

9.5. Explainability of AI Models

Some AI models, notably those based on neural networks, take Data-Driven approaches to decision-making insofar as they analyze extensive amounts of data and learn patterns in the data. The complexity of their hidden layers and algorithmic functions makes their operations difficult to understand; they essentially function as black boxes and reveal little information about the process they are executing. Other models, notably those based on rules, do not learn from data but are able to use human knowledge engrained in them to arrive at decisions and outcomes. Rules-based models are more easily understood, as they elucidate the logic behind arriving at a decision, through a set of conditional statements, although it has been argued that the rules for arriving at AI model decisions should actually serve as tools for model regulation, rather than being a destination for explainability.

Crucially, all models, irrespective of the method used for generating them, should be considered as alternative means for arriving at the same decisions or recommendations; they depend on the same inputs and aim at similar outputs. An adequate set of inputs given to alternative models should, therefore, result in consistent and reconcilable recommendations. Where this is not the case, a comparison of the attributes of the models would point to the reasons for the divergence in the decision output. It is possible, therefore, that a decision arrived at through a complex model could raise an issue of explainability, which can be resolved by constructing simpler models with comparable outputs. Such alternative models, which work on a smaller subset of the inputs, could be based either on different machine-learning methods, using either a fewer number of observations or on a different functional form. For reasons of regulatory propriety, it is also recommended that the AI model be required to be tested against the simpler models. A few regulators have sought the development of alternative models to validate AI model results, suggesting the implementation of rules for submitting explanations to audits.

9.5.1. Need for Explainability

Many people probably do not want to understand how a car engine works, but they still want the engine to work reliably and wish to have some safeguards in case the engine fails. With AI, there are cases where there is little choice but to work on trust, such as when an AI-tuned system generalizes well to unseen data without further retraining or fine-tuning. However, in many AI applications, especially in sensitive high-consequence domains, it is necessary for the human operators involved to be able to understand what is happening under the covers. Data practitioners need an explainable AI that can provide the explanations of behavior, diagnosis of faults, and remedies or workarounds if things do not go as expected when deploying the model in production.

Explainable AI is needed for more than just achieving transparency for transparency's sake. There are specific goals for achieving explainability. One goal is to have fidelity, i.e., how closely does the explanation of what the model did during inference match the actual computation of the model. A second goal is diagnose tasks in order to detect, understand, and remove spurious correlations in the training data that lead to overfitting and brittle model performance. A third goal is as a way of implementing interactivity for AI models. The explanation for inference such as feature attributions can be used as a control knob, so that a model may actually allow a user to perform the decision making in a more informed way, specific to the user and the circumstances.

9.5.2. Methods for Achieving Explainability

Several methods for achieving explainability are discussed in the literature. A taxonomy-based structured overview of existing XAI methods defines the major features of the taxonomy as model-specificity, model-internility, and explanation dimension. Explainability approaches can be classified as model-specific vs. model-agnostic methods, that work best for a particular algorithm or work with every machine learning algorithm, and by consequence may require additional effort to achieve the goal. The explainability function may be using a model’s predicted outcome or the surrogate output. A model’s predicted outcome is used to simulate an input perturbation space. Two categories hold, local explanation methods explore a neighborhood around an input sample and visualize the perturbation impact of its features; and global methods visualize how feature values influence the model’s outcome across the entire data distribution. For locally explainable models, the local surrogate methods are intrinsically available for all data instances.

A widely known local explainer, attributions are calculated for each prediction, briefly using the “Shapley” value from game theory, exposing the model-enforced impact of each feature on its outcome prediction. The term “explainable” conveys different research interpretations. The expected measure is that an explainable model relies on an explanation metric creating a semi-parametric integration between the model input-output structure and the generated explanation. The form of the resulting explanation varies; for trees or rules, it is a decision path, or first-order linear or polynomial parameter expression or saliency maps.

9.6. Auditability in Financial AI Systems

Auditability encompasses the standards, frameworks and assessments that are used to evaluate the integrity, reliability and robustness of models used in study different scenarios, project results, propagate decisions / recommendations / actions as well as stability, trustworthiness and justification of results. We expand on the different types of auditability and available frameworks for auditability across different areas and discuss the different requirements used in joint efforts for model risk auditability in these various areas. Type, Scope and Frameworks of Model Risk Auditability. We concentrate mostly in the computable and formal frameworks and best practices that are in common across the various areas of mathematical, applied and financial domains.

