Prosperity Insight Series

AI FOR RISK-BASED SUPERVISION

Another "Nice to Have" Tool or a Game-Changer

Matei Dohotaru, Marin Prisacaru, Ji Ho Shin and Yasemin Palta

WORLD BANK GROUP
FINANCE FOR DEVELOPMENT
Technology Innovation

AI FOR RISK-BASED SUPERVISION

Another “Nice to Have” Tool or a Game-Changer

January 2025

WORLD BANK GROUP
FINANCE FOR DEVELOPMENT
Technology Innovation

1818 H Street NW
Washington DC 20433
Telephone: 202-473-1000
Internet: www.worldbank.org

Cover image: Veronica Gadea, GCS, World Bank Group.
Accessibility: Will Kemp, GCS, World Bank Group.

Contents

Abbreviations

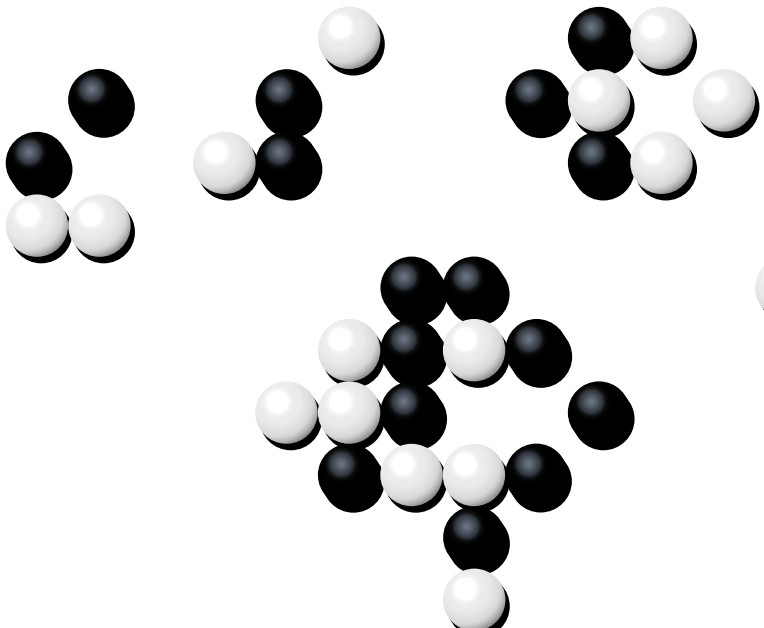
AI	Artificial Intelligence
ASIC	Australian Securities and Investments Commission
ASX	Australian Securities Exchange
ECB	European Central Bank
MAI	The Market Analysis and Intelligence
LLMs	Large Language Models
AML/CFT	Anti-Money Laundering and Combating the Financing of Terrorism
FSAP	Financial Sector Assessment Program
BCP	Basel Core Principles for Effective Banking Supervision
IMF	International Monetary Fund
SupTech	Supervisory Technology
FinTech	Financial Technology
RegTech	Regulatory Technology
NLP	Natural Language Processing
ML	Machine Learning
OCR	Optical Character Recognition

Acknowledgements

This Paper was written by the Financial Stability and Integrity Unit (EFNFS), part of the World Bank's Finance, Competitiveness, and Investment (FCI) Global Practice, and ITS (Information and Technology Solutions) Technology and Innovation Lab. It was prepared by Matei Dohotaru, Marin Prisacaru, Ji Ho Shin, and Yasemin Palta

Peer reviewers: 1) Harish Natarajan, Practice Manager, EFNFI; 2) Gunhild Berg, Lead Financial Sector Specialist, EECF2; 3) Olushola Ibironke Joanne Martins, IT Officer, Business Management II, ITSIO; 4) Mert Ozdag, IT Officer, Engineering I, ITSIO; 5) Francesca Hopwood Road, Centre Head, BIS Innovation Hub – London; 6) Marco Barzanti, MCMFR, IMF; and 7) Emmanuel Ayanfe Crown, ITDAI DA, IMF.

Please note that some sections of the paper have been drafted using World Bank Group approved generative AI technology. The Bank confirms that relevant sections have been reviewed for accuracy and validity.



Executive Summary

Regardless of our individual perspectives on Artificial Intelligence (AI), it could transform our personal and professional lives at an unprecedented pace. It will also impact one of the most regulated and supervised industries in the world – the financial sector. Risk-based supervision (RBS) has been the gold standard for financial sector supervision over the past two decades, promising to aid supervisors in fulfilling their extensive and constantly growing responsibilities with limited resources. Regrettably, progress toward effective implementation of the RBS framework has not been uniform globally. As indicated by the Financial Sector Assessment Program (FSAP) assessments in recent years, most developed countries have managed to establish a robust RBS framework. However, middle- and low-income countries continue to struggle with the effective implementation of RBS.

The remarkable advancement of AI in recent years, both in terms of performance and accessibility, promises to revolutionize numerous industries, including the financial sector. Although attempts to use AI for supervisory purposes began some years ago, they were mainly experimental and limited to a few high-capacity supervisory authorities. However, recently, thanks to a rapid increase in quality and affordability, AI has begun to be considered a mainstream product that can be used by supervisors for most of their business processes. AI could indeed be a game-changer in the realm of financial sector regulation and supervision, introducing functionalities and capabilities that can reengineer supervisory processes, making them proactive and preventive even in countries with limited human resources.

Initial attempts by global supervisors to utilize AI for various supervisory purposes have yielded promising results. Areas that have traditionally suffered from a lack of sufficient resources, such as Anti-Money Laundering and Combating the Financing of Terrorism (AML/CFT), consumer protection, credit risk, and stress testing, among others, can now benefit from a significant boost in automation and big data processing. An increase in productivity is anticipated in all supervisory processes that involve the processing of unstructured data or operations with large quantities of data points.

The benefits of AI for financial sector supervision extend beyond the automation of some manual activities. In the near future, we are likely to see AI end-to-end automated supervisory processes, particularly those that do not require professional judgment and flexibility in decision-making. Moreover, AI can enable supervisors to undertake processes that were previously considered too time-consuming or/and impossible to perform at the previous stage of technological development.

Of course, all these benefits do not come without risks and costs for supervisory authorities. The reengineering of supervisory processes based on AI capabilities will require effort but, more important, it will require the readiness to deal with new risks generated by the use of AI functionalities. Preparing for the transition to AI-supported RBS will require supervisors to reconsider their processes, including the allocation of human resources to activities and how much supervisors can in fact delegate to AI.

There is no expectation that AI will replace humans in the area of financial sector regulation and supervision, even in the long term. Humans will continue to be decision-makers on all critical aspects related to the supervision of the financial sector. However, we foresee an increasing symbiosis between humans and AI in relation to the regulation and supervision of financial institutions. By increasing efficiency, this symbiosis could lead to a much safer financial industry, better consumer protections, and fewer abuses by criminals.

Recognizing that the financial sector and its regulatory bodies are inevitably part of this ongoing evolution, we set out in this paper to examine the tangible impact of AI on the financial sector, with a particular focus on the supervisory perspective and the transition to an effective RBS regime. We also attempted to forecast the medium- and long-term implications of AI on the roles and responsibilities of the financial sector supervisor. This exploration seeks to enhance our understanding of how AI is influencing the financial landscape and its implications for future regulatory practices.

Main challenges faced by the financial sector supervisors

Over the last two decades, a primary goal of financial sector supervisory authorities around the world has been to effectively implement the risk-based supervision framework. An RBS approach is widely regarded as the most effective means to ensure the stability of the financial sector. This goal has transformed over time to become the main challenge for supervisory authorities. Yet despite significant resources allocated for the implementation of the RBS approach, in addition to often considerable technical assistance provided, many supervisors continue to struggle to implement an effective RBS regime.[1] The reasons for the difficulties in implementing an effective RBS framework vary from country to country and encompass both internal and external factors.

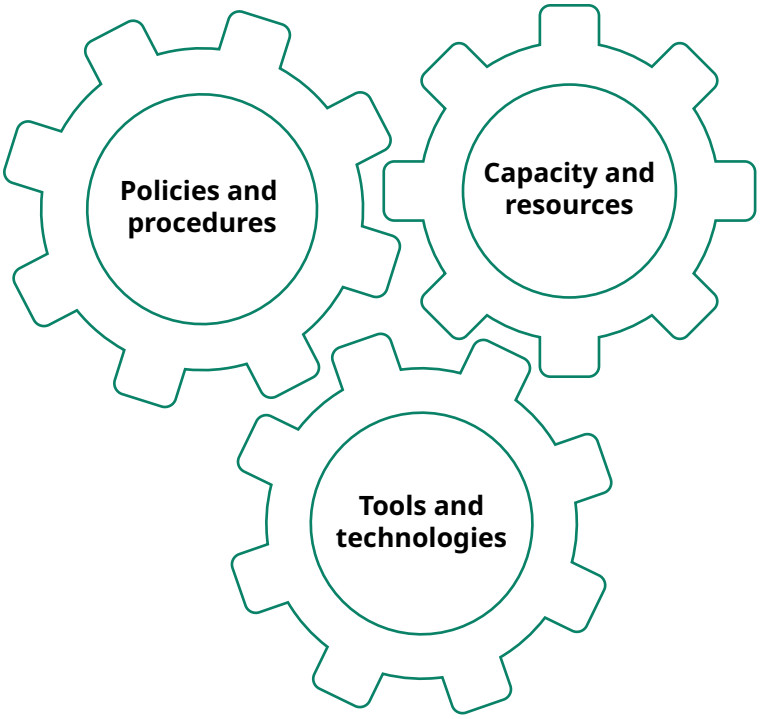
The RBS framework can be disaggregated by different criteria, one of which disaggregates it into three main categories: A. Policies & Procedures, B. Capacity & Resources, and C. Tools & Technologies (Figure 1). We examined these three categories by looking at the assessments of compliance with Basel Core Principles for Effective Banking Supervision (BCP) conducted during FSAPs. As part of this research, publicly available FSAP documents have been reviewed, including the Financial Sector Assessment (WB),[2] the Financial System Stability Assessment (IMF),[3] the Detailed Assessment of Observance as well as International Financial Institutions' publications. It is important to note that the research was mainly focused on prudential supervision and compliance with BCP Principles.

1 Good Supervision: Lessons from the Field, IMF (https://www.imf.org/en/Publications/WP/Issues/2023/09/06/Good-Supervision-Lessons-from-the-Field-538611).

2 World Bank prepares a Financial Sector Assessment (FSA) for its Executive Board.

3 Financial Sector Stability Assessment (FSSA) is prepared for discussion at the IMF Executive Board.

Figure 1 Core elements of Risk-Based Supervision



The findings indicate that jurisdictions have made steady progress on implementation of major Basel regulatory reforms[4]. Enhancements in this area were achieved in part due to extensive technical support provided by the World Bank and the International Monetary Fund (IMF) over recent decades to enable transition to effective RBS. Additionally, these technical assistance projects have significantly reinforced the supervisory capacity, even if the retention of experienced staff remains a challenge in many countries. Our research also revealed a noteworthy trend among most countries that have been unsuccessful in implementing an effective RBS regime: These countries have often either completely overlooked or significantly underestimated the importance of supervisory tools and technologies. This finding underscores the critical role that tools and technologies play in the successful establishment of an RBS framework.

