INTERNATIONAL MONETARY FUND

AI Projects in Financial Supervisory Authorities

A Toolkit for Successful Implementation

Parma Bains, Gabriela Conde, Rangachary Ravikumar, and
Ebru Sonbul Iskender

WP/25/199

2025
OCT

WORKING PAPER

 WP/25/199

IMF Working Paper
Monetary and Capital Markets Department

AI Projects in Financial Supervisory Authorities: Toolkit for a Successful Implementation

Prepared by Parma Bains, Gabriela Conde, Rangachary Ravikumar, and Ebru Sonbul Iskender*

Authorized for distribution by Jay Surti
October 2025

ABSTRACT: This paper discusses the imperative for financial supervisory authorities to enhance their toolkit through the adoption of Artificial Intelligence in response to the growing digitalization of financial services. It aims to assist authorities in safely and effectively overseeing applications of Artificial Intelligence in the financial sector by proposing a tailored project management methodology for implementation of Artificial Intelligence by financial supervisory authorities that address unique risks and align initiatives with strategic goals. Key challenges, including ensuring explainability and mitigating bias, with a focus on stakeholder collaboration, are emphasized, alongside prerequisites for successful deployment, such as robust governance frameworks and adequate resources.

JEL Classification Numbers:	G18
Keywords:	Artificial Intelligence, Machine Learning, Deep Learning, Generative AI, DevOps, MLOps, AI governance, Data governance
Author's E-Mail Address:	pbains@imf.org, gconde@imf.org, rravikumar@imf.org, esonbuliskender@imf.org

* The author(s) would like to thank M. Barzanti, A. Deghi, Z. Gorpe, A. Paduraru, K. Seal, P. Singh, H. Tourpe, K. AlAjmi, and B. Zhao.

WORKING PAPERS

AI Projects in Financial Supervisory Authorities

A Toolkit for Successful Implementation

Prepared by Parma Bains, Gabriela Conde, Rangachary Ravikumar, and Ebru Sonbul Iskender[1]

[1] The author(s) would like to thank M. Barzanti, A. Deghi, Z. Gorpe, A. Paduraru, K. Seal, P. Singh, H. Tourpe, K. AlAjmi, and B. Zhao.

Contents

Glossary

AI	Artificial Intelligence
AML/CFT	Anti-Money Laundering/Combating the Financing of Terrorism
BIS	Bank for International Settlements
CCAF	Cambridge Center for Alternative Finance
CRISP-DM	Cross Industry Standard for Data Mining
D.A.T.A.	Data, Autonomy, Technology, and Accountability
DevOps	Development and Operations
FSI	Financial Stability Institute
GDPR	General Data Protection Regulation
IT	Information Technology
ML	Machine Learning
MLOps	Machine Learning Operations
OECD	Organization for Economic Co-operation and Development
RBS	Risk-Based Supervision

I. Introduction

Financial supervisory authorities must keep upgrading and adapting their toolkit to keep pace with financial sector innovations.[1] Financial institutions have intensified the use of digital technology. By combining vast amounts of data from a wide range of sources with cloud processing and Artificial Intelligence (AI) systems, they are deploying customer centric solutions, streamlining business processes, and redesigning risk management.[2] Financial supervisory authorities face challenges in monitoring this dynamic field. Rapidly evolving technology leaves little room for complacency, compelling them to harness data-driven tools and implement more efficient supervisory approaches to fulfill their mandates.

Adopting new and emerging technology for supervisory processes (or suptech) continues to trend upward, albeit with significant variation across countries. Emerging research suggests that 164 financial authorities from 105 countries have implemented suptech tools.[3] The main areas covered by these initiatives are the prudential supervision of banks and non-bank deposit-taking institutions, consumer protection and market conduct supervision. Other areas that are receiving increasing attention are cyber risk supervision, and ESG reporting alongside securities and insurance supervision.

Nonetheless, authorities from advanced, emerging market and developing economies have uneven rates of suptech adoption, generating concerns of a tiering of supervisory authorities across countries into those that are willing and able to adopt new technologies and others that cannot. In 2024, 75 percent of advanced economies and 58 percent of emerging markets and developing economies had adopted suptech tools, compared to 79 percent and 54 percent, respectively, in 2023.[4] While most deployed tools primarily support descriptive and diagnostic analysis and rely to some extent on manual processes, there is a clear interest in transitioning from traditional methods to more advanced technologies, including AI. A sizable number of respondents are exploring how to incorporate into their supervisory processes (60 percent). While still low as a share of the number of countries int the sample, the use of GenAI more than doubled between 2023 and 2024 (from eight to 19 percent). A recent stocktake conducted among supervisory authorities indicated that 32 out of 42 respondents are experimenting with, using or developing GenAI tools in financial supervision.[5] Use cases mostly relate to process automation and can be grouped into three: categories: (i) basic document processing; (ii) knowledge management; and (iii) document review.

Integrating AI into supervisory processes presents numerous opportunities for financial supervisory authorities. From a supervisory perspective, AI systems can contribute to data management tasks such as validation,

[1] International standards recommend that supervisors have sufficient resources, including human, technological and financial resources, to enable it to conduct effective supervision (see BCBS – Core principles for effective banking supervision, IAIS - Insurance Core Principles and Common Framework for the Supervision of internationally Active Insurance Groups, and IOSCO - Objectives and Principles of Securities Regulation).

[2] An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment (OECD, 2024). AI systems include different subsets such as machine learning, deep learning -labeled as traditional AI-, and novel forms such as generative AI, large language models, or agents. This Note covers all forms of AI and will specify a subset when corresponds. For definitions on different forms of AI, see FSB (2017, 2024)

[3] Cambridge SupTech Lab (2024)

[4] 64 authorities responded to the 2023 survey vis-à-vis 164 in 2024.

[5] Prenio (2025).

consolidation, and visualization. For example, authorities can substitute manual quality controls of completeness, correctness, and consistency of formatting and calculation with AI while continuing to adhere to reporting rules even as AI merges multiple data sources and facilitates their seamless integration. In isolation or when combined with web scraping techniques, AI models can aid anti-money laundering / countering the financing of terrorism (AML/CFT) supervision by detecting suspicious patterns in granular data from various sources. AI models can assist financial supervisory authorities when conducting market surveillance by identifying information in investment advisers' regulatory filings that may warrant initiating enforcement actions. Other applications include identifying misleading marketing or performing real-time monitoring of market transactions.

When applied for prudential purposes, AI systems can enhance the supervision of risks, notably credit and liquidity risks, and besides facilitating the review of documents during licensing and governance assessments. For financial stability purposes, supervisory authorities can leverage AI models to predict liquidity problems affecting participants in financial market infrastructures and to design scenarios to test the resilience of financial institutions.

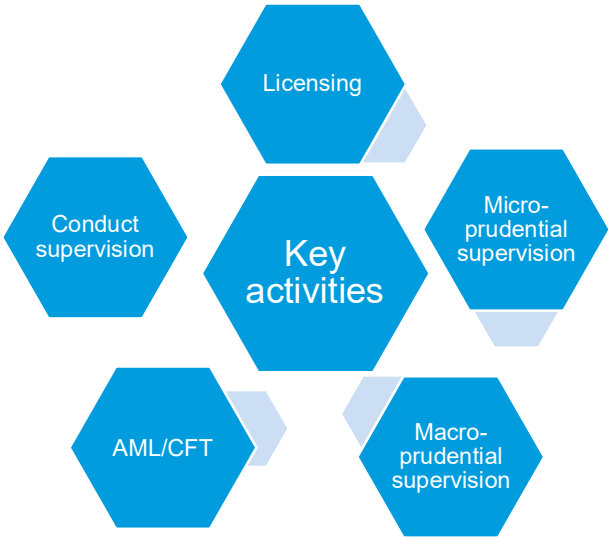
Novel forms of AI offer new possibilities for financial supervisory authorities. More specifically, GenAI enhances capabilities such as information retrieval, content creation, code generation, debugging, and legacy code optimization.[6] These code generation tools enable financial supervisory authorities to accelerate the deployment of AI in traditional use cases, such as fraud detection, monitoring of market activity, and streamlining data management tasks. Additionally, subsets of GenAI, such as Large Language Models (LLMs), may support additional use cases, such as the supervision of corporate governance. For example, LLMs can scan board meeting minutes to identify discussions pertaining to key governance issues such as executive compensation, related party transactions, and risk management.

Financial sector authorities have deployed AI applications to support all key Risk-Based Supervision (RBS) activities. RBS, recognized as the best practice for overseeing financial institutions worldwide, entails several key activities, namely licensing, micro- and macro-prudential supervision, AML/CFT, and conduct supervision (Figure 1). The Bank of England employs AI tools to support macro-financial and macro-prudential surveillance, including its GDP growth forecasts, banking distress, and financial crises. The Bank of Thailand employs AI to support prudential supervision, including analyzing board meeting minutes of supervised financial institutions to assess regulatory compliance. The European Central Bank applies Natural Language Processing (NLP) and AI tools to read fit-and-proper questionnaires and flag issues based on their content, thereby streamlining its authorization process. The Malaysian Securities Commission uses AI tools to monitor the adoption of corporate governance best practices and the quality of disclosures of listed companies on the Malaysia Stock Exchange, as part of its ongoing conduct supervision. A recent study indicates that successfully deployed applications are critical in enhancing the supervisory process. [7]

[6] IMF (2024)

[7] Prenio (2024).