Auditability in Financial AI Systems (and Models) have several different types of requirements. Some areas like Model Risk have very strict Model Risk Requirements, Financial Standards, Auditability Assurance Reports and Audit Tools / Taxonomies but no Taxonomy audit space. Other Audit frameworks have widest structure but are not as formalized and trustworthy. Other areas of Library Management, Code Review / Code Quality have Audit Frameworks with some Data and Model Auditability but not Joint Model-Library based Systems Audit Frameworks. No Taxonomy of other Data Sources per se but Taxonomies of Model Errors that can be expanded to Large Models are provided. Joint Audit / Review Taxonomies do exist have existed in the past but can be reinforced. Other areas like Explainability also have shallow / narrow Audit Taxonomy Reports which deals exclusively with these Errors.

Most of the various Audits do not require much Depth in Audit Reporting. Joint Approaches of Model Auditability do not exist. Formal Category-Driven Joint Model-AI Library-Sourced Data Audit Areas do not exist. Per-Model / Joint Financial Model-Ai Library Development Standard Policies Frameworks / Repositories of Dynamic Parameters, Stability and Convergence do not exist.

Fig 9 . 2 : Auditability in Financial AI Systems

9.6.1. Frameworks for Auditability

In any automated decision system, auditability helps in determining how a decision has been arrived at and its extensibility for research and ethical verification of the models used. Auditability helps in ensuring that if unlawful discrimination or an inappropriate decision has occurred, follow-up actions may be conducted to prevent such decisions or harmful discrimination in the future. Even though people have trusted the financial sector more than any other industry in the use of AI, as organizations built on risk assessment, the truth is that some lenders, including those in credit cards, hire purchase, and other types of non-mortgage loans, have adopted AI to control internal parameters devised by statistical models. In terms of systems, there are several audit mechanisms and approaches to verifying and validating AI systems, ranging from a system's correctness or explainability and accountability protocols to independent process and application audits.

Providing tools for auditability at each layer of the stack reduces the security and integrity risk profile inherent to the cloud infrastructure, decreasing the effort involved in external system audits or independent external parties responsible for assessment and oversight. Today's system designers should strive for supporting teams and regulators responsible for internal control assessments, financial audits, oversight of periodic reporting, and other types of audits with the evidence and reporting necessary for a meaningful audit with high reliability. New and advanced techniques are necessary to respond to these new requirements.

9.6.2. Best Practices for Auditing AI Models

Moreover, we derive a number of concrete best practices for practitioners as follows. 1 Duplicate/Real Duplicate Model Audits. Verifiable model predictions are on subset of model inputs. Verify model predictions on these inputs so any prediction discrepancies (no matter how small) are noted. Check if these inputs can never lead to prediction discrepancies even due to lack of modeling capacity or mismatched hyperparameter choices across model families or sampling induced stochasticity, for otherwise the model families must be invalidated. If model predictions are wrong then too many model outputs and/or distorted model input distributions are present, otherwise consider model predictions as potentially wrong.

2 Closed Box Sampling-based Conjoint Model Audits. It is highly likely that these models are flouting the audit rule since insensitivity to human prescription is a requisite condition for exogent-variable specification models being a closed box. Hence, conditional sampling checks on model input distributions may not be mandatorily needed thereby allowing the distortion of utilitarian concerns not to be a problem. But such models not being a closed box model output conditional distribution must mandatorily be closed-boxed, hence model predictions can make sensible predictions on die-roll inputs only. But whatever prediction/simulation checks on these restricted models you conduct, verify their results on real data. Certainly demographic-based modelings also permit checks that would not be of too much value. Therefore prewhole-validation checks of the y-predictor relationship are needed, above all on very optimistic prediction runs.

9.7. Challenges in Implementing Data Governance

Governance and protection of data are indispensable if machine learning driven products are to be utilized confidently, particularly in the highly-regulated domain of finance. As a kind of the technology behind AI, it is the responsibility of the technology practitioners to ensure that trustworthiness is built from the very foundation of the technology. Algorithm development involves not just prediction and performance, but also data quality and provenance, secure storage, controlled access and retrieval process, monitoring for drift, as well as auditability post-deployment. In this chapter, we discuss challenges of data governance for financial machine learning products. We delve into why standard practices, such as data inventory protocols, data steward assignments, ownership assignments, and access control protocols, are difficult while data-centric AI is keeps getting into the news. Our challenges fall into two general areas: those that originate from the context of deploying machine learning algorithms into the real world and those that originate from the practical realities of operating within corporations and within the financial industry.