While IT tools that can substantially enhance the productivity of supervisors and aid in the transition to an effective RBS framework have become increasingly accessible in recent years, supervisors are not rushing to overhaul outdated

4 Good Supervision: Lessons from the Field, IMF (https://www.imf.org/en/Publications/WP/Issues/2023/09/06/Good-Supervision-Lessons-from-the-Field-538611).

processes that rely heavily on low SupTech generation tools. Most of the recent implementations remain in first- to second- generations of suptech, while fourth SupTech generation tools (e.g. Generative AI) represents only 7.6% of implementation[5]. Although the slower adoption of innovative technologies in the supervision area can be attributed to a general conservatism among supervisors, this disconnect between the potential of technology to enhance supervisory processes and its actual application is likely to widen with the rapid progression of AI technology. Particularly, the much-needed effective implementation of the RBS regime may continue to be hindered if supervisors fail to explore the use of advance technologies to increase their productivity and capacity to identify, measure, and mitigate the risks. This is especially relevant given the widespread adoption of AI in the financial industry, particularly among FinTech companies.

As part of our research, we tried to identify the main challenges faced by financial sector supervisors at all core stages of the supervision cycle. In this context we looked at:

a. **Market entry –** which includes licensing of new entities, opening of new subsidiaries / branches, approval of shareholders, and approval of management and key personnel.

b. **Ongoing supervision –** which includes macro-prudential supervision and micro-prudential supervision (on-site and off-site) distributed on key risk dimensions.

c. **Resolution and market exit –** which includes restructuring and resolution of the financial entities and liquidation.

To identify the main challenges in supervisory activities, the team reviewed the assessments of the compliance with the BCP conducted in recent years by the World Bank and the IMF as part of the FSAPs. Below are presented the main challenges identified by the team, which are specific for most of the supervisory authorities, including developed countries.

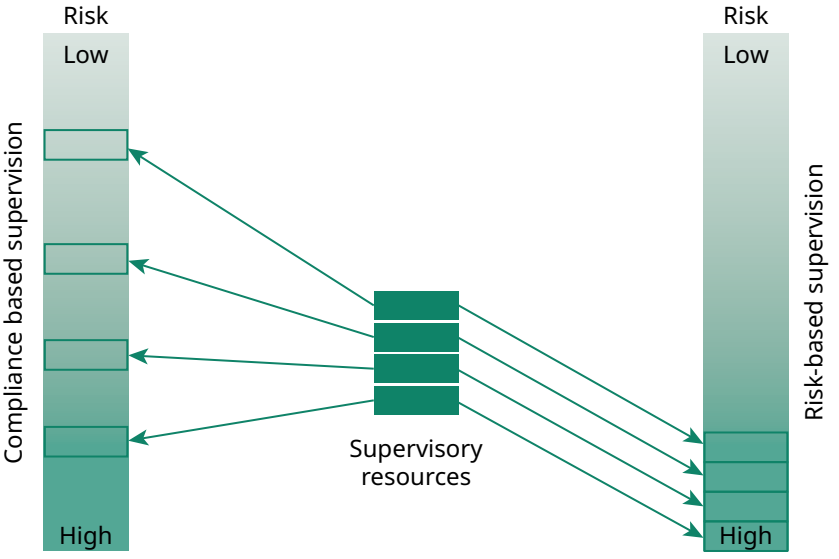
Limited human resources

Each supervisory stage mentioned above encompasses a vast array of risk parameters that can be evaluated and monitored. However, it's important

5 State of Suptech report 2023, Cambridge SupTech Lab (https://lab.ccaf.io/state-of-suptech-report-2023/)

to note that assessing every risk-generating activity for a specific financial institution is an impractical task, even for developed countries, due to the limitations of resources, particularly human resources. This is where the RBS approach comes into play. RBS is a supervisory strategy designed to assist supervisors in fulfilling their responsibilities within the constraints of available resources. The fundamental premise of RBS is that these finite supervisory resources are primarily allocated to the assessment and monitoring of high-risk operations and activities within financial institutions (Figure 2). The foundation of an effective RBS regime lies in the supervisory authority's ability to accurately identify high-risk activities and strategically allocate limited resources for their assessment. However, achieving this balance is not always straightforward and remains a persistent challenge in the implementation of an effective RBS regime. RBS will remain highly relevant and the gold standard for many years, even with the broader adoption of AI. Human decision-makers will continue to play a crucial role in supervisory processes, as it is unlikely that supervisors will ever be able to collect all granular data necessary to fully automate these processes.

Figure 2 Compliance-based Supervision vs Risk-Based Supervision



Even when a supervisory authority successfully identifies risks and allocates resources efficiently, it often finds its resources insufficient to cover the primary risks. This has become particularly evident in recent years as digitalization has led financial institutions to rapidly increase the complexity of their business models and the range of services they offer. Increasing the staff of the supervisory authority is not a viable option in most countries. Such increases

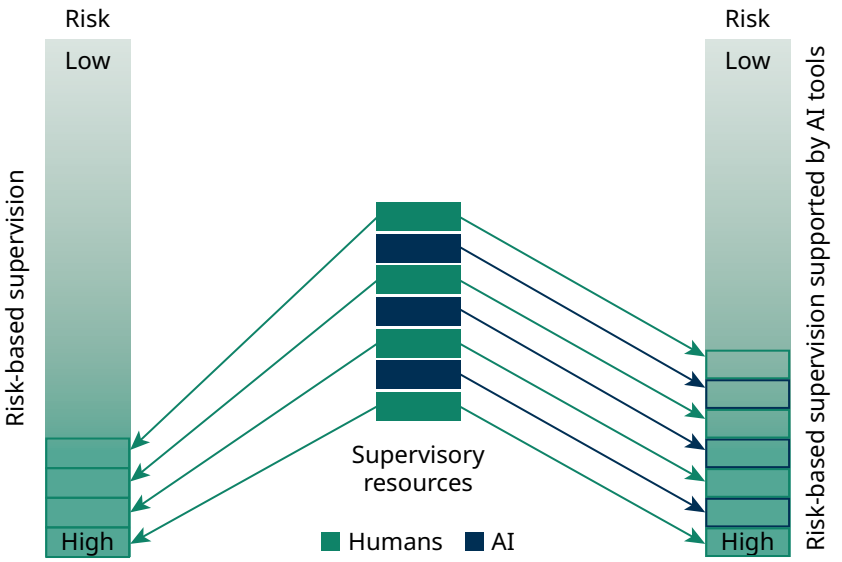
are often unpopular and challenging to implement. Given these constraints, the only feasible solution is to find ways to accomplish more with the existing limited human resources – in other words, to enhance productivity.

A closer look at supervisory processes reveals a significant potential for automation and optimization in many supervisory authorities. While the extent of this potential varies across supervisors in countries at different levels of development and sophistication, it is evident that current supervisory processes often involve manual and repetitive activities. These activities, which do not always require sophisticated qualifications or intense professional judgment, are nonetheless essential for supervisors to reach the stage where professional judgment can be applied. However, these basic and non-complex activities are essential for supervisors to reach the stage where professional judgment can be applied.

Tasks such as collecting, validating, cross-checking, enriching, and requesting data are routine activities that supervisors must undertake before they can apply their expertise to interpret trends, propose remedial actions, or decide on regulatory reactions. In recent years, AI solutions have matured enough to significantly replace most of the non-complex supervisory activities performed by humans, in addition to assisting with more complex processes (e.g., on-site inspection, stress-testing, resolution, and other). Additionally, the accelerated adoption of financial technology (FinTech) and regulatory technology (RegTech) solutions by financial institutions has significantly enhanced data quality and increased their capacity to report more granular data with higher frequency. However, the most important aspect is AI's potential to help supervisors identify high-risk activities and processes within financial institutions, thereby enabling a much more efficient allocation of human resources.

With appropriate implementation of AI solutions by supervisors, there is considerable potential to significantly increase the productivity of supervisors, free up time dedicated to routine activities, and allow them to focus on processes that require intense human involvement. This approach could lead to more effective and efficient supervisory processes, ultimately benefiting the financial sector as a whole (Figure 3).

Figure 3 Risk-Based Supervision vs Risk-Based Supervision supported by AI tools



Data quality and granularity

While professional judgment and, at times, intuition are essential attributes for supervisors, the effective implementation of RBS is fundamentally reliant on data – and a substantial amount of it.

The most frequent challenges related to data in supervisory authorities are:

a. *Data quality* **–** This is a prevalent issue in the data collection process within supervisory authorities. It is crucial to clarify that poor data quality does not necessarily imply intentional misreporting or manipulation of prudential requirements by supervised entities. The reasons for low data quality are manifold and are often rooted in the historical construction of the reporting framework. Some of the more common causes include: i) lack of a data governance policy and team – the absence of a structured approach to managing data quality and a dedicated team to oversee this process can lead to inconsistencies and errors; ii) absence of reporting standards – without clear and uniform reporting standards, data can become fragmented and difficult to compare or analyze; iii) miscommunication between business departments and IT team – effective data management requires seamless collaboration between business and IT teams; and iv) lack of a dedicated reporting tool – many supervisors still use email to collect

data from supervised entities, a method that has numerous deficiencies, particularly concerning data quality.

b. *Data granularity* **–** The lack of sufficient data granularity, especially concerning the most significant risks, directly impacts the ability of supervisors to promptly identify negative trends and proactively address them. Many supervisors have built their reporting around a template-centric approach. While this approach has its merits, it does not provide supervisors with the granularity needed to delve into the data and identify the root cause of specific dynamics without soliciting additional data from the supervised entities. The solution to this challenge lies in transitioning to a data-centric reporting approach which focus on collection and management of raw data. This transition involves a complex overhaul and reengineering of the reporting framework, impacting all key stakeholders. Some supervisory authorities have been hesitant to make this transition due to the lack of tools capable of processing granular data, particularly transactional data. However, it's worth noting that without transitioning to a data-centric reporting approach, leveraging the functionalities of AI is nearly impossible, or at best the benefits will be very marginal. AI tools can both facilitate an effective transition to data-centric reporting and benefit from the acquisition of more granular data for training AI algorithms.

c. *Data redundancy and absence of data governance and the single source of truth* **–** One of the most common consequences of a lack of proper data governance is the gradual escalation of data redundancy and degradation of data quality. Permitting various business units within the supervisory authority to establish their own reporting requirements without coordination or validation from a centralized data governance team may offer flexibility at the initial stage. However, over time this decentralization can become a burden for both supervisors and reporting entities. Data redundancy can escalate rapidly, reaching extreme levels in a relatively short period. And even periodic reviews of the decentralized reporting approach fail to fully address this issue. In addition to increasing data redundancy, which results in additional costs for reporting entities, this approach complicates the creation of a centralized, single source of truth database for key indicators.

d. *Ad hoc reporting burden* **–** the template-based reporting framework often necessitates significant ad hoc reporting. Due to the inability of template-based systems to facilitate in-depth data analysis, the primary method for uncovering the underlying causes of concerning trends

is to solicit extra ad hoc reports from regulated entities. This process substantially escalates compliance expenses. At the same time, this results in a multitude of decentralized data collected via email (in most cases) and stored locally in inconsistent structures and formats, making centralization of this data very difficult.

e. *Endless reporting channels* – Supervisory authorities rarely rely on a single channel for data collection. While multiple channels may seem beneficial from a business continuity standpoint, the reality is that the proliferation of reporting channels is more often a result of convenience and the fragmented evolution of the reporting process. The transition from paper and email to application programming interfaces (APIs) and data pulling has led to a multitude of reporting channels, each built on different technologies. This diversity makes it challenging, if not impossible, to implement uniform reporting standards and validation rules. Apart from numerous other drawbacks that affect the efficiency of data reporting, the abundance of reporting channels introduces significant cyber and data privacy risks. In many developing countries it is the norm to receive reports, including those containing confidential data, via email (including into personal email accounts), rather than being an exception.

f. *Suboptimal collection of unstructured data* – the collection of unstructured data remains significantly inadequate, even in developed countries. An effective RBS regime cannot rely solely on structured data, yet the gathering of unstructured data is often overlooked. This has primarily been due to the lack of cost-effective tools that could assist supervisors in processing and extracting valuable information for identifying, measuring, and monitoring risks from unstructured data. Examples of unstructured data relevant to supervisors include emails, meeting notes, social media posts, customer complaints, and transaction narratives. However, the rapid advancement of AI tools capable of processing unstructured information and integrating it with structured data has been a game changer, leading to a more comprehensive understanding of complex processes.