Figure 1. Key Supervisory Activities under RBS



SOURCE: World Bank Group, *A Roadmap to SupTech Solutions for Low Income (IDA) Countries*.

Financial supervisory authorities encounter numerous challenges in deploying AI solutions despite growing interest and potential benefits. These challenges include developing solutions that are explainable, robust, and capable of preventing discrimination, while operating with limited skilled resources, as well as addressing the incompatibility of legacy systems and existing data and model governance frameworks with the specificities of AI models. Additional challenges arise from the characteristics common across several public sector entities, such as competing objectives, complex organizational structures with multiple specializations and different priorities, and procurement processes that require extensive transparency.[8] Furthermore, embedding AI tools often necessitates broader organizational digital transformation efforts, including adopting a suptech strategy, enhancing data collection practices, or developing a long-term view of the IT supervisory ecosystem.[9] [10]

[8] CeciliaSkingsley, "Sharper supervision in an era of technology races" (speech, Head of BIS Innovation Hub, at the Finance Global Summit, London, 15 April 2024).

[9] Prenio (2024)

[10] Denis (2021)

Limited preparedness and access may hinder AI adoption despite new forms of AI providing greater potential efficiencies and benefits.[11] The deployment of GenAI models necessitates significant resources which are especially scarce in emerging markets and developing economies. Moreover, models such as LLMs, which are trained on diverse data sources, may not be suitable for the specific requirements of the supervisory work. They may require substantial adaptations that necessitate skilled IT personnel and sufficient computing power.

This paper aims to outline the essential prerequisites that financial supervisory authorities should consider when implementing AI initiatives, providing a detailed toolkit for guidance. It builds on the risk considerations of AI identified in previous IMF publications and explores these considerations from the perspective of financial supervisory authorities. This paper does not offer new policy positions but uses existing Fund positions to develop this guide. [12]Section 2 outlines key elements of internal governance, risk management, project management and project team composition that financial supervisory authorities should consider when planning to initiate an AI project.[13] Section 3 emphasizes a project management methodology adapted to the iterative nature of AI initiatives, prioritizing the achievement of business objectives while incorporating techniques to mitigate the unique risks associated with AI. Lastly, Section 4 provides conclusions and recommendations.

[11] In 2024, IMF introduced the AI Preparedness Index, which covers strategic areas for AI readiness. The index is made up of a selected set of macro-structural indicators organized into four categories: (1) digital infrastructure, (2) innovation and economic integration, (3) human capital and labor market policies, and (4) regulation and ethics. The index shows that wealthier economies are generally better prepared than low-income countries to adopt AI with considerable variation across countries. While advanced and some emerging market economies are highly exposed to potential disruptions from AI, yet they are also well positioned to harness the benefits and mitigate the risks of AI. Their preparedness stems from digital infrastructure, human capital, and adaptable regulatory frameworks. On the other hand, low-income countries are underprepared to harness the benefits of AI across all dimensions (see Georgieva, K. (2024), AI Will Transform the Global Economy. Let's Make Sure It Benefits Humanity.).

[12] Risk considerations of AI systems -namely, explainability, bias, data privacy, cybersecurity, and robustness- are explored in depth in AjAjmi et al (2021), Boukherouaa et al (2023). IMF (2024) underscores that AI provides numerous opportunities for supervisors to, for example, generate efficiency gains and conduct real-time monitoring of markets.

[13] For the purposes of this document, an AI project is a development endeavor that utilizes AI technologies such as machine learning, natural language processing, and computer vision to solve a specific problem or achieve a goal, essentially creating a system that can perform tasks typically requiring human intelligence (source: AI Projects: Revolutionizing Efficiency Across Sectors | Lenovo US).

II. Institutional Framework and Approach to AI

An AI project represents the complex interplay of technology, processes, and people, necessitating careful reflection, planning and resource allocation. As a first step, financial supervisory authorities should assess the overarching governance and risk management framework within which AI projects will be executed and monitored. Existing prudential governance and risk management frameworks could be a valuable benchmark for this assessment. Furthermore, it will be essential to strategically evaluate the availability of necessary data and technology for each AI project to empower staff to enhance the supervisory process while adhering to ethical and legal standards. There are interesting early-adopter lessons that can be used for such an evaluation. In addition, the project management methodology, which ideally is kept agile, should align with the iterative nature of AI projects. Finally, a team composition that comprises diverse skills and represents different functions across the financial supervisory authorities is vital for the successful deployment of AI solutions.

The AI governance and Risk Management Frameworks

Following BIS (2025) AI governance can be defined as "the set of principles, responsibilities, structures and, broadly, frameworks that allow for an effective use of this technology while aligning it with organizational goals and risk compliance standards". Four overarching principles present in frameworks issued by governments and international organizations could guide financial supervisory authorities when building an AI governance framework. [14] [15]

First, financial sector authorities should define organizational structures, roles and responsibilities, performance measures and accountability for AI model outcomes. Second, humans should be kept in the loop for AI-augmented decision-making. A critical issue is that models should be designed to avoid any unfair discrimination against individuals or groups. Third, the framework should address the operational management of AI models.

Financial sector authorities should ensure that AI models respect data privacy and quality issues during their lifecycle and identify potential security vulnerabilities while implementing resilience measures. When models are developed by third parties, financial sector authorities should specify in writing the scope of work and activities to be conducted. Financial supervisory authorities should also monitor the performance of models to ensure they continue to meet requirements and record the model's lifecycle. Fourth, the framework should address transparency in communication with all relevant stakeholders.

The governance structure should reflect the complexity and the impact of the use cases. For low-impact use cases, accountability might be retained at the business owner level. Intermediate cases might require a coordinated team that brings together stakeholders involved in their development and governing. For more impactful cases, financial supervisory authorities could consider establishing a board committee that ensures alignment with organizational goals, sets up the risk appetite for AI, assesses AI models along with their

[14] For example: FSB, *Financial Stability Implications of Artificial Intelligence*, XX 2024; IOSCO, *The use of artificial intelligence and machine learning by market intermediaries, and asset managers*, September 2021; and BCBS, *Corporate Governance Principles for Bank (*as it also covers risk management including models), and Principles for Operational Resilience.

[15] ECB, *Guide to internal models*, February 2024; FED and OCC, *Supervisory Guidance on Model Risk Management*, April 2011; Monetary Authority of Singapore, *Technology Risk Management Guidelines*, January 2021; Bank of England, *Model risk management principles for banks*, May 2023; JFSA, *Principles for Model Risk Management*, November 2021.

prioritization, and their approval before deployment. Regardless of the complexity of the use case, financial supervisory authorities should take into consideration that a multi-disciplinary approach is required to ensure effective governance and that a strong linkage with data governance should exist. Whilst each organization's AI needs are unique, the multi-disciplinary team should include legal experts, ethics experts, data scientists, business owners, and other relevant stakeholders. Whichever the adopted framework, the fast-evolving nature of AI models forces financial supervisory authorities to embrace flexibility and agility.

Data governance is the bedrock of successful deployment of AI solutions and contributes to mitigating risks of lack of explainability, robustness, and bias. Canada's Office of the Superintendent of Financial Institutions (OSFI, 2023) defines AI governance as "the structures, systems, and practices an organization has in place for decision-making, accountability, control, risk monitoring and mitigation, and performance reporting."[16] Data governance has a bearing on overall AI governance, and financial supervisory authorities should consider it holistically. Data governance entails policies, processes, and activities aiming to ensure that data is secure, private, accurate, available, and usable. It is a principled approach to managing data during its life cycle, from acquisition to use to disposal. Data governance also involves complying with external standards set by government agencies, standard setting bodies and other stakeholders.[17]

In high-stakes domains, such as finance, laws and regulations, like the European Union's General Data Protection Regulation (EU-GDPR), may impose requirements on models' transparency, trust, explainability, and faithfulness. AI models may introduce bias by systematically and unfairly discriminating against certain individuals or groups of individuals. Both data and algorithms used in optimizing AI models may contribute to bias. While data governance frameworks propose data collection best practices, financial supervisory authorities should also use metrics and audits, integrate AI developing teams in a diverse manner, train staff and monitor systems' performance to further mitigate bias risk. Typical solutions often involve setting up robust data management frameworks, recording data details (such as metadata on sources, versions, and curation), keeping asset inventories (like data catalogs) and metadata registries, adhering to standards, and facilitating efficient data exploration for both humans and machines, for example, through application programming interfaces (APIs) (BIS 2025).

AI systems need to be robust in preventing security and data breaches. AI systems are like any other IT systems. Therefore, all the general IT and cybersecurity measures are equally applicable to them. Identification, protection, detection, response, and recovery aspects need to be implemented effectively. Governance and risk management play a key role. It is also important to follow the three-lines of defense approach -business level controls, risk management, and independent audit- to manage such risks and achieve a better outcome.

[16] Financial Industry Forum on Artificial Intelligence: A Canadian Perspective on Responsible AI - Office of the Superintendent of Financial Institutions (osfi-bsif.gc.ca)

[17] In 2013, the BCBS published its Principles for Effective Data Aggregation and Risk Reporting, which provide useful guidance for financial authorities when designing, maintaining, and monitoring data and IT architecture, data ownership and data quality across their lifecycle. While the Principles apply to banks' group risk management processes, they are also applicable for financial and operational processes and those outsourced to third parties. A report published in 2023 by the BCBS on the progress in the application of the principles showed that banks are still struggling with their adoption. While notable improvements have been made in many key areas, banks still present significant challenges with fragmented IT architecture, legacy systems and manual processes not fit for purpose. Lack of prioritization, insufficient ownership by the board and senior management, as well as challenges with updating data and IT infrastructure largely explain the delay in complying with the principles. The report also found that while new technologies such as AI might potentially help overcome persistent data management challenges, banks still lack quality data that is a prerequisite for profiting from digitalization projects.