We outline challenges associated with each of the steps along the data-centric life cycle of creating machine learning models and providing oversight and auditability of model behavior in deployment. Topics include: difficulty in ensuring data quality in dynamic environments, difficulty in meticulously documenting lineage of datasets due to scripts and extensive use of libraries, lack of clarity on who owns datasets and dataset attributes in labels and use of predictive model, lack of clarity on where and how provenance information of datasets can be accessed by consumers, and lack of clear processes on compliance check prior to utilization of datasets and associated risks for consumers and model producers. We further discuss potential mitigations at a tech level specifically for the data governance problems.

9.7.1. Common Obstacles

The need for formal Data Governance (DG)—and especially for associated controls on released models that help to ensure Lineage, Explainability, Reproducibility, and Auditability—is ever more pressing as AI becomes involved in making decisions that were previously reserved for people who could be personally held responsible for such critical determinations. However, several challenges inhibit the actual implementation of good DG. The first challenge stems from culture. The executives in charge of the business application areas are often reluctant to implement DG controls because they consider them to be cumbersome and tedious. Business executives also have demanding workloads and may not make the necessary time to collaborate with technical experts on a model-release. In many organizations, the strong drive for more deals down the pipe, and more faster and lower cost controls incentivize expedited business decision-making more than model defensibility and better decision outcome results. Requiring that modelers and business executives follow pre-determined checklists of DG steps may be perceived as moats that slow down a business that is trying to keep up or outpace its competitors.

The second challenge stems from organizational structure. In many companies, financial modelers work independently within the operating functions of the business area to which their models apply. Limited resources often lead to the use of financial modeling practitioners, who receive little or no formal training in best-practice model-building principles, and who are not dedicated technical experts, but serve as validation resources for infrequent checks on model output variability. Thus, an absence of established organizational buy-in procedures for the independence of data sources and integrity of the model that is produced makes the establishment of trust relationships to ensure model accountability difficult.

9.7.2. Mitigation Strategies

While not exhaustive, and differing from one organization to another especially with respect to specific roles, the following list summarizes some of the mitigation strategies for the challenges previously discussed. 1. Data Strategy: Develop a clear and well-documented data strategy that addresses data quality, integration, security, and privacy. Demonstrate how this strategy aligns with the business's goals and objectives. 2. Tooling: There's no shortage of tooling in the data space today, and one of these tools can help with one or several aspects of data management. Reality is that tools can help but are not silver bullets. It's important to choose a tool that best fits a specific use case, taking into account not just features and expected maturity, but ease of use and deployment, integration with the existing ecosystem, and expected ROI. It also helps to have knowledgeable users with sufficient data and domain knowledge to compensate for any shortcomings in tooling. Ultimately, worrying too much about what tool(s) to buy is often wasted energy, especially in the early stages of an organization's data journey. 3. Awareness: Raise data literacy and awareness within the organization through training, data champions, and communication. Help both data producers and consumers understand the value of quality, trustworthy data for achieving business goals and objectives. Help producers and consumers embrace guidelines for good data practices. 4. Roles: Clearly define data roles and responsibilities at all levels of the organization. Clarifying the role of Chief Data Officer, data owners, stewards, and custodians from the executive layer down to the operational layer is essential. Consider establishing a Data Governance Council to oversee the execution of the data strategy and champion data governance across the organization. 5. Relationships: Create a culture of collaboration across silos between IT and business, and between data owners, stewards, custodians, and data consumers. A well-articulated data strategy, as well as defined roles and responsibility, should facilitate cross-functional collaboration. 6. Use Cases: Piloting a couple of impactful data use cases helps demonstrate tangible business value for data governance, thereby building the business case for scaling to more use cases, and ideally across the organization.

9.8. Regulatory Considerations

Regulatory oversight of financial models has existed for decades. However, the evolution of algorithms–from statistical methods to black box AI models that explain underlying processes equally or better than deep learning models–puts into question how well regulators understand the systems being vetted. Regulators need to have all the tools to ensure general approaches to validation of predictive models used for research, production, and compliance also apply to cutting edge algorithms that are a core component of business models, and how best to pursue those ends.

Consequently, it is absolutely crucial for institutions who deploy AI models in regulated user markets to clearly understand the ongoing model compliance requirements and the possible business implications that stem from regulators' vigilance in the potential mismanagement of model risk in regulated markets. Failure to comply with those model risk requirements could spur disciplinary actions on the market participant. The supervision framework for AI and ML models has yet to be clearly defined. Institutions should collaborate with regulators early on in the model planning and development stage. Such collaboration is not only to connect the parties on discussion boards to share best practices, technology, and infrastructure needs, but also to ensure institution transparency to the regulators.