Deficient analytics

To improve the decision-making process, data collected by supervisory authorities must be meticulously processed and analyzed. While it is understood

that regulators do not simply collect data without analysis, it is crucial to emphasize the importance of considering trends, dynamics, and deviations, as well as conducting complex simulation scenarios. These steps are essential for enhancing financial stability. An effective RBS regime requires early identification of systemic risks, implementation of preventive measures, and a consistently forward-looking supervisory approach. Achieving these objectives necessitates robust analytical capabilities to ensure comprehensive and proactive supervision.

Regrettably, backward-looking analytics still remain the primary analytical activities in many supervisory authorities, particularly in developing countries. Analytics that primarily focus on understanding past dynamics, rather than predicting future dynamics, even approximately, are ill-equipped to support a preventive decision-making process.

It is important to recognize that the current situation is not due to a lack of understanding by supervisors about the importance of forward-looking analytics, but rather the absence of suitable analytical tools and resources to implement it. Even within the limits of structured data, where there are ample high-quality analytical tools available, most analytics are performed using basic flat table processing tools. The situation is even more dire for unstructured data, where analytical processes are almost entirely manual.

Big data analytics, predictive analytics, network analytics, and others are essential analytical functionalities for a supervisory regime that can adapt flexibly to the constantly changing risk environment in the financial sector, the increasing complexity of financial institutions' business models, and amendments to international standards and rules.

Outdated processes

According to the Core Principles for Effective Banking Supervision, key supervisory processes should be explicitly outlined in the internal policies, procedures, and manuals of the supervisory authority. This ensures consistency, proportionality, and a non-discriminatory approach to supervisory actions. While most supervisory processes are detailed in on-site and off-site supervision manuals, these processes are rarely calibrated to the potential of advanced IT solutions that could support them. The potential of AI as a supporter of on-site and off-site supervisory processes is often overlooked.

As a result, these processes are designed with the assumption of using outdated IT solutions, which offer limited support to the productivity of supervisors. This creates a vicious cycle: internal manuals cannot reference tools that do not yet exist within a supervisory authority, and at the same time, implementing advanced AI-based solutions to follow process designed for a different technological era makes little to no sense.

Consequently, we have de facto processes that may not align with existing written procedures. Information collected as part of our technical advisory (TA) activities in the SupTech area indicate that individual teams frequently attempt to increase their productivity by leveraging available tools, often resorting to demo or free versions, which come with their own set of challenges and risks. A common example is the use of free versions of ChatGPT or similar products. While these AI tools can facilitate many supervisory processes, especially those related to unstructured data, they raise significant concerns regarding data confidentiality and output quality. Additionally, free versions of AI tools have other limitations, such as a limited number of queries per day, that may significantly restrict the practical application of these tools for supervisory processes, as the work of supervisors is not linear. Moreover, most free AI tools do not have memory and are not able to connect new queries with previous queries or provided information.

Given the rapid advancements in AI, it is inevitable that supervisors will need to undertake a substantial review and overhaul of their supervisory procedures and manuals. This will likely result in a redistribution of resources, particularly in terms of on-site and off-site supervision, but will also require a specification of the role of AI in the supervisory activities. We can expect a significant increase in the off-site supervision role as data-centric reporting, unstructured data, open data, and AI tools provide new opportunities for the off-site team to identify risks without leaving their offices or virtual space. However, on-site supervision will continue to play a crucial role in establishing and maintaining an effective RBS regime, although the responsibilities may undergo significant revisions for targeted activities.

Limited expertise

As previously noted, the majority of analytics in supervisory authorities are still conducted using basic tools for processing two-dimensional tables with limited data points. This situation is not solely due to the high cost of more

sophisticated analytical tools, despite increased access to such tools in recent years, but also is the result of a lack of expertise within the supervisory authority to operate these advanced analytical tools. There is a valid concern that investing in advanced analytical tools may have a limited impact on productivity due to a lack of necessary expertise. It is generally very difficult, even in high-income countries, to hire and retain staff with strong expertise in financial sector supervision, as well as skills in data science, data engineering, and data analytics. This concern is compounded by the noncompetitive remuneration policy in many supervisory authorities. Retaining staff with expertise in financial sector risks and advanced data analytics is a constant challenge for supervisors, as the private sector offers more flexible remuneration policies and can employ more aggressive recruitment strategies. These factors are forcing decision-makers to pause or postpone digital transformation projects in supervisory authorities, which inevitably impacts the effectiveness of the RBS regime.

However, as with other challenges mentioned above, AI tools could be a game changer. Their main innovation comes from so-called low-code or no-code tools. While these tools do not represent a universal solution for every problem, their progress in recent years has been substantial. Tasks that required solid knowledge of coding or processing and transforming big data three to five years ago can now be performed with drag-and-drop actions or even by writing the task (prompt) in a format understandable to humans. Also, major IT companies (e.g., Meta, Alphabet, and others) have invested in and open-sourced critical AI frameworks. These libraries, much like low-code tools, support the development of AI solutions with fewer resources.

This rapid transformation of advanced data analytics tools, supported by AI functionalities, can significantly expand the pool of experts who can effectively use them. Even data analytics processes with above-average complexity can be performed without a single line of code from the direct user. However, for this to happen, proper data architecture and governance in the financial authority remains a precondition for implementing advanced analytical tools.

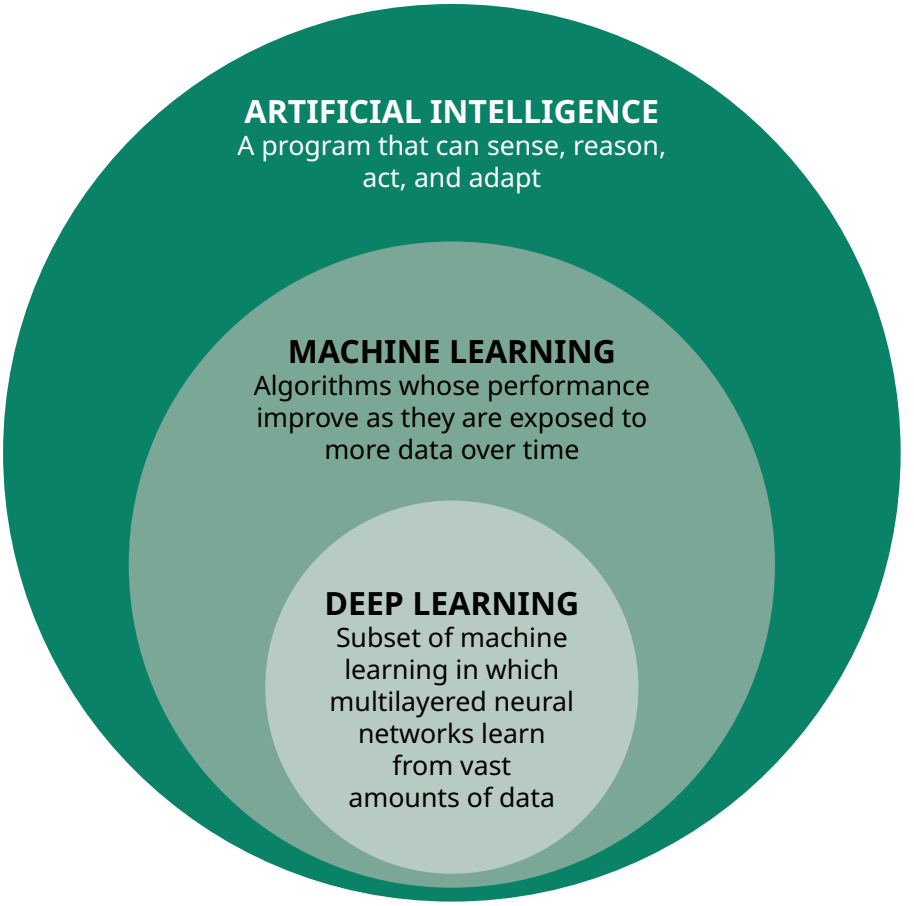
Empowering financial supervisors with AI capabilities

Financial supervision has always been data driven. The risk-based supervision (RBS) approach depends on data pipelines that ensure a continuous flow of sufficient, relevant, accurate, and timely information to supervisory authorities. These authorities must possess the necessary skills, practices, and technologies to derive value from the data and effectively support the RBS model. Although supervisory authorities have been early adopters of traditional data technologies, these tools are now struggling to meet the emerging needs of supervisors. This is largely due to the overwhelming volume of data that must be collected and processed for effective financial supervision, as discussed in the previous chapter. Under these circumstances, artificial intelligence (AI) stands out as the most capable technology for managing the complexity and diversity of data that currently overwhelms supervisory authorities.

AI holds transformative potential, offering robust solutions that can significantly enhance both the efficiency and efficacy of financial supervision. According to EU AI Act,[6] AI is a software that "infers from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments." AI is broadly understood to include various types and subtypes, such as machine learning (ML), deep learning, natural language processing (NLP), computer vision, and generative AI. Each offers distinct capabilities, making AI models well-suited for specific tasks, many of which are also relevant to the work of supervisory authorities.

6 https://artificialintelligenceact.com/article-3-definitions/

Figure 4 ML is sub part of AI



Source : Medium (2020)

Machine learning models in particular have been long utilized by supervisors to enhance data analytics by learning from historical data to support informed decision-making and predict future trends. Generative AI on the other hand, known for its ability to process and generate text in natural language, only recently started to attract significant interest. A McKinsey study[7] analyzing the impact of AI across 16 business functions found that approximately 75 percent of the value from generative AI use cases is concentrated in four key areas: customer operations, marketing and sales, software engineering, and research and development. Supervisory authorities, engaged in similar-in-nature activities, stand to benefit greatly from the adoption of AI technologies.

7 McKinsey. 2023. *The economic potential of generative AI: The next productivity frontier*, https://www.mckinsey.com/.

Considering the numerous challenges that supervisory authorities encounter, there are extensive opportunities to leverage the vast potential of AI. This chapter explores the foundational capabilities of AI that can be harnessed to meet the evolving needs of financial oversight, detailing how AI can address the challenges faced by supervisory authorities.

Addressing challenges related to human resources

AI has the potential to significantly enhance Risk-based supervision (RBS) by automating time-consuming tasks, which allows supervisory authorities to focus resources on high-risk activities. AI technologies like machine learning (ML), optical character recognition (OCR), and natural language processing (NLP) can transform how tasks are performed and improve supervisory efficiency.

ML models can detect anomalies, such as suspicious transactions or unusual activities, which are critical for anti-money laundering and countering the financing of terrorism (AML/CFT) supervision. They also assist in monitoring ownership changes, aiding shareholder transparency, and grouping entities based on risk profiles, which allows supervisors to allocate resources more efficiently toward higher-risk entities.