Cybersecurity in AI needs to be considered at various stages, namely design, development or procurement, deployment, and operations. Cybersecurity in AI cannot and should not be an afterthought. Cybersecurity in AI is a new and emerging topic where standard setters are currently engaged in developing principles and guidance. For example, the International Organization on Standardization (ISO) is developing a new standard called *ISO/IEC 27090* where the ongoing work is likely to be published in 2025.

If financial supervisory authorities decide to use or rely on third parties for developing or deploying AI, a third-party risk management framework must be in place including policies, processes, risk management activities, due diligence, ongoing monitoring, independent assurance, business continuity arrangements, disaster recovery arrangements and exit plans. When evaluating third-party data sets used by AI systems, it is important to clearly understand data quality, training data sources, data ownership, and traceability. Regarding AI model attributes, it is necessary to clarify the type of model, learning method, biases that may be present, autonomy level, and the extent of human oversight employed. In this way, financial supervisory authorities should align the identified third-party AI risks with the current corporate risk appetite before making decisions on legal agreements, ensuring that the risk appetite is adequately factored in by the conditions and duties of third-party service providers.

Lessons from the D.A.T.A. framework

Lessons learned from a framework developed to predict the success of big data initiatives could guide authorities to strategically reflect on AI adoption. According to a representative study, as many as 85 percent of these initiatives fail because executives cannot accurately assess project risks in advance.[18] They proposed a framework, known as D.A.T.A., that helps to predict the success of a big data initiative. To operationalize it, the authors suggest answering the following questions sequentially.

Do we have access to data that is valuable and rare? Can employees use data to create solutions? Can our technology provide the solution? Is our solution in accordance with laws and ethics?

These questions represent the four components (D.A.T.A.), i.e., data, autonomy, technology, and accountability. The initiative's success depends on the final score obtained by answering the D.A.T.A. questions. Each question requires a simple yes or no answer, with each yes earning one point. If the score reaches four, the project has a high likelihood of success. A score of three suggests that the project requires critical adjustments to achieve success, while a score of two indicates it is unlikely to succeed. A score of one makes success very unlikely, and if the project scores zero, it should be terminated immediately.

Agile Methodologies and DevOps/MLOps

Project management for software development has traditionally followed a Waterfall Method. The approach is largely suitable for clearly defined processes that identify dependencies of the work, the budget, and the deadline, all of which are important for large projects. It is particularly useful for public authorities as it helps to allocate budget clearly and transparently and gain buy-in from within the organization and external stakeholders. Substantial changes after development starts can be expensive and are, therefore, typically avoided. Testing occurs prior to the market launch of a product and consequently issues that arise during the testing period can delay the project and be expensive to remedy. Importantly, with software technology

[18] Lund Pedersen and Ritter (2020)

developing rapidly, relevant market updates cannot be integrated into the design once the project is in the development phase.

Agile methodologies based on the "Manifesto for Agile Software Development"[19] provide a framework for AI project teams to respond quickly and continuously improve their models through a focus on flexibility and adaptability during design and a level of pragmatism in terms of delivery of the final product.[20] Agile methodology requires an iterative approach with small steps (so-called 'sprints') followed by regular and frequent feedback, segmenting a larger project into smaller pieces. These approaches can foster collaboration, communication, and trust by onboarding relevant stakeholders to ensure that AI project teams can create solutions that are both innovative and aligned with business objectives. The aim is to deliver faster feedback cycles that can identify problems early with the flexibility to make changes throughout the project, thereby reducing risks and the likelihood of costly mistakes and project delays. Agile methodologies are increasingly gaining popularity across economic sectors and have, instead of the Waterfall Method, been identified as the preferred alternative.[21]

Agile methodologies had a decisive influence on the emergence of the DevOps model that integrates Development and Operations within IT departments. DevOps is a response to the traditional software development model that assumes an organizational and functional separation between those who write code and those who deploy and support it. Under a DevOps model, development and operations teams integrate to improve and shorten the software development lifecycle. By leveraging the expertise of IT professionals and enabling teams to build, test, and release software faster and more reliably, DevOps can contribute to a successful deployment of supervisory AI tools. DevSecOps is an extension of the DevOps practice that integrates security testing at every stage of the software development process.[22]

Machine learning operations (MLOps) is a recent response to AI projects' specificities. It involves applying DevOps principles and best practices to deploying, monitoring, and managing ML models and data. MLOps aims to improve AI/ML workflow by automating steps such as data ingestion, data preparation, modeling, evaluation, and deployment. For example, MLOps can contribute to data preparation by linking information across domains and sources, such as geospatial information, business registers, and financial statements, or by performing data comparisons against benchmarks.[23] AI models can degrade quickly if the data they receive post-deployment differs from the data used during their training. MLOps can mitigate this risk by incorporating procedures that facilitate continuous monitoring and model retraining.

[19] Manifesto for Agile Software Development

[20] What is Agile Software Development | TechFAR Hub Handbook | USDS.gov

[21] Digital regulatory reporting 2023 (bankofengland.co.uk)

[22] Under the DevSecOps approach, programmers should ensure that the code is free of security vulnerabilities while security practitioners test the software further before releasing it. It is recommended that teams implement DevOps and continuous integration before adoptin DevSecOps.

[23] IFC (2024)

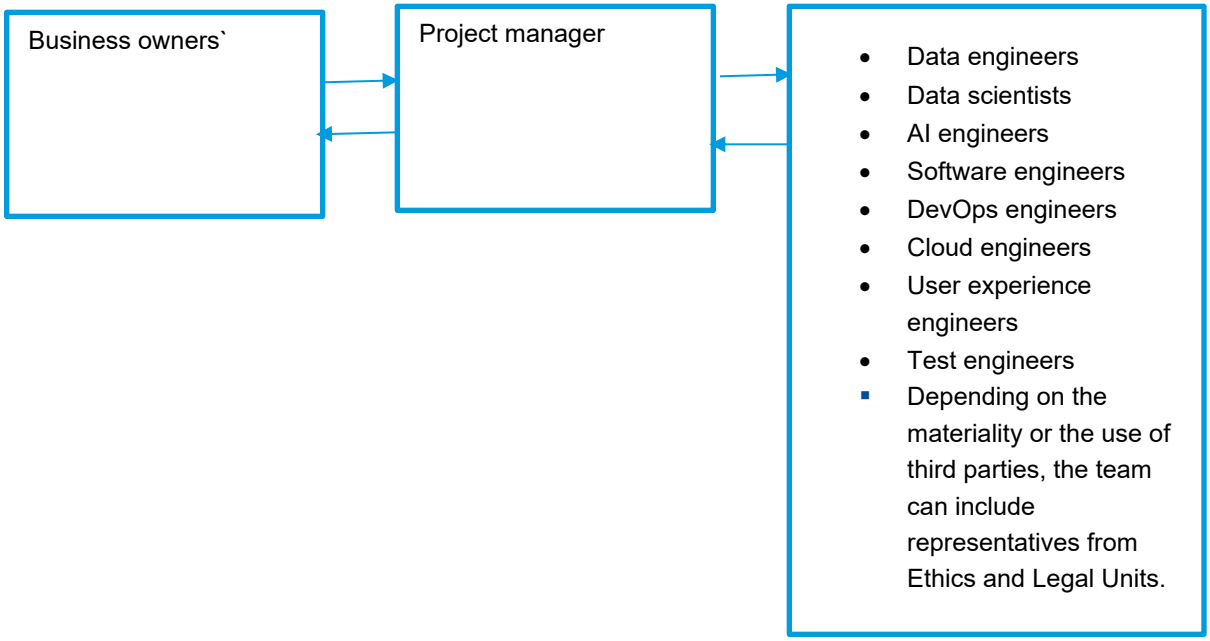
Project Team

AI projects require a dedicated team that fulfills different roles and represents various functions of the organization. The core project team should include business representatives, solution designers, and a project manager or connector (Figure 3 and Table 1). Business representatives or business owners are responsible for defining the use case, selecting key success criteria, and performing operational tasks before and after deployment. Supervisors from either general or specialized units can fulfill this role. They should assess potential business benefits and risks associated with data-driven initiatives, ensuring alignment with strategic objectives.

Data engineers, data scientists, AI engineers, and software engineers contribute to the design of the solution by managing the data, developing the AI models, and ensuring compatibility with existing software. By understanding how business units operate and create value, the project manager bridges the gap between the business objectives and the solution design. The project manager's role is critical for guaranteeing the timely delivery of the project. Depending on the project's size and complexity, a team member may take on multiple roles, or additional roles may be required, such as DevOps engineers, cloud engineers, user experience engineers, and test engineers.

When specialized resources are scarce, financial sector authorities should prioritize the roles of business owner, data scientist and software engineer. In this scenario, data scientists will be responsible for preparing the data and guaranteeing its quality and usability. Software engineers would assume the responsibilities of designing the AI system used in production, while monitoring its deployment and performance. Additionally, other stakeholders, such as the Legal or Ethics Units, may participate in high-materiality cases when determining whether to deploy the models. They also play a key role in the review of risk management frameworks. Legal experts should participate to ensure that the organization's data practices align with legal obligations while minimizing legal risks. Representatives from Legal Units are critical to review contracts, when AI systems development is trusted to third parties. Ethics should contribute to the development of ethical principles and guidelines for data handling

Figure 2. Project Team



SOURCE: Authors based on MINDTITAL How to put together a machine learning team for an AI project (mindtitan.com).