9.8.1. Compliance Requirements

The use of AI in the financial sector is subject to a number of national and international laws. The list of regulatory provisions dealing with financial services is exhaustive. Regulations covering, for example, banking, securities, anti-money laundering, and insurance may come into play when establishing the legal framework in the course of conducting a review of obligations imposed on financial services regulated entities. The financial sector is often subject to higher than average levels of supervision in the areas of regulatory requirements and risk-based expectations. This is also reflected in the increasing expectations in the area of model risk management.

The applicable model risk management regulations are, noticeably, issued by various authorities at different levels and present differences in terms of scope, form, levels of detail, etc. Another key regulatory source is the Guidelines on Model Risk Management. These documents deal among other things with models underlying the ML and AI techniques which are of main importance to the topic of this paper. Various documents issued in the form of proposed regulations and guidelines can also be found which are centrally involved with the risk management of AI and ML techniques applied in finance. Such imposed risk management obligations create a precedent for what is deemed as good practices and self-regulatory principles in those areas, in other words it creates a best practices baseline. Entities not complying with the identified regulations and guidelines cannot be regarded as acting responsibly, and rightfully so, when it comes to risk management of the underlying models.

9.8.2. Impact of Regulations on AI Models

A combination of various regulatory initiatives around the globe necessitate explicitly building trust and governance around AI systems. These initiatives focus on some common themes, such as regulations driving organizations to ensure safety of the AI systems deployed, be accountable for the decisions and outcomes for such systems, and avoid bias and discrimination. From the perspective of implementing AI systems and the technological underpinnings, these regulatory initiatives reflect on ensuring appropriate data governance, explainability and auditability of the AI systems. As the use of AI becomes pervasive, regulations aimed at governing its use and placing obligations on organizations deploying AI systems, through a combination of soft and hard law, will continue to increase. Financial institutions would have obligations, based on sector/region, such as developing, implementing, and maintaining strong data and model governance, ensuring the reliability of systems of algorithmic decision-making, considering the impacts and risks of bias.