OCR technologies are beneficial for digitalizing paper documents, enabling the extraction and analysis of structured data. This is especially useful in processes like market entry, where paper-based documents are prevalent. OCR can also streamline the review of meeting minutes from supervised institutions' management bodies, enhancing oversight and saving time.

NLP models enhance supervision by analyzing and comparing submitted documents such as internal regulations and policies of supervised entities. When trained on the regulatory framework, these models can assist in assessing compliance or noncompliance, significantly speeding up the document review process.

Additionally, ML models can assess market sentiment or the reputation of specific entities, products, or topics. This information can be integrated into early warning systems, allowing supervisory authorities to identify potential market conduct issues before they escalate, thereby improving overall oversight and proactive risk management.

Addressing challenges related to data quality and granularity

Though AI cannot resolve all data-related issues, such as improving data granularity or eliminating the burden of ad hoc reporting, it can significantly assist in addressing key challenges, particularly around data quality and the extraction of insights from unstructured data.

- **Improving Data Quality:** Machine learning models can help validate data by identifying outliers and ensuring data integrity. A rule-based layer should be implemented as a first line of defense to catch common errors. Additionally, ML techniques such as named entity recognition (NER) and named entity disambiguation (NED) can extract structured data from unstructured sources (e.g., identify unique entities such as individuals and companies from web data).
- **Addressing Data Redundancy:** While AI models like NER and NED cannot directly eliminate data redundancy across various internal sources, they can correlate information from multiple systems, helping to create enriched entity relationships. They can also flag inconsistencies in data related to the same entity or event across different datasets.

The availability of granular supervisory data, enhanced by AI-driven validation for improved quality, not only boosts short-term supervisory effectiveness but also supports the long-term continuous refinement and improvement of AI models.

Addressing challenges related to deficient data analytics

Predictive machine learning models such as linear regression have long been used to analyze data, but their effectiveness depends heavily on the quality, relevance, and completeness of the input data. A key challenge is identifying all relevant factors necessary for accurate predictions. ML models excel at uncovering hidden patterns and relationships within the data that might otherwise go unnoticed.

As AI continues to evolve, more advanced ML models can help supervisors to apply nonlinear relationships to predict and anticipate systemic risks, which are often better suited to complex, real-world problems. Techniques such as random forests, and gradient boosting machines offer significant advantages

by building on decision trees to capture intricate relationships between variables. These models can account for curved patterns in the data, improving the accuracy and robustness of predictions. As a result, they offer more reliable insights compared to traditional, linear models, particularly in scenarios where data relationships are not straightforward.

Addressing challenges related to outdated processes

While it may be challenging for AI models to manage the documentation of supervisory processes, applying proper IT governance is essential for maintaining transparency and control over the IT solutions and software tools that support these processes. This transparency ensures that process documentation remains updated and accurate, preventing any AI tool usage from going undocumented. By systematically documenting these tools and their applications, organizations can avoid gaps in operating procedures and internal regulations, ultimately improving compliance and operational efficiency.

Addressing challenges related to limited expertise

Supervisory authorities often face challenges due to a lack of expertise in operating advanced analytical tools. AI technologies, particularly those powered by NLP and ML, can help address this limitation by simplifying data navigation and extraction from large repositories, making complex data more accessible and actionable.

For instance, translating natural language into structured query language (SQL) queries – an otherwise complicated task – can be managed by advanced AI models. Specifically, transformer-based models and sequence-to-sequence (Seq2Seq) models can be trained to map natural language inputs into structured SQL queries. These models can automatically convert user queries into the required SQL format, thus reducing the need for specialized technical expertise. However, the effectiveness of such AI models depends heavily on the quality of the underlying data architecture documentation. A well-structured enterprise data catalog, which defines and describes data types using natural language, is crucial for ensuring that AI models operate efficiently. With the right AI tools and data governance in place, financial authorities can bridge the expertise gap and better manage their data analytics needs, even with limited internal data science skills.

Foundational capabilities of AI that can be harnessed by supervisory authorities

The value proposition of AI for supervisory authorities lies in its unmatched ability to process and analyze massive volumes of data, surpassing the limitations of traditional technologies. Some of these foundational capabilities of AI are outlined below.

Enhanced Data Processing and Analysis

AI excels in handling large volumes of data quickly and efficiently. Through advanced algorithms and machine learning, AI systems can process and analyze data sets far beyond the capability of human supervisors or traditional data processing tools. This includes not only structured data such as transaction histories and compliance reports, but also unstructured data like emails, social media posts, and news articles. AI algorithms can identify patterns and anomalies within this data, providing insights that were previously inaccessible due to the volume and complexity of the information involved.

For financial authorities, this means an unprecedented level of insight into market trends, consumer behavior, and potential regulatory violations. AI can automate the initial stages of data analysis and flag potential issues for human review. This can allow supervisors to allocate their resources more effectively and focus on high-priority risks rather than routine data processing.

Predictive Analytics and Risk Forecasting

One of the most powerful applications of AI in financial supervision is predictive analytics. AI systems can forecast potential market disruptions and financial risks by analyzing historical data and identifying trends that may indicate future problems. This proactive approach to risk management enables supervisory authorities to take preventive measures before issues escalate into serious threats.

Machine learning models can be trained to predict outcomes based on specific indicators such as credit defaults or unusual trading patterns. These models can be continually improved as they process more data, increasing their accuracy over time. For supervisors, this means being better equipped to manage the complexities of the financial markets, ensuring stability, and protecting consumers from potential harm.

Automation of Routine Tasks

AI can automate many of the routine and repetitive tasks that currently consume a significant portion of supervisors' time. This includes document analysis, report generation, and monitoring compliance with regulatory frameworks. AI can execute these functions quickly and with minimal errors, significantly speeding up processes while enhancing their accuracy. This provides financial authorities with the reliable data essential for making informed decisions.

For instance, natural language processing (NLP) technologies are adept at interpreting communications with financial entities, consumers, and other market participants. Domain-specific chatbots or AI assistants can offer expert support to both external and internal stakeholders, assist in collecting necessary data and documents, validate data accuracy, and even verify the authenticity of documents during all interactions with the supervisory authority.

Enhanced Access to Knowledge

AI significantly enhances decision-making processes by providing detailed analyses and model-driven predictions that inform supervisors' choices. By adopting AI tools, financial authorities gain access to a deeper and broader set of data insights. Large language models (LLMs) such as GPT-4, Llama 3, or Gemini enable supervisors to quickly summarize extensive documents or extract relevant topics from both internal knowledge bases and external sources. These capabilities provide supervisors with unprecedented access to comprehensive knowledge, aiding them in making more informed decisions based on detailed risk assessments and trend analysis that consider a wide array of variables and potential outcomes.

Continuous Learning and Adaptation

AI systems are capable of continuous learning and adaptation. When designed for this purpose, they can update their models and strategies as they encounter new data and outcomes, reflecting new information and changing conditions. The AI models can be fine-tuned or retrained to update the models and strategies to reflect new information, changing conditions, and user feedback. This capability allows financial supervision to evolve in response to new challenges and opportunities.

The foundational capabilities of AI offer profound opportunities for supervisory authorities to enhance their supervisory functions. By harnessing AI's potential to analyze large data sets, predict future trends, automate routine tasks, and support complex decision-making processes, supervisors can significantly advance their capabilities and manage the challenges faced. As AI technology continues to evolve, it will play an increasingly critical role in shaping the future of financial supervision, driving innovations that ensure market integrity and protect consumer interests.

Use case of AI in supporting activities of the supervisory authorities

AI is revolutionizing a multitude of sectors, and financial supervisory oversight is no exception. Globally, financial supervisors are harnessing these advanced technologies to elevate their digital transformation as the integration of AI into RBS represents a significant leap forward for overseeing the banking sector.

Financial Institutions are expected to provide large amounts of data points through regulatory returns at multiple frequencies, and AI's advanced data collection and processing capabilities allow for the analysis of vast quantities of both quantitative and qualitative data, which can uncover trends and anomalies that might elude traditional methods.

Predictive risk modeling is another area where AI shines, utilizing historical data to forecast potential future risks. This predictive power enables authorities to adopt a proactive stance, anticipating and mitigating risks before they materialize. Automated risk scoring and dynamic assessments has the potential to further streamline the supervisory process, allowing for real-time adjustments as new data comes in.

Moreover, AI can generate sophisticated reports and visualizations, making complex risk profiles easily understandable. Decision support systems augmented by AI assist authorities in making well-informed choices, while AI-driven action plans offer customized strategies to address specific risks identified in financial institutions.

Instances of AI adoption by financial authorities include regions such as the North America, Asia, and Europe, among others. Similarly, central banks in Asia are leveraging big data to evaluate micro-supervisory risks, drawing on information from financial statements and news articles to improve credit scoring

models and early warning systems for loan defaults, as well as to detect unusual transactions. Meanwhile, the Dutch central bank utilizes daily transactional data to establish network and operational benchmarks, as well as to monitor liquidity movements within TARGET2 and other financial market infrastructures. This extensive data collection aids in identifying cyclical trends, establishing a normative baseline for monitoring and risk assessment. Market deviations from this baseline are scrutinized for potential signs of heightened risk.[8]

Globally, financial supervisors are increasingly integrating AI and ML technologies into their operations to bolster regulatory supervision, enhance risk management, and improve decision-making. These technologies offer critical insights and predictive analytics, helping financial authorities to identify potential threats and ensure the stability and integrity of financial markets.[9] The following examples dive into the specifics of AI adoption journeys and the technological capabilities of these solutions.

Australian Securities and Investments Commission

The Australian Securities and Investments Commission **(**ASIC) also introduced The Market Analysis and Intelligence (MAI) system for market surveillance. The MAI system offers generation of real-time alerts that identify market anomalies for potential investigation or detection at the time of execution. These alerts are integrated into the daily workflows of the staff and can lead to detailed investigations and analyses. Additionally, the system provides historical analytical capabilities that allow for complete market reporting and the assessment of broad and complex thematic risks. These historical analyses are performed within an integrated environment that can be connected to other cases or to issues prioritized by surveillance staff.

Specifically, the system is equipped with real-time data streams from all significant Australian capital markets, including the Australian Securities Exchange (ASX) and Chi-X, covering equity and equity derivative transactions. These live data streams are enhanced by nightly data feeds, which include information on listed futures and derivatives products. To broaden the scope of

8 Asian Development Bank. 2022. *Building Regulatory and Supervisory Technology Ecosystems For Asia's Financial Stability and Sustainable Development*, https://www.adb.org/sites/default/files/publication/820686/regulatory-technology-ecosystems-asia-financial-stability.pdf.

9 University of Cambridge. "Cambridge SupTech Lab." The Suptech Marketplace. Accessed April 29, 2024. https://ccaf.io/suptechlab/solutions_tracker/solutions.

visibility into Australia's capital market activities, the MAI system incorporates daily information from over-the-counter derivatives transactions, sourced from designated trade repositories. This data is integrated into a post-trade analytics framework, underpinned by a combination of programs and software.[10,11]

This framework is in the process of being expanded into a big data platform such as Elasticsearch, SPARK, and Kibana, along with other supplementary technologies. This expansion aimed to enable the visualization of time-series data through various modalities such as reports, dashboards, and networks, and to provide inherent support for machine learning, alert generation, and comprehensive searches across large data sets. The post-trade analytics environment offers ASIC a continually evolving and increasingly detailed view of Australian financial markets, further enriched by the addition of more data for deeper insights. ASIC has access to a suite of algorithms, analytics, and reports based on this dataset, which highlight entities or trades of interest.