Table 1. Roles and Responsibilities in an AI Project Team

Role	Responsibilities
Business owner	
	Defines the business case and explains why it is relevant
	Makes a gap analysis, describing the differences between the process today and how it will look after the project is implemented
	Select the relevant metrics and KPI that will measure the success of the AI project
	Set up the processes to keep the AI project aligned with business after deployment phase has ended
	Give access to the team to the required systems for the project
Project manager	
	Understands and explains the functioning of the business units and how they create value
	Keeps the team productive
	Explains expectations to each team member
	Plans and establishes project goals
	Set milestones and tracks progress
	Ensures timely delivery

Role	Responsibilities
Design team	
Data engineer	Prepares the data for the data scientist by analyzing and organizing raw data and creating datasets
	Is responsible for the quality and usability of datasets
	Collaborates with both data scientist and AI engineer
Data scientist	Selects the data representation methods
	Cleans the data
	Trains AI models
	Evaluates whether the results of the models align with business KPIs and technical indicators
	Visualizes and communicates the results to stakeholders by telling a story with data
	Researches and implements AI algorithms and tools
AI engineer	Focuses on building and designing AI systems used in production
	Oversees the deployment and monitors AI models
Software engineer	Develops, tests, and assures quality of software
	Conducts systems risk and reliability analysis
	Performs maintenance and integration with existing software
	Identifies and assesses new technologies prior to implementation
Representative from Ethics	Reviews risk management framework
	Contributes to draft principles /guidelines for data management
Representative from Legal	Ensures data practices align with legal obligations
	Reviews third-party agreements

SOURCE: Authors based on How to put together a machine learning team for an AI project (mindtitan.com).

III. Managing AI Projects

This section discusses the necessity of aligning AI initiatives with business objectives by defining clear use cases and performance metrics. It introduces the Cross Industry Standard Process for Data Mining (CRISP-DM) framework as a structured methodology for guiding AI project management through its iterative phases, from project foundation to deployment. It discusses the vital importance of data and model governance, addressing cybersecurity risks, data quality, and organizational culture as significant challenges that financial supervisory authorities must navigate to successfully implement AI solutions.

Methodology for managing an AI project

CRISP-DM's attributes are structured to serve several project management goals. First, managing an AI project requires a methodology adjusted to the challenges posed by big data. CRISP-DM has been widely used for managing projects based on data exploitation for over 20 years. Second, the CRISP-DM methodology is particularly useful to explain the results of the project to different stakeholders. Third, by ensuring a granular estimation of resources needed to execute the project, CRISP-DM eases monitoring and accountability.

While the methodology continues to be a valid approach to data-centered projects, it might require adjustments when data availability is limited, and business objectives are unclear. By executing an initial assessment with the D.A.T.A. framework, financial supervisory authorities will reduce the need for adjustments.

For example, let us take a situation where financial supervisory authorities cannot clearly define the business objectives at the outset of the AI project. In that case, they should include exploratory activities in the methodology to identify these objectives in a data-driven manner.[24] Data generated through simulations can be instrumental in developing an AI tool for monitoring liquidity risk. In addition, activities related to data source exploration can help discover new sources of data when availability is limited. These exploratory activities must ensure compliance and seamless integration with IT infrastructure, as well as data and AI governance requirements and specific activities related to software development.

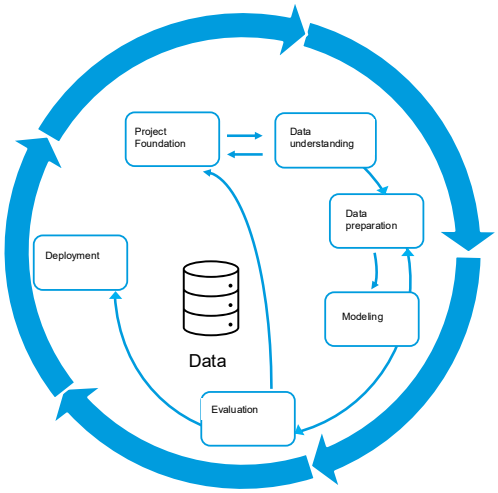
Originally, CRISP-DM consisted of six phases: (1) Business Understanding, which for this paper we have interpreted broadly and will call Project Foundation; (2) Data Understanding; (3) Data Preparation; (4) Modelling; (5) Evaluation, and (6) Deployment. However, the execution of phases is not strictly sequential. Instead, as some phases will produce findings that will require redoing part of the work, an iterative approach can seamlessly integrate the distinct phases of the project.

Agile methodologies provide for an iterative and incremental approach that eases the organization of phases and enables a quick and efficient response to challenges encountered during the project. To better reflect risk considerations and integration with IT infrastructure, the initial phase needs to expand its scope and change its denomination accordingly (Figure 4). This framework can be executed flexibly, allowing movement between different phases as needed. The arrows, indicating requirements between phases, highlight their interconnections, while the outer circle represents the framework's cyclic nature.

[24] CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories | IEEE Journals & Magazine | IEEE Xplore includes examples of different projects with activities related to goal exploration, data source exploration, and data value exploration.

Each of the six phases is described in the following subsections.

Figure 3. The six phases of an AI project



SOURCE: Authors based on Data Science Process Alliance https://www.datascience-pm.com/crisp-dm-2/).

Project Foundation

A clear and thorough statement of the business case is critical to ensure the successful implementation of an AI solution. During this phase, the project team should justify the project's contribution to the financial supervisory authorities' business objectives, namely the execution of one or more activities within risk-based supervision. This involves selecting performance metrics and assessing risks and challenges.

Next, the team must produce a project plan. Answering three questions should help keep the project aligned with business needs. What is the goal? How should results be measured? What is expected from the project?[25] The team might find it useful to search for similarities in systems deployed by other financial supervisory authorities, published research, or code repositories. For example, the Bank for International Settlements (BIS) and the Cambridge Centre for Alternative Finance periodically publish surveys of AI systems deployed by financial supervisory authorities.[26] The European Central Bank (ECB) created a suptech platform named

[25] Example of application of these questions to an AI project aiming to use NLP techniques in fit-and-proper assessments. **Which is the goal?** Authorities aim to enhance the efficiency and accuracy of fit-and-proper assessments. **How to measure its results?** The results can be measured by comparing the time taken and the accuracy of assessments before and after its implementation. Metrics could include the reduction in time spent on manual reviews, the number of issues correctly identified, and the decrease in oversight errors. **What is expected from the project?** It is expected that the deployed tool streamlines the fit-and-proper assessment process through automatic machine reading of banks' questionnaires, supported by textual analysis, process automation and data analytics capabilities. This will improve the quality of assessments by reducing manual workload and enabling supervisors to focus on more complex cases.

[26] IFC (2024), Cambridge Suptech Lab (2023)

"Virtual Lab," which is a cloud-based system designed to facilitate remote collaboration between the ECB and national regulators. This platform will enable secure data sharing and project collaboration, utilizing advanced AI, machine learning, and deep learning to foster a culture of innovation and data-driven decision-making (ECB 2024).

While exploiting data is a key characteristic of AI projects, it is important that the team does not propose vague and broad definitions such as 'let's do something with all our data' or implement AI.' Additionally, the team should have clarity on the questions that the project will not answer and whether follow-up initiatives will be necessary for deeper analysis. Furthermore, financial supervisory authorities should aim "low at first" selecting projects with a modest scope. Those projects will help to build a learning curve and gain trust, shaping more ambitious implementations in the future.

Performance metrics play a key role when deciding whether to deploy an AI solution. Metrics will include Key Performance Indicators (KPIs) and technical indicators, aiming to assess whether the project achieved business objectives and evaluate AI model performance, respectively. Examples of KPIs in AI projects aiming at measuring efficiency include: an increase by a certain percentage in daily loan files reviewed, or a minimum reduction in the average of days required to process fit-and-proper documents. To monitor improvement in effectiveness, financial supervisory authorities can compare, for example, the output of an AI model against the outcome of inspections. Statistical measures such as accuracy, precision, recall, or root mean squared error are commonly used as technical indicators.[27] When selecting these technical indicators, the team should consider the type of AI model to deploy and their potential limitations for specific business cases. For example, measures such as accuracy, precision and recall are relevant for classifying potentially defaulting debtors.

Assessing implementation challenges is a key activity in the Project Foundation phase. Initial results of the D.A.T.A. framework should be complemented by a thorough assessment of human resources, IT infrastructure, and cybersecurity requirements.[28] Data and model governance frameworks reflecting the specificities of AI models are vital to deploy solutions that are explainable, robust and prevent discrimination. Funding and organizational culture are critical to AI deployment and should be carefully assessed. The rigid nature of public budgets may not be fully adapted to the iterative nature of AI projects, and the organization may present resistance to new solutions.

Implementing AI solutions requires a multidisciplinary team, a conducive IT infrastructure, and a compliance function aligned with cybersecurity best practices. Authorities often invoke the lack of staff with data analytics and IT skills as factors hindering the broader adoption of suptech tools.[29] The team responsible for the deployment should exhibit expertise in AI, IT, and domain specific skills corresponding to the role assigned in the project.