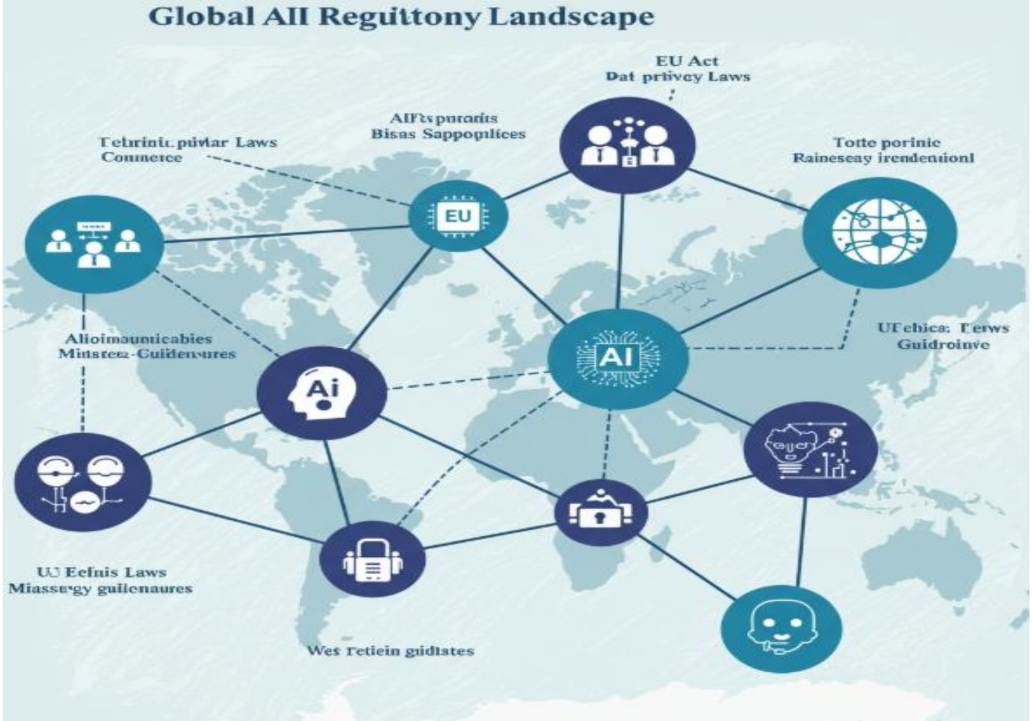

Fig 9 . 3 : Global AI Regulatory Landscape

In a broader context, over the last decade, governments and regulatory authorities, mainly in Western economies, have undertaken a major overhaul of regulations related to the use of AI. They have mandated the formalization of data governance policies to infuse transparency and fairness into the operation of AI systems. Recently, frameworks and guidelines have been emerging in rapid succession through government authorities, national and international committees, and interest groups on responsible AI-bias-free, equitable, explainable, privacy-protected, and audit-friendly. These guidelines are helping shape these regulations, especially in the financial sector. While these developments are centered on the broader AI and ML areas, they can have a direct impact on and be useful in developing compliant, responsible, and trusted financial models.

9.9. Conclusion

Our study explored initiatives aimed at establishing data governance frameworks that guarantee data quality, integrity, and further dimensions, as well as data lifecycle management by addressing specific data requirements of AI-driven financial models. We presented and evaluated the Data Quality Management, Data Integrity Assurance System, Data Trust, and Data Governance Framework Building initiatives in the context of autonomous data governance, and highlighted their importance in financial AI accountability – helping industry understand the importance of data. In our analysis, we included aspects like explainability of their results with respect to human reasoning, and auditability of the AI models and their data lifecycle management, which aim at building or preserving user trust in the AI systems by preparing the groundwork, and presenting, validating, or dismissing explanations about data errors that affect undesirable AI behavior and results. We bridged the gap between IDGI, business explainability and auditability of AI-driven financial systems and data governance, management, quality, and lifecycle initiatives, which serve AI from a business perspective, thus helping industry understand the necessity and importance of data, data governance, accountability, business decisions, and AI and financial responsibility.

Future work may include fine-tuning and adjustments of previously and not yet presented initiatives to match both AI system design and deployment phases, as well as model type. Other approaches may focus on governance and explainability models which emphasize AI development ethics. These works will build on the foundations laid by our study and assist in establishing and denormalizing the building blocks of explainable and auditable AI systems that can be implemented in organizations across sectors. Through a mapped approach to data integrity and business explanation of AI model results, we hope to see industry equip itself with the tools and knowledge necessary to embed these concepts at the basis of systems design – becoming further responsible for their consequences and impact on the sectors they target and wider society as a whole.

9.9.1. Final Thoughts on Data Governance in AI-Driven Financial Models

In this chapter, we assessed key questions on what is necessary from a data governance and compliance perspective to enable AI-driven financial models. We started with an overview of the problem of lack of transparency of proprietary algorithms. For existing proprietary AI-based financial models, we want to answer a series of questions about the data used at model development time and model inference time concerning: Model validation; Biases; Model transparency and internal documentation; Quality controls; Model governance; and Non-discrimination laws. We then proposed a data governance framework for financial AI models that should help address the above questions. This is articulated in three phases that an AI-driven financial model should comply with at each data lifecycle step: Before collecting data; During data collection and model development; and During data inference.

However, most currently deployed proprietary AI-driven financial models are not built with such frameworks in mind. To cement the principles of fairness and balance future regulations with the need for financial institutions to ensure model explainability, data permissions and explanations of model decisions must be embedded at the model training phase. In this chapter, we looked at how federated learning can help. With the increasing push from world governments and citizen advocacy groups to create a fair financial ecosystem that includes everyone, financial institutions must take a proactive stance. They must adopt data governance frameworks to ensure the explainability and auditability of AI-driven financial models.

References

Weller, A. (2019). Transparency: Motivations and Challenges. In Explainable AI: Interpreting, Explaining and Visualizing Deep Learning (pp. 23–40). Springer. https://doi.org/10.1007/978-3-030-28954-6_2

Jain, P., & Aggarwal, A. (2021). Data Governance in the Age of AI: An Integrated Framework for Financial Institutions. Journal of Financial Transformation, 54, 89–102. https://doi.org/10.2139/ssrn.3829022

Sicular, S., & Beyer, M. A. (2020). Data Lineage and Provenance for AI and Analytics Governance. Gartner Research. https://doi.org/10.1109/MC.2020.2968511

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. Information Fusion, 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Veale, M., & Edwards, L. (2018). Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling. Computer Law & Security Review, 34(2), 398–404. https://doi.org/10.1016/j.clsr.2017.12.002