ASIC also leveraged a natural language processing (NLP) tool specifically for advertising, which can streamline the process of identifying non-compliant content within advertising materials. Given the extensive amount of such materials, ASIC's responsibility to ensure adherence to both the Australian Consumer Law and the National Credit Code in advertising can only be managed by either examining a random selection of content or by responding to complaints from consumers.[12]

European Central Bank

The European Central Bank (ECB) has utilized artificial intelligence to oversee banking operations, including the launch of a forward-looking digital strategy to improve its analytical functions. A suite of SupTech tools has been adopted to monitor the intricate European Union banking industry and handle a growing volume of data and responsibilities. Attention has also been given to the end-users of this technology; over the past three years, 14 applications and platforms have been created, benefiting over 3,500 users within the ECB as well as national supervisory authorities.

10 Broeders, Dirk, and Jermy Prenio. 2018. "Innovative Technology in Financial Supervision (Suptech)." BIS, July 2018. https://www.bis.org/fsi/publ/insights9.pdf.

11 OECD. 2021. "OECD Business and Finance Outlook 2021: AI in Business and Finance. Paris: OECD Publishing." https://doi.org/10.1787/ba682899-en.

12 inCoNet. 2020. "SupTech Tools for Market Conduct Supervisors - Finconet." FinCoNet, November 2020. https://www.finconet.org/FinCoNet-Report-SupTech-Tools_Final.pdf.

The ECB has invested in sophisticated infrastructure to enhance the scaling of AI applications through the Virtual Lab. This cloud-enabled collaborative platform boasts machine learning capabilities and offers a conducive environment for the sharing and refinement of code. The Virtual Lab has not only fostered improved collaboration among European banking supervisors but has also supported the adoption of cutting-edge technologies such as generative AI, which has been popularized by tools like ChatGPT.[13]

Athena, an AI enabled tool leveraging natural language processing and developed by the ECB, is designed to identify gaps in supervision and regulation. Athena helps the supervisor to compare the internal assessment, linking various dimensions and data sources to augment internal and external views.[14] Supervisors can use advanced topic modelling and sentiment analysis capabilities and receive alerts when a predefined set of conditions appears in the market news that ECB receives daily.

Monetary Authority of Singapore

The Monetary Authority of Singapore (MAS) explored the development of an integrated surveillance platform to collate and aggregate data and information from various sources (e.g., news, financials, macroeconomic indicators), enable sense making by applying NLP/sentiment analysis, and facilitate in-depth analysis and risk identification using machine-learning techniques.[15] AI and ML techniques are being used to build predictive models that forecast financial market trends, credit risk, and other relevant indicators, aiding in proactive decision-making.

MAS applied data analytics and artificial intelligence to assess financial risks, monitor compliance with safe distancing measures, and execute credit grading. MAS has gathered data for analysis and representation on a monitoring dashboard and has employed natural language processing to interpret international news and consumer feedback. MAS is also in the process of

13 https://www.bankingsupervision.europa.eu/press/interviews/date/2024/html/ssm.in240226~c6f7fc9251.en.html

14 European Central Bank (ECB). 2021. Presentation slides - ECB banking supervision, May 2021. https://www.bankingsupervision.europa.eu/press/speeches/date/2021/html/ssm.sp210527_annex~bc0fd7b801.en.pdf.

15 Financial Stability Board (FSB). 2020. "The Use of Supervisory and Regulatory Technology by Authorities and Regulated Institutions." 9 October 2020, https://www.fsb.org/wp-content/uploads/P091020.pdf.

developing a unified surveillance platform to consolidate data and implement machine learning for more profound financial risk analysis. And MAS has merged text analysis with quantitative analytics of financial metrics and displayed the results on a single automated dashboard, providing supervisors with a panoramic view of the reports. This relatively straightforward tool has increased the efficiency of auditing financial statements. Following a pilot in one department, MAS is considering extending its use to additional supervisory departments. MAS is in the process of developing a data analytics system designed to sift through the approximately 3,000 suspicious transaction reports (STRs) filed monthly by financial institutions and pertaining to money laundering and terrorist financing risks.[16,17]

MAS has also used automation tools that required NLP to compile international news and keep updated on COVID-19 related events.[18] NLP was applied to evaluate consumer feedback on issues stemming from the pandemic, as well as to monitor vulnerabilities within different customer and product segments. MAS also compiled weekly data from regulated institutions to monitor the uptake of credit relief measures as the pandemic evolved. The data collection was designed to be nimble, allowing for swift iterative improvements and adaptability. The processes for data gathering and transformation were automated, with the results visualized for effective monitoring.

Further Explorations and Testing

There are numerous exploratory initiatives currently underway to use AI to assist financial supervision. Due to recent advancements in large language models, various supervisory bodies are looking to implement AI to harness its capabilities within their operations, as generative AI holds promise in aiding supervisors by streamlining their daily activities.

In 2023 the ECB identified over 40 potential applications for AI from supervisory feedback and has developed multiple proofs of concept that showcase generative AI's capabilities. The applications the supervisor aimed to tackle

16 Broeders, Dirk, and Jermy Prenio. 2018.

17 Asian Development Bank (ADB). 2022. "Building Regulatory and Supervisory Technology Ecosystems: For Asia's Financial Stability and Sustainable Development." Asian Development Bank (ADB), August 2022. https://www.adb.org/sites/default/files/publication/820686/regulatory-technology-ecosystems-asia-financial-stability.pdf.

18 Financial Stability Board (FSB). 2020.

include providing immediate responses to inquiries about supervisory methods, complete with direct references to internal procedures, and the ability to convert questions posed in natural language into code to identify specific data points.[19]

An insightful example of the successful use of AI to aid financial supervision was recently initiated by De Nederlandsche Bank (DNB), the central bank of the Netherlands. In 2023 a dedicated team of data scientists, software engineers, experts in large language models, and regulatory professionals at the central bank embarked on an intensive project to develop a generative AI tool named ChatDNB, utilizing Microsoft's Azure OpenAI Service. The team crafted a proof of concept using five publicly available annual reports as the textual foundation for training. This early version demonstrated the tool's potential and garnered attention from supervisory staff. Capitalizing on this momentum, the team shifted its focus to the creation of a pilot product and embarked on construction of a prototype chatbot that was fed with an extensive dataset exceeding 10,000 pages from the central bank's Open Book Supervision, a vast array of laws, regulations, and supervisory data. The chatbot was engineered to field questions based on information in the documents, and a pilot utilizing this dataset yielded positive reactions from both supervisors and policymakers. This encouraging response, coupled with a strong interest from the institutions under supervision, propelled the team to advance the development of their product within the confines of the central bank.

ChatDNB, as it now stands, functions as a chat interface capable of addressing inquiries pertaining to select documents and is exclusively accessible to central bank personnel. The generative AI tool has streamlined the process of locating pertinent information and documents, thereby enhancing staff proficiency. Moreover, the tool is adept at providing interpretations of regulations when their original text is ambiguous. For instance, for users seeking to understand the considerations necessary when establishing a supervisory requirement for a bank, ChatDNB is equipped to respond with the applicable regulation and any published interpretations thereof. From an organizational standpoint, ChatDNB has the potential to bolster the consistency of supervisory activities.[20]

19 McCaul, Elizabeth. 2024. "From Data to Decisions: AI and Supervision." European Central Bank - Banking supervision, February 26, 2024. https://www.bankingsupervision.europa.eu/press/interviews/date/2024/html/ssm.in240226~c6f7fc9251.en.html.

20 Central Banking Staffs. 2024. "Initiative of the Year: The Netherlands Bank's Chatdnb." Central Banking , March 19, 2024. https://www.centralbanking.com/awards/7960892/initiative-of-the-year-the-netherlands-banks-chatdnb.

AI related risks and concerns

Artificial intelligence continues to demonstrate its enormous potential across various sectors, including financial supervision, yet its adoption is still often met with hesitation. This chapter explores the key reasons behind the reluctance of supervisory authorities to integrate AI, with a focus on the risks and concerns associated with its use.

The emerging AI regulations and guidelines, such as the UN Principles for the Ethical Use of AI, the EU AI Act, the US Blueprint for an AI Bill of Rights, and the NIST AI Risk Management Framework, all focus on addressing the risks associated with AI usage, including bias, transparency, accountability, and safety. These frameworks aim to mitigate the potential harms related to algorithmic discrimination, data privacy violations, and lack of explainability in AI decision-making. They emphasize the importance of ensuring that AI systems are developed and deployed responsibly, with mechanisms in place to protect individuals' rights and prevent unintended consequences. These same risks are key concerns for supervisory authorities when considering AI adoption.

AI-related concerns refer to the possibility that the expected benefits of using AI technology for effective RBS may not be fully realized. The primary concerns include:

- **Bias and stereotyping:** AI, especially LLMs, can inadvertently generate biased or prejudiced content, including sexist, racist, or homophobic content. These biases stem from the large datasets used for training these models, which may contain biased information. This is a significant concern for consumer-facing applications such as chatbots, where harmful stereotypes can be propagated to produce skewed results. Also, such biases are also highly relevant and critical for the financial sector, especially given the increasing focus of supervisors on gender diversity and financial inclusion. In the area of supervision, these concerns may lead to incorrect

risk assessments for certain types of products, customers, or geographical locations. Consequently, policy decisions by supervisory authorities could exacerbate financial exclusion or restrict access to finance for categories of customers who, in fact, do not pose any additional risks. However, it is important to acknowledge that risk assessments based on purely human judgment can also be biased in different ways. Data-based approaches to risk assessment, when properly designed and implemented, have the potential to reduce human biases by relying on more objective data. This highlights the importance of developing AI systems with fairness and transparency in mind to mitigate biases and enhance decision-making processes.

- **Robustness and stability:** AI systems can perform inconsistently when encountering data that is significantly different from their training sets. LLM-based AI systems may generate responses that seem accurate and confident but are factually incorrect, a phenomenon known as „hallucination." Accurate source citing is another challenge for these systems, potentially leading to unexpected or erroneous outputs. This is particularly problematic in real-world applications that require strict adherence to specific regulations and laws. AI-powered chatbots assisting supervised institutions or customers in identifying and interpreting relevant regulatory requirements are directly impacted by these issues.[21] In the area of supervision, this may lead to a disproportionate reaction by the supervisory authority to certain events. As a result, valuable time may be lost during which supervisory mitigation measures could have restored financial stability with relatively limited damage.

- **Interpretability:** The outputs provided by AI systems are not always transparent, explainable, or verifiable, which complicates their use in environments where clear justification of decisions is required, such as in risk-based supervision. A lack of full transparency in the output can also lead to delayed or disproportionate decisions by the supervisory authority. However, it is also important to consider that an overly transparent algorithm for identifying at-risk banks could inadvertently show banks how to manipulate data inputs to avoid being flagged as at-risk. Therefore, finding a balance between transparency and security is crucial to ensure effective and fair supervision.

21 Gov.UK. 2024. Generative AI Framework for UK Government, https://www.gov.uk/government/publications/generative-ai-framework-for-hmg/generative-ai-framework-for-hmg-html#principles

- **Compliance with regulatory requirements:** Laws and regulations governing AI are emerging across various jurisdictions, including the EU and US. These rules define where and how AI technology can be used, as well as the responsibilities of developers, providers, beneficiaries, and users of AI systems. Since these frameworks are still evolving, supervisory authorities are cautious about adopting technologies that may not comply with newly established regulations. Also, there may be additional costs associated with monitoring and maintaining the compliance.
- **Weakened internal knowledge:** The reliance on AI for tasks traditionally performed by human personnel may create a knowledge gap within supervisory authorities. This dependency on technology can lead to a loss of critical expertise and institutional knowledge. Excessive reliance on AI in supervision processes may undermine supervisors' abilities to exercise professional judgment and skepticism, which are critical for an effective RBS regime.