Following recommendations from the AI governance framework, other stakeholders, such as auditors and compliance officers, should also have a sufficient level of understanding of the workings of highly complex models to audit, oversee, challenge, and approve their use. Senior managers should understand and follow the development and implementation of AI models as they may be accountable for their use.

[27] Data Science Performance Metrics for Everyone | by Andrew Long | Towards Data Science; Art of Choosing Metrics in Supervised Models Part 1 | by Saeed Garmsiri | Towards Data Science

[28] For a thorough description of IT legacy infrastructure in financial supervisory authorities see *World* Bank (2020).

[29] Cambridge Center Suptech Lab (2023).

Financial supervisory authorities can address limitations in skilled workforce with a multi-pronged strategy. They can enable the flow of talent and education across the organization by combining training with internship programs between financial supervisory authorities' functions. Additionally, they could design training programs that cater to all stakeholders involved in deploying AI models. For example, a basic programme may be sufficient for business owners, while a more technical and detailed programme is needed for specialist staff, such as those involved in developing AI models (BdE 2023). Other alternatives include recruiting staff with data science skills and partnering with external experts to provide training while internal capacities are developed (World Bank, 2025). Over a long-term horizon, financial supervisory authorities should consider recognizing a data analytics career path to foster retention and ensure continuity in developing AI solutions.

Inadequacy of IT infrastructure also limits countries' capabilities to profit from frontier technologies such as AI. Financial supervisory authorities may experience other technological difficulties, such as legacy systems not prepared to incorporate AI models, lack of the necessary processing power or storage capacity to train large datasets, dependence on cloud service providers, and IT systems designed to work in silos. Ideally, the underlying technology to conduct advanced analytics should include three components: infrastructure, data platform and data processing (Figure 5).[30]

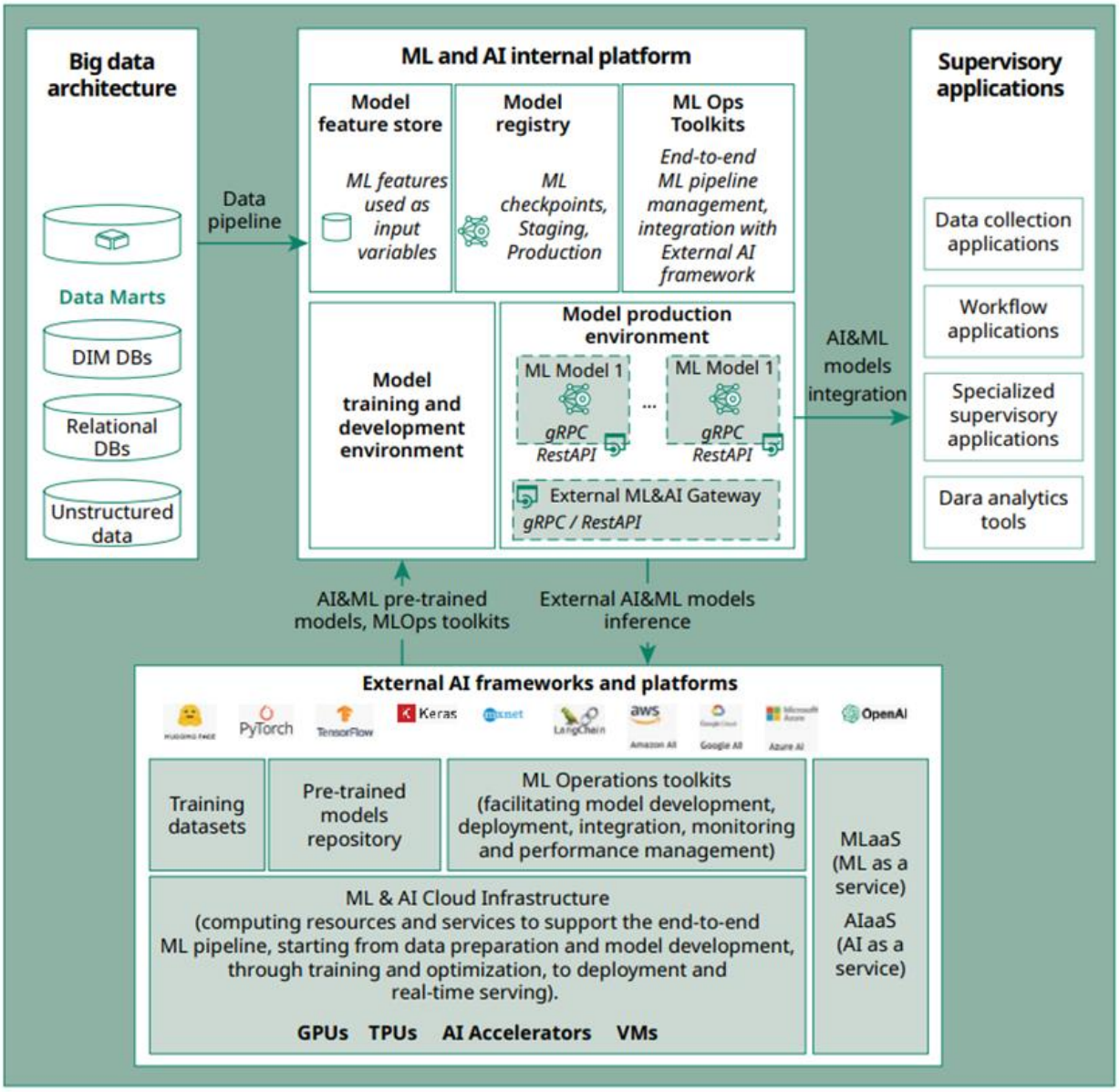
To address infrastructure challenges, supervisory authorities can leverage cloud-based AI platforms. Cloud solutions offer specialized AI infrastructure, including scalable computing resources, dedicated machine learning platforms, and pre-built models. These cloud solutions enable organizations to train and deploy AI models on demand, eliminating the need for substantial initial investments in physical infrastructure. The decision to deploy AI solutions in the cloud needs to be based on the criticality of the applications. Non-critical AI solutions are typically cloud deployed and in the case of critical applications whether to fully deploy in cloud or in a hybrid arrangement or on-premises solution will depend on multiple factors including technical complexity, IT architecture, computing resources required, sensitivity and confidentiality of data used, availability of skills within the organization, regulatory environment and legal requirements.

One of the challenges in adopting cloud for AI deployment is stickiness – it becomes difficult to move AI solutions developed using the provisions of one cloud service provider to another environment. Multi-cloud arrangements are also gaining ground to avoid over dependence.[31] Whether financial sector authorities decide to deploy AI solutions in their premises or in the cloud, cybersecurity requirements must be contemplated in the plan, reflecting their impact on the various stages of the project.

[30] NIST, NIST Big Data Interoperability Framework: Volume 6, Reference Architecture, June 2018, p. 16 (https://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-6r1.pdf).

[31] For further guidance on implementing cloud, see ABS (2024)

Figure 4. Technological infrastructure for advanced analytics



SOURCE: Dohotaru et al (2025)

Public organizations exhibit unique characteristics when it comes to funding and culture. AI solutions introduce changes in financial supervisory authorities' processes and may face resistance. The team should address risks of AI solutions that do not seamlessly integrate within the supervisory process or fail to incorporate the qualitative aspect of supervisory judgment. AI solutions should be viewed as enhancing supervisory capabilities rather than as tools for reducing staff numbers.

To mitigate resistance, the project plan should include a clear communication of AI benefits and ensure knowledge transfer among team members. While the methodology is flexible enough to operate under different

organizational structures, such as a dedicated Project Management Office, an Agile Team, or data scientists working in supervisory units, it is important that the AI project plan contemplates that diversity and assesses its risks. Public budgets are oriented to the execution of projects where there is certainty on the deployment date and the configuration of the deliverable solution. Aiming to meet transparency requirements, projects funded with fixed budgets may impose restrictions on the iterative nature of AI projects that the team needs to assess. A fine-tuned estimation of project resources helps mitigate this challenge.

The Project Foundation phase concludes with a detailed project plan for subsequent phases. As a result of this phase, the team produces a plan that clearly defines the business problem that the AI project intends to solve, the criteria for verifying the achievement of objectives, a challenge assessment, technical and functional requirements, and resource estimation.

Data Understanding

The Data Understanding phase focuses on identifying, collecting, and analyzing the data relevant to the business problem at hand. While proprietary data might seem abundant, some projects may require collecting additional data from external sources. For example, data scientists could profit from diverse data sources such as transactional data, employment history, utility payments and other relevant data to complement data from credit bureaus, thereby enhancing the prediction of the probability of default.

Typically, data relevant to solving business problems does not conform to a unified structure and exists in silos. Using siloed data exposes financial supervisory authorities to data bias that could affect model outcomes. As a best practice, these authorities should consider developing a unified data repository prior to the development of AI models. This practice might ease future AI projects and help implement recommendations from data governance frameworks as repositories centralize data access, ownership, stewardship, data ethics, and metadata. Additionally, collecting diverse data sets might help financial service authorities reduce biases arising in underrepresented data.

Collecting data from external sources might raise privacy concerns. When exercising their supervisory powers, financial supervisory authorities regularly collect personal data that does not require explicit consumer consent. However, deploying AI models might require collecting large amounts of data from non-traditional sources such as Internet of Things devices, smartphones, and web tracking means whereby individuals might not be fully aware of providing personal information. These concerns can affect, for example, the estimation of banking customers' probability of default or the identification of mis-selling practices in the insurance sector. When third parties develop models, financial supervisory authorities may inadvertently share information that could lead to the identification of individuals or be used for purposes beyond the original intent.[32] In addition, when third parties develop models using proprietary data, financial supervisory authorities might not be able to verify whether they collected data properly.

Various techniques aligned with data governance best practices may help financial supervisory authorities address data privacy concerns. Anonymization is the process of removing personal identifiers that could lead to the identification of an individual. However, significant privacy risks remain.[33] EU GDPR defines pseudonymization as "the processing of personal data in such a way that the data can no longer be attributed

[32] A2ii et al (2022)

[33] Sweeney_Article.pdf (epic.org)

to a specific subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organizational measures to ensure non-attribution to an identified or identifiable individual." Data masking substitutes confidential data, keeping structure and format, and guaranteeing that reverse engineering does not allow tracking back original data values. Also, privacy technologies might help financial supervisory authorities to share data with external developers. [34]

Conducting data quality checks and removing correlated features leads to identifying key variables of a problem. Once data privacy issues have been solved, data should be described by examining the format (structured, semi-structured, unstructured, synthetic) and identifying relevant features. Next, by exploring the data, data scientists should articulate the relationship between data and visualize it with a clear focus on answering the business question.[35] Multicollinearity and high-correlated features are usually removed in this phase. However, they may be preserved for further exploration later in the project. Lastly, data quality should be verified by checking whether it meets quality thresholds. Most present data quality issues are missing data, typographical errors when entering data, measurement errors, coding inconsistencies, and bad metadata.

Data Preparation

The Data Preparation phase prepares the dataset for modeling. This phase involves data selection, cleaning, feature engineering, integration, and formatting. It begins by selecting the relevant dataset along with their respective columns and rows. Data cleaning primarily involves determining how to handle missing data (by ignoring, filling blanks manually, or using computed values), noisy data (through binning, clustering, applying an algorithm, or manual removal), and inconsistent data. Data cleaning is typically performed once at the beginning of the project by either data scientists or data engineers. Feature engineering involves selecting the attributes relevant for building the AI model. Additionally, new attributes can be created by combining existing data, e.g., combining two or more datasets to enrich the original. During this phase, data scientists/data engineers should adhere to the previously established steps for cleaning and preparing data. The final step involves formatting the data according to the requirements of the AI model. Examples of actions during this step include converting categorical values into numerical and vice versa. The team should document decisions made during this phase, as this documentation may help explain potential inconsistencies in subsequent phases.

Modeling

During modeling, data scientists estimate and technically validate the AI models designed to address the business problem. AI models should be clearly defined regarding their intended purpose and scope. They often have dynamic and complex behaviors, necessitating a broader understanding of their operational context. The nature of the business problem guides the selection of the appropriate family of models.

Clustering is suitable for grouping items by similarity; regression is used for predicting numeric outcomes by associating features with outcomes; classification assigns labels, such as suspicious activity or spam mail; time

[34] Privacy technologies, based on the concept of privacy by design, refers to a set of technologies that aim to protect sensitive data and maintain the confidentiality and integrity of data, while also allowing use of that data. (AlAjmi, 2021).. There are several privacy technologies that could be used in relation to AI, the most promising of which is federated learning. Federated learning operates on the notion that the training of a model occurs in a decentralized manner with only model updates or summary results being shared back to central servers (Bains and Gaidosch, 2025).

[35] See as examples of visualizations: 0.2a Embedded AI — Exploratory Data Analysis (EDA) | by David Such | Jul, 2024 | AI Advances (gopubby.com)

series analysis identifies temporal structure in data; and textual analysis encompasses various tasks such as translation, document representation, bag of words, and text summarization. Each family may include one or more models, and ideally, data scientists should generate various alternatives within and across these families. For example, models such as regression trees, random forest and artificial neural networks are suitable for estimating probability of default.[36] Clustering, neural networks, logistic regression, and random forest can support the modelling of suspicious transactions (FATF 2021).

Code repositories provide access to numerous pre-trained models, which can facilitate the modeling process by reducing the need to build models from scratch. Next, data scientists should partition the data into separate disjoint sets for training and testing the models, ensuring the test design is justified. Once the model is estimated, metrics selected during the Project Foundation phase and calculated on the test set will serve to validate the model technically.

When modeling, data is often partitioned into three disjoint sets: training, validation, and testing sets. It is crucial that these sets belong to the same underlying distribution; otherwise, the model may not generalize effectively. The training set will feed the selected algorithm to estimate the AI model parameters. Models are often trained iteratively, with a performance measure calculated in each iteration to reflect the model's error. This measure is used to update model parameters and reduce the estimation errors when applied to the training dataset. More complex models, such as deep learning models, include parameters (known as hyperparameters) that are not estimated but calibrated to minimize the model's estimated error. The optimal value of the hyperparameters is selected using the validation set. Techniques such as data augmentation, dropout regularization, early stopping, and gradient checking are employed to calibrate hyperparameters, thereby enhancing model performance beyond the training sample.[37] Data scientists can use ensemble models in machine learning to combine predictions from multiple models, improving overall performance (ensemble methods). However, this often results in a trade-off with explainability.

Data scientists should select a statistical validation technique tailored to the problem's specific requirements, the size and complexity of the dataset, and the desired level of confidence in the model's performance. The ultimate objective is to ensure that the model performs well on new data. Various types of sampling are regularly applied to validate models.

- In the **holdout method**, the training set is used to learn the model parameters, while the validation set is utilized to calibrate the hyperparameters and fine-tune the model. After the model is estimated and fine-tuned, the test dataset calculates the generalization error (i.e., the error resulting from applying the model to new data). This method is easy to implement but may not be appropriate for small datasets, datasets with imbalanced classes (for example, suspicious and non-suspicious transactions), or complex models.
- **Cross-validation** involves in splitting the dataset into k subsets. Iteratively one dataset is used as validation set and the remaining k-1 subsets are used for training. The process is repeated k times, which each subset used once as the validation set. The average of performance measures across the k iterations is used as the validation error. While more computationally intensive, cross-validation

[36] Petropoulos et al (2018) provide an example of the use of machine learning models for estimating probability of default.

[37] A comprehensive description of these techniques is beyond the scope of this document. For further details, please refer to Machine Learning Mastery.

provides a more reliable estimate of the model performance than the holdout method. This is the preferred method when training deep learning models.

- **Bootstrap sampling** involves randomly selecting samples from the dataset with replacements. Multiple bootstrap samples are used to train and test the model, and the results are averaged. This method can help to reduce the variance of the performance estimates. It could be used, for example, for models aiming to detect fraud in a credit card database.
- **Stratified sampling** implies sampling the dataset in a way that ensures that each class is represented equally in the training and the testing sets. It is useful for datasets with imbalanced classes. Financial supervisory authorities could apply stratified sampling, for example, when building models to predict credit default.
- In **Rolling time series cross validation** rolling train set and every validation set are the same size, giving more balance to how each fold becomes weighted when averaging the final error. This method is recommended to validate time series models as it prevents skewed errors and respect the structural time dependence.[38]

Encouraging critical assessment of model outputs is vital. When practical, data scientists should compare AI model outputs against challenger models. These models receive the same inputs as the estimated model, but their predictions are logged for further analysis. For example, Financial Transactions and Reports Analysis Centre of Canada (FINTRAC) is developing a challenger model using machine learning and heuristic methods and comparing its performance against that of the model in production for overseeing AML/CFT (Coelho 2019).

Data scientists should maintain thorough documentation of AI models estimated and validated during this phase. Guided by principles of rigorous documentation, data accuracy, integrity, and quality assessment, data scientists should record design choices, data sources, validation techniques, and decision-making process. Documentation will contribute to ensure transparency, facilitating reproducibility, auditing and knowledge transfer across the organization.

Evaluation

Model performance must be both technically and functionally robust. The Evaluation phase focuses on assessing the model's performance in relation to the business problem and its objectives. Data scientists should consider the KPIs selected during the Project Foundation phase and also other critical requirements, such as regulatory constraints.

To ensure alignment with ethical and legal requirements, the evaluation must assess whether model predictions include bias. By applying metrics and tools, data scientists can detect biased predictions and understand the type of bias (e.g., disproportionally selecting large institutions or underrepresenting some fraud modalities). When choosing metrics, it is crucial to consider the context, as different metrics are appropriate for various situations. Additionally, combining multiple metrics can provide a more comprehensive understanding of potential biases. Metrics often used in bias detection include:

[38] For further details, please refer to Model Validation Techniques for Time Series | Towards Data Science. Adapting Model Valuation in the Age of AI provides a sample of validation techniques in Large Language Models, a subset of GenAI.

- **Confusion Matrix Analysis:** representing counts from predicted and actual values (true positives, true negatives, false positives, and false negatives), the matrix helps to assess how the model's performance varies across distinct groups.
- **Disparate Impact** measures the ratio of positive outcomes for a minority group to that of a majority group.
- **Statistical Parity Difference** checks whether distinct groups have equal probabilities of being correctly classified.
- **Equality of Opportunity** requires equal outcomes only within the subset of records belonging to the positive class. For example, equality of opportunity requires that the individuals in group A who are qualified to be hired are just as likely to be chosen as individuals in group B who are qualified to be hired.
- **Predictive Equality** measures that the false positive rates are equal across distinct groups.
- **Calibration** measures if the predicted probabilities of an outcome are accurate across distinct groups.