AI-related risks refer to the potential adverse business impacts that may arise from the deployment and use of AI by supervisory authorities. The key risks include:

- **Data privacy risks:** The use of AI models involves risks related to data privacy, particularly when sensitive and confidential data are processed, and potentially exposed through cloud-based or on-premises AI systems. Furthermore, strict data privacy laws might limit certain AI applications or implementations. Some AI models such as LLMs may inadvertently disclose sensitive information used for model training.
- **Intellectual property risks:** These risks arise from the way AI systems are trained, developed, and used, as well as from the data and content they generate or consume. They include concerns about ownership of AI-generated content, potential copyright infringement from the use of unlicensed data, and the legal ambiguity surrounding the patentability of AI-driven innovations.
- **Cybersecurity threats:** The integration of AI into financial supervision systems increases vulnerability to cybersecurity threats such as phishing, sensitive data extraction, data poisoning, and complex fraud. It is crucial for supervisory authorities to carefully weigh the cybersecurity risks against the business value provided by AI technologies.[22]

22 Federal Office for Information Security (Germany). 2024. Generative AI Models Opportunities and Risks for Industry and Authorities, Germany's Federal Office for Information Security (April 2024), https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Generative_AI_Models.pdf?__blob=publicationFile&v=4.

- **Misuse of AI:** AI tools, including AI-powered chatbots, could be misused to find and exploit loopholes in regulations, potentially facilitating fraudulent activities while ostensibly adhering to existing regulations.
- **Prompt injection:** This is a risk commonly associated with generative AI models, where the model can be manipulated to generate specific, potentially harmful content. It poses substantial risks when such tools are used in external communications with legal consequences for supervisory authorities.
- **Data poisoning:** This type of attack targets the training process of an LLM, where malicious actors intentionally introduce corrupt or biased data into the model's training dataset. The aim is to degrade the model's performance, introduce specific biases, or cause it to generate incorrect, harmful, or manipulated outputs. When using user feedback collected from publicly accessible AI-powered chatbots, extra caution must be taken to prevent such attacks.

The reluctance of supervisory authorities to adopt AI is rooted in valid concerns, as outlined above. To balance the benefits against the risks, careful planning, transparent communication, and a robust AI governance framework are essential. Fortunately, there are already established guidelines for implementing effective AI governance, including the UN Principles for the Ethical Use of AI, the EU AI Act, the US Blueprint for an AI Bill of Rights, and the NIST AI Risk Management Framework. A key element of these frameworks is comprehensive risk management for AI systems throughout their entire lifecycle. Additionally, professional organizations like ISACA are developing specialized audit programs for AI, including the assessment of AI algorithms. And last but not least, supervisory authorities are in a position to decide on when and how to apply the "human in the loop" principle in the development and use of AI. Some regulations may ask for this explicitly, like EU AI Act requirements for "high-risks AI systems."[23] All of these measures promote a risk-informed approach to AI adoption by supervisory authorities. As a result, AI adoption is expected to accelerate as organizations strengthen controls, gain experience, and build trust in AI systems.

23 EU Artificial Intelligence Act. https://artificialintelligenceact.eu/article/14/.

AI implementation challenges

Implementing AI within supervisory authorities holds great potential, yet it also creates challenges that must be addressed for successful adoption. The nature and severity of these challenges can vary across jurisdictions depending on factors such as the organization's digital maturity, current data governance practices and those applied over the years, and legal requirements for data protection and cross-border data processing. Although developing countries may face more intense challenges, there is not necessarily a direct correlation between the level of a country's economic development and the potential for it to implement innovative AI-based solutions. This is because, in many cases, supervisory authorities in developing countries enjoy greater financial independence compared to other state authorities. Additionally, these countries benefit from significant technical assistance and other forms of support from international organizations, including the World Bank. Below are some of the key challenges commonly faced by supervisory authorities.

Data availability and quality

AI is a data-driven technology that processes data, learns from it, and enables automated decision-making and problem-solving. The performance of AI models is as good as the data fed to them during model training, as well as model inference. Many supervisory entities discover important constraints when it comes to providing relevant, sufficient, and veracious data to build and run AI models. In particular, challenges are related to the following:

- **Relevant data:** refers to data that directly aligns with the objectives of the AI model and the specific tasks it aims to perform. For supervisory authorities, finding data that truly reflects the context of financial supervision or regulatory compliance can be challenging. Irrelevant or outdated data can distort AI predictions or decision-making, leading to inaccurate results. The challenge lies in identifying the right datasets from vast repositories, ensuring that the data represents the environment the AI system will be applied to. This requires close collaboration between domain experts and data scientists to ensure that the correct data is selected for training AI models.
- **Sufficient data:** machine learning algorithms, especially deep learning models, require large datasets to learn effectively and make accurate predictions. For supervisory authorities, gathering enough data can be challenging, particularly when it involves historical or sensitive information from regulated entities. Supervisory authorities may face challenges in data collection due to privacy laws, limited data-sharing agreements, or incomplete records, which hinder the model's ability to achieve reliable outcomes. Also, the limited data available for training a model in a specific language may lead to a significant reduction in the quality of the AI output.
- **Veracious data:** refers to data that is accurate, truthful, and free from bias. The integrity of AI models depends heavily on the quality of the data they are trained on. Supervisory authorities may face challenges in ensuring that the data they use is clean, reliable, and not skewed by errors or fraud. Incorrect or biased data can lead to misleading results and can undermine the trust in AI systems. Ensuring data veracity requires rigorous data validation, cleansing processes, and mechanisms to detect and correct biases before feeding the data into AI systems.
- **Accessible data:** supervisory authorities may face challenges in accessing the required data due to data silos, legacy systems, or restrictive data governance policies. Additionally, cross-border data regulations or sector-specific laws may limit access to certain datasets. Ensuring that data is both technically and legally accessible is critical for successful AI implementation. Data accessibility challenges can hinder AI systems' ability to operate in real time and impact the model's performance due to limited access to the context information required for accurate inference.

The above challenges can be addressed by reviewing and strengthening the organization's data governance framework. In addition to standard actions such

as establishing effective data management structures, processes, policies, and procedures, particular attention should be paid to the following:

- **Well-documented and communicated data models:** ensure a clear understanding of what data the supervisory authority holds and foster a common understanding among all involved parties, including supervised entities. This helps in identifying relevant data for developing and running AI models.
- **Clear data lineage:** implement tools and processes to track the flow, transformations, and usage of data throughout its lifecycle. This transparency helps ensure that data remains accurate, reliable, and accessible at every stage, while also aiding in compliance with regulatory requirements. This helps with selecting proper data sources for both AI models training, as well as for model inference.
- **Regular data quality assessments:** conduct routine data quality audits to ensure that data is veracious (accurate and reliable). This includes establishing mechanisms for ongoing validation and cleaning of datasets to maintain data integrity. Specialized AI models can be employed for this task.
- **Data accessibility and integration:** implement modern data management solutions that eliminate data silos and provide secure, real-time access to data. This ensures that AI systems can readily access sufficient and relevant data while complying with data protection and privacy laws.
- **External data sources integration:** establish formal data-sharing agreements with external entities, including government agencies, to ensure secure and ethical access to external datasets, when necessary. This helps to enrich the data model and provide more relevant data for AI training and model inference.

Addressing data-related challenges for AI adoption may take time for supervisory authorities and may involve complex changes, such as revising regulatory reporting to gather the necessary data. While pursuing these long-term objectives, authorities should also explore AI opportunities that can be leveraged with the data currently available. Despite some limitations, supervisory authorities still hold a significant amount of data, even if not all of it is immediately suitable for AI development. By focusing on what can be achieved with existing data, they can make progress toward AI adoption while working on overcoming long-term challenges.

Skills and expertise

Another significant challenge to AI adoption in supervisory authorities is the availability of the necessary skills and expertise. AI systems require a deep understanding of machine learning, data science, and algorithmic modeling, along with domain-specific knowledge of financial supervision and regulatory frameworks. However, many supervisory authorities face talent shortages in these specialized areas. Traditional skill sets in these organizations are often centered around legal, regulatory, and financial analysis, with limited exposure to advanced data analytics or AI technologies. This gap in expertise makes it difficult for authorities to effectively develop, implement, and oversee AI models, which are needed for critical functions such as risk assessment, fraud detection, and regulatory compliance. Moreover, even when AI specialists are brought on board, integrating them into the existing organizational structure can be challenging, as financial supervisors may not fully understand the capabilities and limitations of AI technologies. This misalignment can result in underutilization or improper application of AI tools, further hindering progress.

To overcome this challenge, supervisory authorities should invest in a comprehensive capacity-building strategy. This includes recruiting specialized talent in data science, AI, and machine learning, as well as offering continuous training to existing staff to develop foundational AI skills. Partnering with external experts, academic institutions, or technology companies can provide short-term access to expertise while internal capabilities are developed. Additionally, fostering a collaborative work environment where AI specialists and domain experts (e.g., financial regulators) work together can bridge the knowledge gap, ensuring that AI systems are applied effectively. Implementing mentorship programs and cross-training initiatives can further support knowledge transfer, helping staff adapt to new AI-driven approaches in supervision and regulation.

IT infrastructure capacity

One of the critical challenges for AI adoption in supervisory authorities is the IT infrastructure capacity needed to train, fine-tune, and run AI models. Training AI models, especially large machine learning and deep learning models, requires substantial computational resources, including powerful processors (e.g., GPUs or TPUs), large amounts of memory, and fast storage systems. This process is resource-intensive and can take days or even weeks, depending on the size

and complexity of the model, as well as the amount of training data involved. Supervisory authorities may not have the necessary on-premises infrastructure to support the high-performance computing demands of training AI models. Even fine-tuning pre-trained models or running AI models in production can strain traditional IT systems, which are often optimized for regulatory tasks and data storage rather than large-scale computations. Moreover, maintaining such infrastructure requires ongoing investments in hardware, energy, cooling, and IT staff, making it costly and logistically challenging for supervisory authorities to build and manage.

To overcome these infrastructure limitations, supervisory authorities can turn to cloud-based AI platforms. Leading cloud providers such as Amazon Web Services, Google Cloud Platform, and Microsoft Azure offer AI-specific infrastructure, including scalable compute resources (GPUs, TPUs), dedicated machine learning platforms, and pre-built models. These cloud solutions allow organizations to train and run AI models on demand, without the need for large up-front investments in physical infrastructure. Supervisory authorities can benefit from hybrid cloud strategies, combining cloud-based AI resources with existing on-premises systems. For instance, initial training and fine-tuning of large models can be performed in the cloud, while lighter, inference-based tasks can be handled locally. Cloud-based solutions also offer flexible pricing models, allowing organizations to pay only for the resources they use, making them cost-effective. Additionally, major cloud providers offer integrated tools for data security and compliance, which is crucial for the sensitive data handled by supervisory authorities. For example, cloud platforms enable organizations to select the geographic region where their data is stored, facilitating compliance with data localization regulations. Moreover, data sanitization tools and hybrid cloud architectures can be employed to ensure that sensitive data is processed exclusively on local infrastructure, further enhancing security. These technologies can enable developing countries to leapfrog from a low level of IT infrastructure development to cutting-edge technologies without the need for significant investment in expensive physical infrastructure.