Another key aspect to assess during the Evaluation phase is explainability, where prioritizing simpler models is essential, even if it comes at the expense of predictive power. The level of explainability in an AI system depends on factors such as what needs to be explained, who needs the explanation, the materiality of the use case, and the complexity of the model. Financial supervisory authorities should consider whether the objective is to explain the significance of each variable in the model, how the entire model works or how it contributes to the final decision (i.e., automated or assisted).
While explainable AI methods (e.g. LIME, Shapley) might help understand the inner workings of AI models, they should not be perceived as a panacea. Data scientists should check whether the methods provide clarity, simplicity, broadness, completeness, and soundness. Those desirable properties might be challenged when results are complex to interpret, high computational costs are required, explanations are inconsistent and not robust across different instances or involve assumptions that might not hold in all cases. [39]

The recipient of the data is a key factor in determining the required level of explanation. An explanation that may suffice for a head of supervision may be insufficient for a data scientist or a business owner. When stakeholders with various levels of expertise are involved, data scientists should always prefer simpler models as opposed to explainable methods. Higher levels of explainability are required for high-materiality applications such as stress testing models. When explainability is hardly attainable given the intrinsic complexity of the model as it is the case for deep neural networks, financial supervisory authorities should consider whether their use is appropriate.

When summarizing the results of the phase, data scientists should articulate how the model benefits the supervisory process to an audience of business owners and other stakeholders who may lack data science expertise. Reviewing the work completed in previous phases can help identify any missing activities or necessary corrections. The phase concludes with a decision regarding whether the model can proceed to the next phase, requires adjustments, or necessitates initiating a new project.

[39] For details on Explainable methods see Boukherouaa et al. (2021). Explainable Generative AI (GenXAI): A Survey, Conceptualization, and Research Agenda describes explainable methods for GenAI models.

Deployment

Integrating the AI model within the existing IT infrastructure and developing a plan for monitoring and maintenance are essential components of the deployment phase. Key questions that should guide the deployment phase include: Who is going to use the model? For what purpose? How is it going to be used? Does the model run in real time? And does the model need to be rerun with new data?

It is crucial for the AI application to integrate with other supervisory tools seamlessly. Validation and performance monitoring is crucial for AI models to assess their robustness against unexpected inputs. Additionally, models can deteriorate over time as new data may be drawn from a different distribution than that used during training. For example, training data used to build a suspicious activity detection model six months ago may not account for new types of activities that have emerged in the last three months, given criminals adapt their behavior to avoid detection (Coelho 2019). To monitor model performance, data scientists can apply generic statistical metrics that measure the distance between probability distributions such as Kolmogorov-Smirnov or Wassertein distances (Treveil 2021). The team should propose a comprehensive monitoring and maintenance plan to ensure that the model continues to produce robust results with new data. Additionally, business owners should evaluate whether AI models continue to serve the supervisory process and if the benefits of the model outweigh the cost of development and deployment.

For monitoring purposes, the selected model should update an inventory of deployed models, providing varying levels of information based on the complexity and the overall adoption of models by financial supervisory authorities. In high-impact cases such as stress testing models, the inventory should include its purpose and use, limitations and assumptions, findings from validation as well as data related to governance such as who conducted the validation and when.

To ensure the AI model is accessible to end users, the project team should consider the governance structure and demonstrate alignment with the selected objectives of the supervisory process. The project team should explain the main conclusions of the project and the process leading to them. Results should be documented in a summary report, as a narrative that illustrates alignment with business objectives, highlights, and value created for the organization and proposes actionable recommendations. The report should summarize the conclusions of the Modeling and Evaluation phases, stating the model purpose, synthesizing results, and detailing major limitations and key assumptions.

Furthermore, the project team should recognize that the adopted AI governance structure will influence who participates in the deployment decision. Lastly, this phase requires a final review aimed at evaluating what went well, what could have been better, and how to enhance future efforts. Recent advancements in AI project management emphasize the importance of building up data pipelines in accordance with MLOps practices during this phase.

How the methodology can contribute to the supervisory process

The AI project management described in this paper can also be utilized by supervisory authorities to evaluate whether supervised entities comply with regulatory requirements related to model risk management, as well as AI and data governance. A thorough understanding of the distinct phases of an AI project and the corresponding governance framework will enable financial supervisory authorities to assess whether supervised entities have deployed these solutions in accordance with international standards, local regulations, and best practices.

The deployment of AI solutions by financial supervisory authorities will help ensure that their skills and supervisory practices remain up to date while counting on the technological resources to process information from supervised entities efficiently.[40] Secondly, by adopting AI solutions, financial supervisory authorities can not only streamline information processing but also significantly enhance their supervisory skills in areas such as data and model governance. Financial sector authorities can leverage best practices in data governance to better assess whether supervised entities have adequate controls on data quality throughout its lifecycle and whether the IT infrastructure supports data aggregation and risk reporting.[41]

Experience gained from establishing an AI governance structure can assist financial supervisory authorities in evaluating board's understanding of AI solutions, their limitations, and the associated risks of deployment. Additionally, by implementing model validation techniques, financial supervisory authorities will be better equipped to assess models' outcomes and their behavior under normal and stressed conditions in supervised entities. Fourthly, implementing bias mitigation strategies in deployed solutions can help financial supervisory authorities ensure that consumers receive fair treatment in AI-priced products. Finally, supervisory authorities can assess whether supervised entities maintain comprehensive documentation of AI solutions, including design choices, data sources, and decision-making processes.

[41] Principles for effective risk data aggregation and risk reporting (bis.org)

IV. Conclusions

AI solutions can assist financial supervisory authorities in keeping up with the rapidly digitalizing financial sector. By efficiently processing vast amounts of data from diverse sources and detecting hidden patterns, AI solutions can enhance the capabilities of financial supervisory authorities. Freed from routine tasks such as compiling, validating, and summarizing information, authorities can focus on the key value-added activity of exercising supervisory judgment.

Ongoing trends in adoption suggest that financial sector authorities will need to further incorporate AI into their supervisory kit. For example, AI can transform capital markets by making them more efficient but also more volatile and financial sector authorities must prepare for this new world (IMF 2024). AI can facilitate activities such as real-time monitoring, process automation, and forward-looking modelling, thereby increasing the efficiency and effectiveness of the supervisory process.

An AI project represents a complex interplay between technology, processes and people, requiring careful reflection, planning and resources. Financial supervisory authorities should consider whether they possess the necessary IT infrastructure and the human resources to deploy an AI solution aligned with their supervisory and ethical goals before undertaking any AI project. Beginning with small-scope projects will enable financial supervisory authorities to establish a learning curve, which is essential for larger projects. Knowledge transfer is a critical component of planning, regardless of whether projects are developed in-house or outsourced. The project team should comprise a diverse composition of representatives from various functions across the organization, possessing a range of skills. An iterative approach that enables financial supervisory authorities to assimilate lessons learned during the project's execution quickly is pivotal.

Successfully implementing AI projects within financial supervisory authorities depends on several key components. Selecting a business case aligned with the objectives of RBS is crucial for ensuring organizational buy-in. AI solutions should augment supervisory capabilities, allowing them to focus on tasks where supervisory judgment is critical. Adopting data and AI governance frameworks will ensure that AI models comply with legal and regulatory requirements, align with organizational objectives, and uphold societal values while providing appropriate checks and balances throughout the model lifecycle. Depending on the materiality of the use case, numerous stakeholders with different skills may be involved in the decision to deploy an AI solution. While each financial sector authority is unique, legal, ethical, technical, cultural, and business perspectives should be factored in to warrant a responsible adoption of AI.

Additionally, AI solutions introduce new sources of cyber risk; therefore, cybersecurity must be considered from design to deployment. Financial supervisory authorities should also evaluate the challenges associated with developing or deploying AI solutions that involve third parties. Good practices in third-party risk management—such as policies, processes, risk management activities, due diligence, ongoing monitoring, independent assurance, exit plans, business continuity arrangements, and disaster recovery arrangements—should be in place.

Financial supervisory authorities should strengthen their foundational elements by implementing risk management practices tailored to the unique challenges of AI applications. This includes ensuring an appropriate level of explainability based on the use case's materiality and legal constraints, while favoring interpretable solutions. To combat discrimination, robust data collection, diverse AI development teams, and

continuous monitoring are essential. Authorities must also provide training on data privacy and governance to address skill gaps and may need to reorganize teams to leverage IT expertise. Evaluating data repositories for their capacity to handle unstructured data and ensuring legacy systems have sufficient processing power is crucial, with infrastructure updates typically preceding AI model deployment. Finally, applying privacy techniques can facilitate compliance with data protection laws when sharing data with third parties.
The final component for successful implementation is a data-centric project management methodology incorporating risk mitigating techniques. Many data-driven initiatives fail due to decision-makers' inability to assess project risks accurately. Consequently, a clear definition of the use case aligned with business objectives is essential to prevent failures. Subsequent stages will involve proper data treatment, optimizing the AI model, and assessing its outcome from technical and business perspectives. Deployment should proceed only if the model meets the objectives set in the initial project stage. The deployment stage should encompass monitoring and maintenance to mitigate the risk of decaying performance.