Legal and regulatory constraints

Supervisory authorities face significant challenges when implementing AI due to evolving legal and regulatory frameworks surrounding AI technologies. Emerging regulations such as the EU AI Act and data protection laws impose strict requirements on how AI systems can be developed, deployed, and used,

especially in high-risk areas like finance. Authorities must ensure AI systems comply with privacy laws, data security regulations, and fairness principles, particularly when processing sensitive data. The complexity is heightened by cross-border operations, where AI systems must adhere to different jurisdictions' regulations. Additionally, AI systems' decision-making processes may lack transparency, leading to potential conflicts with regulations requiring explainability and accountability.

To address these challenges, supervisory authorities should establish a robust AI governance framework that ensures compliance with relevant regulations. Collaborating with legal experts and regulators during AI system development can help navigate complex legal landscapes. Supervisors should also invest in explainable AI (XAI) technologies to ensure the transparency and accountability of AI models. Compliance officers should be involved early in the process to align AI initiatives with regulatory standards, including data protection and fairness requirements. Supervisors can also use regulatory sandboxes, which allow for controlled experimentation with AI systems while ensuring compliance, enabling safe testing and innovation without breaching legal constraints.

Building trust in AI systems

Building trust in AI systems is a major challenge for supervisory authorities, as AI models often operate as "black boxes" with internal decision-making processes that are difficult to understand. This lack of transparency can make it hard for users to trust AI-generated recommendations or decisions, particularly in high-stakes areas like financial supervision. Concerns over bias, inaccuracies, and the potential for AI to make errors can further erode trust in these systems. This can lead to reluctance among supervisors to use and trust the deployed AI models.

To build trust, supervisory authorities should prioritize transparency and explainability in AI systems by adopting Explainable AI models that offer insight into how decisions are made. Regular audits of AI systems for bias, fairness, and accuracy should be conducted, with clear accountability structures in place. Adversarial testing is a good practice to ensure robustness of the developed AI models. Supervisory authorities can also foster trust by gradually introducing AI, starting with lower-risk applications and demonstrating consistent, reliable results over time. Additionally, authorities should communicate openly with stakeholders, including the industry and consumers, about AI's role and

limitations in decision-making processes, thereby promoting understanding and confidence in AI-driven supervision.

In conclusion, while supervisory authorities face significant challenges in adopting AI – ranging from data quality and IT infrastructure to skills gaps, regulatory concerns, and financial constraints – there are practical solutions to overcome these barriers. By investing in talent development, leveraging cloud-based AI infrastructure, and adhering to strong data governance practices, supervisors can harness AI's full potential. Building trust and ensuring compliance with emerging regulations are critical to sustainable AI implementation. The next section will explore strategic approaches to AI adoption, providing a roadmap for successful integration of AI within supervisory frameworks.

AI adoption strategies

There are many facets of AI adoption that must be managed by any organization deciding to harness AI technology. These include data governance, infrastructure readiness, skills and expertise development, ethical and regulatory compliance, and change management. Additionally, organizations must focus on building a supportive culture, managing expectations across all levels, and developing sourcing strategies that balance in-house capabilities with external partnerships. Each of these elements plays a critical role in ensuring a smooth and effective AI adoption process, enabling the organization to fully leverage the benefits of AI technology.

Though standard best practices for change governance should be applied by supervisory authorities, this chapter offers additional insights that are specific to AI adoption, addressing unique challenges and opportunities that arise in this context.

Managing C-suite expectations

Enthusiasm from the C-suite for adopting AI technologies often comes with high expectations but a limited understanding of specific objectives. To address this, supervisory authorities must ensure that top executives are well-informed about both the capabilities and the limitations of AI. Hosting strategic alignment workshops and targeted educational sessions can help bridge this knowledge gap, aligning AI initiatives with clear, measurable goals. This prevents unrealistic expectations and ensures that AI tools are deployed to solve well-defined problems. It's also important to emphasize that AI systems often require time to learn and improve, so managing the expectations curve accordingly is crucial to long-term success.

Cultural and organizational resistance

Resistance to AI within an organization often arises from fear of change and uncertainty about its impact on job roles. To address these concerns, supervisory authorities should implement comprehensive change management strategies. This involves clear communication about the benefits of AI and how it complements, rather than replaces, human roles. Detailed plans for workforce transition, including upskilling and reskilling initiatives, are essential to easing anxieties. Introducing pilot programs that allow employees to interact with AI in a controlled, low-risk environment can further help demonstrate AI's value and effectiveness, fostering greater acceptance and smoother integration into daily operations.

Balancing speed and scale of AI adoption

Supervisory authorities must carefully balance the speed of AI adoption with its scalability. While there is often pressure to move quickly to capture the benefits of AI, rushing can lead to incomplete solutions that may not align with long-term goals. On the other hand, starting with large-scale implementations can strain resources and introduce unnecessary complexity.

To manage this balance, authorities should adopt a phased approach, guided by the "think big – act small" principle. Begin by deploying AI in focused, high-impact areas where early wins can demonstrate value and build momentum. These initial successes provide valuable insights into AI's potential and limitations, shaping more strategic and scalable initiatives in the future.

Simultaneously, supervisors should develop a long-term AI roadmap that ensures AI investments are aligned with the broader enterprise architecture. This approach will allow for smooth scalability as the organization progresses along its AI journey, ensuring the system remains sustainable and adaptable.

To support this process, a more technical reference architecture for an enterprise AI platform, previously discussed with several supervisory authorities during World Bank technical assistance projects, is included in Annex 1. This framework provides guidance for integrating AI into supervisory operations while ensuring both immediate and future scalability.

AI sourcing strategies

Supervisory authorities must strike a balance between leveraging internal resources and external expertise when adopting AI technologies. While developing AI capabilities in-house offers greater control, customization, and alignment with organizational goals, it requires significant investment in specialized talent and infrastructure. Conversely, relying entirely on external providers can lead to dependency, lack of customization, and potential misalignment with specific supervisory needs.

To balance this, authorities should adopt a hybrid sourcing strategy that combines both internal and external resources throughout the AI model lifecycle. Early in the AI adoption journey, external resources – such as third-party AI models, cloud infrastructure, and consultancy services – can be leveraged to quickly build foundational capabilities and address immediate needs. These external partnerships can provide access to advanced tools, pre-trained models, and expertise that may not be available internally.

At the same time, it is essential to gradually build internal capabilities for model development, fine-tuning, and management. Training in-house teams of data scientists and AI engineers ensures long-term sustainability, customization, and independence. Over time, as internal capacity grows, authorities can shift from relying heavily on external resources to managing more of the AI lifecycle in-house.

This hybrid approach allows for quicker AI adoption while maintaining flexibility and control over the technology's future evolution. The ML & AI reference architecture in Annex 1 illustrates this sourcing model for AI technologies. A well-designed sourcing strategy roadmap, aligned with the organization's AI objectives, should define the roles of both internal and external resources at each phase of the AI model lifecycle, from initial proof of concept to large-scale deployment and continuous optimization.

AI adoption varies insignificantly between developed and developing countries, as cloud-based solutions offer a level of democratization in access to advanced AI technologies. Cloud services enable both developed and developing countries to access cutting-edge AI tools and infrastructure through pay-per-use models, reducing the need for significant up-front investments. This allows even resource-constrained organizations in developing countries to experiment

with and deploy AI models. Moreover, while developed countries may benefit from a wealth of historical data for training robust AI models, some developing countries have also accumulated high-quality, reliable data over the years, positioning them well for effective AI training. However, developed countries often have a broader range of options when it comes to sourcing and deploying AI solutions. They may opt for expensive on-premises infrastructure to maintain control and meet specific regulatory needs. This option is generally less accessible to developing countries due to financial constraints, making cloud-based AI platforms more attractive and practical for them. Nevertheless, with the right strategic planning, developing countries can leverage cloud solutions to close the gap, using AI to enhance their supervisory capabilities and achieve a level of technical sophistication comparable to their developed country counterparts.

Looking into the future

Given the substantial resources dedicated to the development of AI models, it is reasonable to expect that the pace of AI advancement will only accelerate. This progress will usher in new opportunities and applications for nearly every industry, including finance. While the future of AI functionality remains uncertain due to various dynamic and unpredictable factors such as regulation, computational power, and competition, we venture to forecast the influence of AI on supervisory activities in the coming years. It is important to emphasize that this is strictly the vision of the authors, and that at this stage there are no implemented and tested solutions that would perform these functionalities.

Below are our main visions with regard to the anticipated impact of the AI on supervision:

Real time supervision

The majority of supervisory practices, even in developed countries, operate on a post-factum basis. Supervisors receive information about transactions, operations, and decisions made or conducted by the entities they supervise with a delay, which can be substantial. Even the most basic reports are received with a significant time lag. For example, a monthly balance sheet report is submitted to the supervisor 10, 15, or even more days after the reporting period, and this balance sheet contains transactions that occurred at the beginning of the previous month. When we factor in the time required for validation, consolidation, processing, and analysis, we can conclude that when supervisors write their conclusions based on the analyzed data, the external reality may have changed significantly.

Despite the elimination of numerous obstacles that previously hindered the reduction of time between event instances of reporting in recent years – including but not limited to a lack computational power, high data storage costs, slow data transmission capacity, and a lack of sufficient reporting tools – the time gap has not been significantly reduced. The primary reason for this

stagnation is the supervisory authorities' limited capacity to process real-time or near real-time data. It would be pointless to require financial institutions to report real-time data if supervisors are only capable of reviewing this data two to three weeks later.

AI holds the potential to revolutionize this scenario by analyzing real-time data almost instantaneously, including outside of standard working hours. In the initial phase, AI tools may be tasked with analyzing real-time data and flagging transactions and operations that may necessitate human intervention and decision-making. However, from a medium-term perspective, AI could be entrusted with conducting preliminary investigations of red flags or anomalies (e.g., frauds detection, ML/TF, related parties transactions, and others). AI tools would be permitted to request additional information from the financial institutions (e.g., Know Your Customer files, contracts, decisions) and even liaise with risk offices or other staff in the institution to obtain explanations and clarifications. In this context, an alert to the supervisors would only be generated if the preliminary investigation still requires human intervention. Incorporating a preliminary investigation function into an AI tool will significantly augment the capacity of supervisors to process real-time transactions and, of course, will provide supervisors more time to savor their morning coffee.

With the adoption of AI tools anticipated to occur on both sides (i.e., supervisors and financial institutions), it is projected that in the long term, the AI tool of the supervisory authority will likely interact with the AI tool of the supervised entities to obtain clarifications on transactions on a 24/7/365 basis. Real-time supervision will prove particularly beneficial for areas such as consumer protection, AML/CFT, resolution and pre-resolution stages, market conduct, liquidity risk (bank run), fraud detection, and related parties' transactions.

However, in practice, not all supervisory processes will necessitate a transition to real-time supervision. Numerous areas of supervision will continue with post-factum supervision, including with significant delays, despite the technological capability to offer real-time supervision.

Data integration

Comprehensive and integrated data is essential for effective supervision, providing a complete and nuanced perspective on specific events or entities.

Fragmented or narrow-scope data, by contrast, can be unhelpful and even potentially misleading, directing attention away from actual risks and vulnerabilities. There has been substantial recent progress in the integration of structured data by supervisory authorities, with many supervisors successfully consolidating such data into efficient data warehouses, including cloud-based solutions. However, it is important to recognize that structured data alone does not encompass all the information needed for supervisors to execute their duties. A substantial amount of valuable and sometimes crucial information resides within unstructured data, such as reports, letters, emails, and diagrams, all of which are integral to informed decision-making processes.

The task of integrating structured and unstructured data in a manner in which they enhance one another and aid supervisors in timely risk identification is not a trivial one. However, given the rapid advancements in AI capabilities, we can anticipate that this complex task will become commonplace in the coming years. The efficiency and quality of merging structured and unstructured data, as well as the extraction of valuable insights from it, are expected to grow exponentially. Concurrently, the accessibility of such solutions, particularly in terms of cost, is also projected to increase.

In terms of practical application, nearly all supervisory domains stand to gain from these advancements. This ranges from AI tools capable of reading or monitoring in real-time board minutes, to AI tools designed to scrutinize advertising materials from financial institutions on social media and identify potentially misleading information for consumers. The integration of structured and unstructured data will pave the way for new opportunities for supervisors, facilitating a shift from reactive to proactive supervision.

Forward-looking modeling

Forecasting the future is a challenging endeavor across all sectors, and it becomes exceedingly complex within the financial sector. The close interconnection of national financial systems with the global market means that many pivotal elements shaping domestic financial dynamics are beyond the control of local authorities. Hence, any predictive modeling must consider both domestic and international influences, along with their likelihoods. Achieving accuracy in this process demands a substantial investment in human resources. Additionally, these predictive models require regular reviews and adjustments to

align them with actual occurrences, a process that is both time-consuming and complicated.

Due to the substantial human resources required, most predictive models are significantly simplified and only incorporate a limited set of parameters and variables. Consequently, these models often lack accuracy and proper calibration.

AI tools present a promising solution to these challenges. They can process vast amounts of data at a fraction of the cost and recalibrate existing models based on actual data much more rapidly than humans.

Key supervisory activities that stand to gain from these capabilities include stress testing (especially macroprudential), early warning indicators, crisis simulation, and validation of recovery and contingency plans. In terms of credit risk, these AI models can be utilized to estimate a debtor's default probability. They can also validate banks' internal capital adequacy assessment processes and internal liquidity adequacy assessment processes.

Full (end-to-end) automation of some processes

While most AI solutions are currently designed to assist humans, particularly in processes where human intervention is crucial, we foresee a future where certain supervisory processes will be fully automated, eliminating the need for human involvement. This will be particularly feasible with the rapid development of Autonomous AI agents that will incorporate capabilities such as learning, memory, decision-making, and adaptability. At the initial stage of development this end-to-end automation will not apply to complex procedures such as on-site inspections or the execution of resolution plans. However, in the long term, even more complex procedures can benefit from the use of Autonomous AI agents.

Processes that can be automated from end-to-end at the initial stage are typically those that require limited discretion and are largely based on box-checking. For instance, the horizontal reassignment of a financial institution's administrator is one such process. In some jurisdictions, even when an administrator is reassigned laterally (e.g., a branch director reassigned elsewhere), approval from the supervisor is still required. This approval process, which usually takes a simplified form and does not include an interview or

knowledge testing of the candidate (as has been done previously), is primarily a box-checking activity that could likely be automated with AI tools.

This end-to-end automation will also affect other processes that do not involve complex professional judgment. It may also be applied to new processes that are currently not conducted due to a lack of human resources. An example could be verifying whether supervised entities are disclosing information to the public in accordance with legal requirements. Checking the information disclosed by these entities on their websites is time-consuming but not particularly sophisticated, and can be conducted even by individuals with limited supervision expertise. These types of activities will likely be the first to be fully automated.

In the long term, even more complex supervisory processes could be subject to full automation. The motivation to automate some processes end-to-end may stem from authorities' desire to reduce corruption factors. In some cases, the risk of an AI algorithm generating occasional fixable errors may be more acceptable than having a high vulnerability to corruption cases, which could lead to severe reputational damage.

Dark web integration

The dark web serves as a valuable information source for authorities, especially in areas such as AML/CFT, fraud monitoring, consumer protection, and market conduct supervision. Despite the fact that a significant portion of data on the dark web lacks credibility, it is also a hub for trading stolen credit card information, exposing compromised personal information from private and public databases, and circulating information about new fraud schemes more rapidly than on the surface web.

At present, supervisors utilize minimal information from the dark web, primarily due to its low credibility and the substantial effort required to monitor and filter the information. However, these challenges can be addressed by AI tools. AI solutions can perform 24/7/365 monitoring of the dark web for any information that may indicate potential vulnerabilities in the financial sector, such as customer data of financial institutions, information about a financial institutions' internal control system, or information about AML/CFT flags used by financial institutions or authorities.

Combining information from the dark web with data collected by supervisors from traditional channels and public sources can help identify vulnerabilities that criminals could exploit at an early stage. In the coming years, AI may evolve to conduct unsupervised discussions with potential criminals on the dark web, fully emulating human behavior to extract additional information or document data potentially related to money laundering or terror financing.

Real time consultation/feedback

Instances where a supervisory authority is asked to provide consultation or feedback to a financial institution, customers of these institutions, the media, or other parties are quite common. While these activities may not be critical for maintaining financial stability, they are still part of supervisors' responsibilities and can be time-consuming. For example, interpreting a legal provision for a financial institution can be a lengthy process, even if a similar request was recently addressed for a different institution. The same applies to routine information requests from the general public, media, or other national authorities.

Real-time consultation and feedback to various stakeholders is a fundamental feature of AI tools. Despite occasional frustrations with AI chatbots that fail to answer simple questions correctly, we are rapidly moving toward a scenario where we will request a human operator to connect us with an AI chatbot for quicker issue resolution.

These AI capabilities will also revolutionize how supervisors interact with their stakeholders. The mandate for AI to handle processes without any human intervention will expand rapidly. We will quickly transition from AI tools that assist us in finding the right phone number to AI tools that autonomously handle the resolution of complex customer complaint cases in real time.

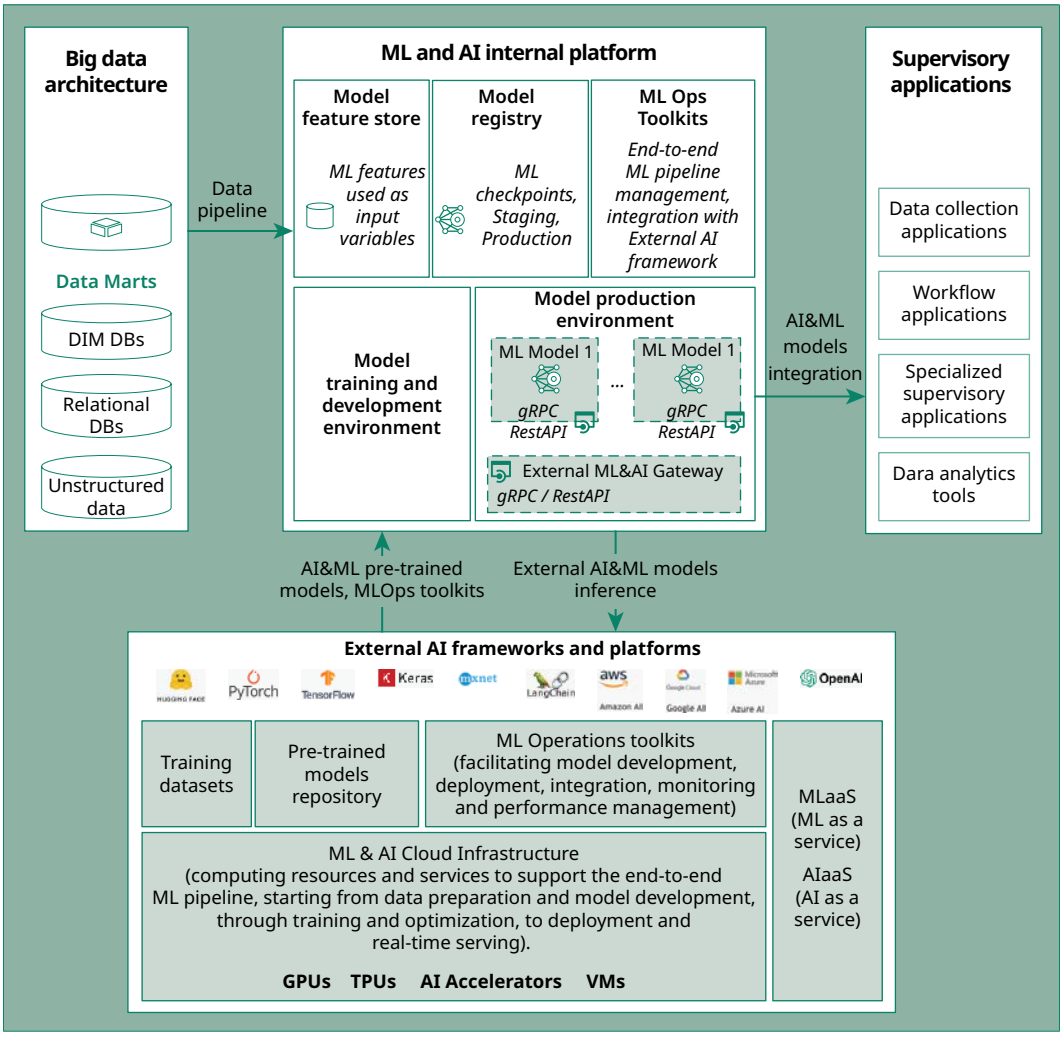
While the rapid advancement of AI has sparked fear in many, primarily due to concerns about job displacement, the prospect of AI replacing financial sector supervision is not imminent. This assertion is backed by several factors, the most significant being the extreme complexity of financial sector supervision activities and the critical role the financial sector plays in any national economy.

Despite the strides made in AI, we are still a considerable distance from entrusting the reins of the financial sector to AI. However, what we are likely

to witness, and what this research aims to capture, is an increasing symbiosis between supervisors and AI tools.

The reliance of supervisors on AI solutions is set to grow exponentially, unlocking potential in data analytics, modeling, and forecasting that were previously inconceivable. This evolution will not only enhance the efficiency and effectiveness of financial supervision but will also pave the way for innovative approaches to managing financial risks and ensuring economic stability.

Annex 1- ML & AI enterprise reference architecture



Key Flows of the AI Reference Architecture

External AI Framework and Platforms

The organization can download various resources for internal use from external AI frameworks and platforms, including pre-trained machine learning models, ML operations toolkits, and training datasets (especially when in-house data is insufficient). These resources can also be accessed and utilized directly on the external platform's ML & AI cloud infrastructure.

Feature Store and Data Lake Integration

The Feature Store, built on top of the Data Lake, stores static features that serve as input variables for ML models. These features can be extracted using time-based batch jobs or event-based triggers. The Feature Store acts as a data mart specifically designed for ML models, providing real-time data for models in production and batch data for models in development. It also supports versioning and tracking of features, allowing data scientists to analyze model performance over time. Examples of stored features include financial institution names, types of financial products, and historical financial indicators.

ML Model Integration and API Access

ML models are integrated into the organization's application architecture as microservices, accessible by different supervisory applications via gRPC or RESTful APIs. Some input parameters for the models are included in the request, while others are retrieved from the Feature Store via a feature-serving API. Once all inputs are defined, the model processes them, and the output is sent back to the requesting supervisory application. Additional orchestration capabilities can chain multiple models to improve performance, with load-balancing ensuring scalability and optimal performance.

Third-Party ML & AI Services

Certain ML & AI models can be provided by third-party vendors and hosted on their cloud platforms. These may include public shared services or private, tailored models for the organization. Access to these models is managed via inference APIs, with requests moderated through an External ML & AI Gateway to ensure governance, including authentication, access control, monitoring, and other management tasks essential for large-scale application consistency.


WORLD BANK GROUP


FINANCE FOR
DEVELOPMENT


Technology
Innovation