While the availability of AI tools for streamlining processes in financial supervisory authorities has advanced significantly, emerging markets and developing economies still lag in their adoption. For example, novel forms of AI can improve the efficiency of information processing and facilitate code generation, thereby mitigating the risk of insufficient engineering skills. However, deploying them requires significant resources and careful consideration of their risks. While an appropriate project management methodology can assist in deploying AI solutions aligned with supervisory objectives, organizational buy-in, and adequate human and IT resources are essential.

References

Access to Insurance Initiative (A2ii), the Financial Stability Institute (FSI) of the Bank for International Settlements (BIS) and the International Association of Insurance Supervisors (IAIS), 2022: "A2ii FSI IAIS Joint note on SupTech in insurance supervision", https://www.iais.org/uploads/2022/12/A2ii-FSI-IAIS-Joint-note-on-SupTech-in-insurance-supervision.pdf

Adrian, T., M. Moretti, A. Carvalho, H. Chon, F. Melo, K. Seal, and J. Surti, 2023, "Good supervision – lessons from the field" IMF Working Paper 23/181, International Monetary Fund, Washington, DC.

Ahmed, H., 2022, "Developing an Artificial Intelligence Governance Framework", https://www.isaca.org/resources/news-and-trends/newsletters/atisaca/2022/volume-38/developing-an-artificial-intelligence-governance-framework

AlAjmi, K., E.B. Boukherouaa, J. Deodoro, A. Farias, E. S. Iskender, A. Mirestean, R. Ravikumar, G, Shabsigh, 2021. "Powering the Digital Economy: Opportunities and Risks of Artificial Intelligence in Finance" IMF Departmental Paper 21/024, International Monetary Fund, Washington, DC.

Amazon AWS (2024): "What is DevOps - DevOps Models Explained?", https://aws.amazon.com/devops/what-is-devops/

——— (2024): "What is MLOps? - Machine Learning Operations Explained", https://aws.amazon.com/what-is/mlops/

Association of Banks of Singapore (ABS), 2024: "ABS Cloud Computing Implementation Guide 3.0 for the Financial Industry in Singapore", https://www.abs.org.sg/docs/library/abs-cloud-computing-implementation-guide---july-2024.pdf

Bains, P. and Gaidosch, T, 2025. "Privacy Technologies and the Digital Economy". Working Paper No. 2025/060, International Monetary Fund, Washington, DC. https://www.imf.org/en/Publications/WP/Issues/2025/03/28/Privacy-Technologies-The-Digital-Economy-565415

Bank for International Settlements (BIS), 2025: "Governance of AI adoption in central banks", Consultative Group on Risk Management, January, https://www.bis.org/publ/othp90.pdf

Boukherouaa, E. B. and G, Shabsigh, 2023. "Generative Artificial Intelligence in Finance: Risk Considerations" IMF Fintech Note 23/006, International Monetary Fund, Washington, DC.

Cambridge Suptech Lab, 2023, State of Suptech Report 2023, Cambridge: University of Cambridge. Available at www.cambridgeSuptechlab.org/SOS

Broeders, D. and J. Prenio, 2018: "Innovative technology in financial supervision (suptech) – the experience of early users", FSI Insights on policy implementation, no 9, July, https://www.bis.org/fsi/publ/insights9.htm

Cerutti, E., A. Garcia Pascual, Y. Kido, L. Li, G. Melina, M. M. Tavares, and Ph. Wingender, 2025, "The global impact of AI: mind the gap" IMF Working Paper 25/076, International Monetary Fund, Washington, D.C.

Coelho, R., M. De Simoni and and J. Prenio, 2019: "Suptech applications for anti-money laundering", FSI Insights on policy implementation, no 18, August, https://www.bis.org/fsi/publ/insights18.pdf

Denis, E., 2021, "The promises and pitfalls of SupTech for corporate governance-related enforcement", *OECD Going Digital Toolkit Notes*, No. 10, OECD Publishing, Paris, https://doi.org/10.1787/9f0b8883-en.

Di Castri, S., S. Hohl, A. Kulenkapmpff and J Prenio, 2019: "The suptech generations", FSI Insights on policy implementation, no 19, October, https://www.bis.org/fsi/publ/insights19.htm

Dohotaru, M., M. Prisacaru., J. H. Shin and Y. Palta, 2025, "AI for risk-based supervision – Another "Nice to Have" Tool or a Game-Changer", https://documents1.worldbank.org/curated/en/099021725190042788/pdf/P503970-2c607262-8673-4797-b5a6-52df7187e543.pdf

European Banking Authority (EBA), 2020, "Report on big data and advanced analytics", https://www.eba.europa.eu/sites/default/files/document_library/Final%20Report%20on%20Big%20Data%20and%20Advanced%20Analytics.pdf

European Central Bank (ECB), 2024, "From data to decisions: AI and supervision", Article by Elizabeth McCaul, member of the Supervisory Board of the ECB, for Revue Banque, European Central Bank (ECB), 2024, "From data to decisions: AI and supervision", Article by Elizabeth McCaul, member of the Supervisory Board of the ECB, for Revue Banque, https://www.bankingsupervision.europa.eu/press/interviews/date/2024/html/ssm.in240226~c6f7fc9251.en.html

Financial Action Task Force (FATF), 2021, "Opportunities and challenges of new technologies for AML/CFT", https://www.fatf-gafi.org/content/dam/fatf-gafi/guidance/Opportunities-Challenges-of-New-Technologies-for-AML-CFT.pdf.coredownload.pdf

Financial Stability Board (FSB), 2017: "Artificial intelligence and machine learning in financial services", https://www.fsb.org/2017/11/artificial-intelligence-and-machine-learning-in-financial-service/

——— (2024): "The Financial Stability Implications of Artificial Intelligence", https://www.fsb.org/2024/11/the-financial-stability-implications-of-artificial-intelligence/

International Monetary Fund. 2024. Global Financial Stability Report: Steadying the Course: Uncertainty, Artificial Intelligence, and Financial Stability. Washington, DC. October.

Irving Fisher Committee on Central Bank Statistics (IFC) (2024): "Granular data: new horizons and challenges", IFC Report, no 61, https://www.bis.org/ifc/publ/ifcb61.htm

——— (2025): "Governance and implementation of artificial intelligence in central banks", IFC Report, no 18, https://www.bis.org/ifc/publ/ifc_report_18.pdf

Lund Pedersen, C., & Ritter, T. (2020). "Use This Framework to Predict the Success of Your Big Data Project". *Harvard Business Review Digital Articles*. https://hbr.org/2020/02/use-this-framework-to-predict-the-success-of-your-big-data-project?ab=hero-subleft-2

Maleki F, Muthukrishnan N, Ovens K, Reinhold C, Forghani R., (2020), "Machine Learning Algorithm Validation: From Essentials to Advanced Applications and Implications for Regulatory Certification and Deployment", https://pubmed.ncbi.nlm.nih.gov/33038994/

Monetary Authority of Singapore (MAS), 2020, "Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector",

https://www.mas.gov.sg/~/media/MAS/News%20and%20Publications/Monographs%20and%20Information%20Papers/FEAT%20Principles%20Final.pdf

Office of the Superintendent of Financial Institutions (2023), "Financial Industry Forum on Artificial Intelligence: A Canadian Perspective on Responsible AI", https://globalriskinstitute.org/mp-files/financial-industry-forum-on-artificial-intelligence-a-canadian-perspective-on-responsible-ai.pdf/

Organization for Economic Cooperation and Development (OECD), 2021, "Tools for trustworthy ai - A framework to compare implementation tools for trustworthy ai systems", OECD Digital Economy Papers N0. 312

Personal Data Protection Commission of Singapore, 2020, "Model Artificial Governance Framework (Second Edition)", https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf

Petropoulus A., V. Siakoulis, E. Stavroulakis and A. Klamargias, 2018, "A robust machine learning approach for credit risk analysis of large loan level datasets using deep learning and extreme gradient boosting", Ninth IFC Conference on "Are post-crisis statistical initiatives completed?", https://www.bis.org/ifc/publ/ifcb49_49.pdf

Prenio J. (2024): "Peering through the hype – assessing suptech tools' transition from experimentation to supervision", FSI Insights on policy implementation, no 58, June

Prenio, J. (2025): "Starting with the basics: a stocktake of gen AI applications in supervision", FSI Briefs, no. 26, June

Thompson, S. (2023), "Managing Machine Learning Projects – From design to deployment"

Treveil, M. & the Dataiku Team (2021), "Introducing MLOps – How to Scale Machine Learning in the Enterprise"

UNCTAD (2021), "Digital Economy Report 2021 - Cross-border data flows and development: For whom the data flow", https://unctad.org/system/files/official-document/der2021_en.pdf

Verhulst, S. and F. Schüür (2023), "Six reasons why Data Governance is the bedrock for AI Governance", UNICEF, https://data.unicef.org/data-for-action/interwoven-realms-data-and-ai-governance/#:~:text=This%20essay%20delves%20into%20the%20intertwined%20nature,unattainable%20without%20a%20comprehensive%20and%20robust%20framework.

World Bank. A Roadmap to SupTech Solutions for Low-Income (IDA) Countries. Washington, DC: World Bank, 2020. https://documents1.worldbank.org/curated/en/108411602047902677/pdf/A-Roadmap-to-SupTech-Solutions-for-Low-Income-IDA-Countries.pdf

World Economic Forum prepared in collaboration with the Info-communications Media Development Authority of Singapore, 2020, "Companion to the Model AI Governance Framework – Implementation and Self-Assessment Guide for Organizations", https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGIsago.pdf

PUBLICATIONS